AI and Electoral Campaigns

Rafael Rubio Núñez
Frederico Franco Alvim
Vitor de Andrade Monteiro

WILEY Blackwell

AI and Electoral Campaigns

AI and Electoral Campaigns

Rafael Rubio Núñez

Frederico Franco Alvim

Vitor de Andrade Monteiro

WILEY Blackwell

Copyright © 2026 by John Wiley & Sons, Inc.

All rights reserved, including rights for text and data mining and training of artificial intelligence technologies or similar technologies.

Published by John Wiley & Sons, Inc., Hoboken, New Jersey.
Published simultaneously in Canada.

No part of this publication may be reproduced, stored in a retrieval system, or transmitted in any form or by any means, electronic, mechanical, photocopying, recording, scanning, or otherwise, except as permitted under Section 107 or 108 of the 1976 United States Copyright Act, without either the prior written permission of the Publisher, or authorization through payment of the appropriate per-copy fee to the Copyright Clearance Center, Inc., 222 Rosewood Drive, Danvers, MA 01923, (978) 750-8400, fax (978) 750-4470, or on the web at www.copyright.com. Requests to the Publisher for permission should be addressed to the Permissions Department, John Wiley & Sons, Inc., 111 River Street, Hoboken, NJ 07030, (201) 748-6011, fax (201) 748-6008, or online at http://www.wiley.com/go/permission.

The manufacturer's authorized representative according to the EU General Product Safety Regulation is Wiley-VCH GmbH, Boschstr. 12, 69469 Weinheim, Germany, e-mail: Product_Safety@wiley.com.

Trademarks: Wiley and the Wiley logo are trademarks or registered trademarks of John Wiley & Sons, Inc. and/or its affiliates in the United States and other countries and may not be used without written permission. All other trademarks are the property of their respective owners. John Wiley & Sons, Inc. is not associated with any product or vendor mentioned in this book.

Limit of Liability/Disclaimer of Warranty: While the publisher and the authors have used their best efforts in preparing this work, including a review of the content of the work, neither the publisher nor the authors make any representations or warranties with respect to the accuracy or completeness of the contents of this work and specifically disclaim all warranties, including without limitation any implied warranties of merchantability or fitness for a particular purpose. No warranty may be created or extended by sales representatives, written sales materials or promotional statements for this work. The fact that an organization, website, or product is referred to in this work as a citation and/or potential source of further information does not mean that the publisher and authors endorse the information or services the organization, website, or product may provide or recommendations it may make. This work is sold with the understanding that the publisher is not engaged in rendering professional services. The advice and strategies contained herein may not be suitable for your situation. You should consult with a specialist where appropriate. Further, readers should be aware that websites listed in this work may have changed or disappeared between when this work was written and when it is read. Neither the publisher nor authors shall be liable for any loss of profit or any other commercial damages, including but not limited to special, incidental, consequential, or other damages.

For general information on our other products and services or for technical support, please contact our Customer Care Department within the United States at (800) 762-2974, outside the United States at (317) 572-3993 or fax (317) 572-4002.

Wiley also publishes its books in a variety of electronic formats. Some content that appears in print may not be available in electronic formats. For more information about Wiley products, visit our web site at www.wiley.com.

Library of Congress Cataloging-in-Publication Data Applied for:

Paperback ISBN: 9781394311781
ePDF: 9781394311804
ePub: 9781394311798
Obook: 9781394311811

Cover Design and Image: Wiley

Set in 10.5/13pt STIXTwoText by Straive, Pondicherry, India

Contents

About the Authors

RAFAEL RUBIO NÚÑEZ

Doctor of Constitutional Law and Professor of Constitutional Law at the Complutense University of Madrid. Specialized in electoral law, parliamentary law, and the fundamental rights of participation and freedom of expression. Visiting researcher at Georgetown University, Harvard University, George Washington University, Scuola Superiore Sant'Anna, and Dublin City University. He has been President of the Council for Transparency and Participation of the Madrid regional government, Deputy Director of the Center for Political and Constitutional Studies, and a member of the Spanish government's Open Government Forum as well as the Venice Commission. He is currently a Researcher at the ICCAL Lab at the Max Planck Institute of the University of Heilderberg (25/27), co-director of the Complutense Observatory of Disinformation, and co-principal investigator of the project: "Institutional and regulatory safeguards, electoral authorities and digital supervision against interference, hostile narratives, targeted advertising and polarization". He is a member of the Academic Committee of the Global Network for Electoral Justice, of the Advisory Board of the Media Observatory, of the Editorial Committee of the Revista del Estado (Colombia), of the Scientific Committee of the Center for the Study of Political Parties (National University of Distance Education; UNED), and of the Research Group on Technology and Democracy (i+dem) of the Complutense University of Madrid. Winner of two Electoral Justice Innovation Awards, in the categories of "Transparency and Strengthening the Image of the Electoral Justice" and "Specific progress in a topic relevant to the GNEJ" (2023).

FREDERICO FRANCO ALVIM

Doctor in Legal and Social Sciences (University of Argentine Social Museum; UMSA); doctoral student in Political Science (University of Lisbon); Master of Law (Methodist University of Piracicaba; UNIMEP)

with a Capes PROSUP scholarship; specialist in Electoral Law (National Autonomous University of Mexico; UNAM); specialist in Electoral Law and Procedure (Federal University of Goiás; UFG); specialist in Judicial Power with an emphasis on Electoral Law (Electoral Judicial School of Mato Grosso; EJE/MT); Diploma in "Disinformation and Electoral Campaigns" (Complutense University of Madrid). Coordinator of the Supreme Court's Fight against Disinformation; former Head of the Special Office of the General Secretariat of the Presidency and of the Special Office for Combating Disinformation of the Superior Electoral Court. Founding member of the Brazilian Academy of Electoral and Political Law (ABRADEP). Author of the following books, among others: *Manual de Direito Eleitoral* (Fórum, 2012), *Curso de Direito Eleitoral* (Juruá, 2016), *Cobertura Política e Integridade Eleitoral: Efeitos da Mídia Nas Eleições* (Habitus, 2018), *Abuso de Poder Nas Competições Eleitorais* (Fórum, 2024), and *Crise Democrática e Justiça Eleitoral: Desafios, Perspectivas e Pistas de Ação* (TSE, 2021). Co-author of the following books, among others: *Dicionário das Eleições* (Juruá, 2020), *Guerras Cognitivas Na Arena Eleitoral: O Controle Judicial da Desinformação* (Lumen Juris, 2023), *Glosario Contra la Desinformación* (INE, 2022), and *Inteligência Artificial para Eleições (Mais) Livres, Justas e Eficientes* (Lumen Juris, 2025). Awarded the "90 Years of Electoral Justice" medal by the President of the Brazilian Superior Electoral Court (2022). Winner of first Electoral Justice Innovation Award, in the category "Transparency and Strengthening the Image of the Electoral Justice" (2023).

VITOR DE ANDRADE MONTEIRO

Doctoral candidate in Constitutional Law at the Complutense University of Madrid; Master of Public Law (Federal University of Alagoas); Visiting researcher at the Max Planck Institute for Comparative Public Law and International Law (Heidelberg, Germany); Specialist in Procedural Law (Escola Superior da Magistratura do Estado de Alagoas; ESMAL); Diploma in "Disinformation and Electoral Campaigns" (Complutense University of Madrid). Former Brazilian Superior Electoral Court (TSE) International Affairs Advisor; former member of the Special Office for Combating Disinformation of the TSE of Brazil. Judicial Analyst at the Brazilian Electoral Court; member of the Brazilian Academy of Electoral and Political Law (ABRADEP), of the American Conference of Subnational Electoral Organizations for

Electoral Transparency (CAOESTE), of the Complutense Observatory of Electoral Disinformation, and of the Research Group on Technology and Democracy (i+dem) at the Complutense University of Madrid. He is the author and co-author of several academic articles and books. Winner of first Electoral Justice Innovation Award, in the category "Transparency and Strengthening the Image of the Electoral Justice" (2023).

Prologue

Before beginning this prologue, I would like to emphasize that for an academic work to be truly relevant, it must, in my opinion, meet two conditions: first, it must be formative and informative. It requires research, explanation and depth; second, it must be timely, that is, it must respond to a need to express or solve a problem. This book fulfills both conditions. Artificial intelligence (AI) is now part of our lives, whether we like it or not, and in that sense we must not only live with it as spectators or even victims of AI; we must get to know it, work with it, and more importantly, make it work for us. There is no better way to do this than through trial and error, that is why I want to be very clear about how this prologue was constructed and I am very clear in using the term "constructed" because more than writing it, constructing it was precisely what I did. When Dr. Rafael Rubio honored me by inviting me to write the foreword to this work, I thought of doing something new and not simply reading the work and summarizing its relevance. I endeavored to work with AI, thinking that in the near future most written works will have some elements of AI in its confection.

First, I went through the text carefully reading its content. This facilitated the next step of engaging in a conversation with ChatGPT, with the help of Karen Garzón and Yerutí Mendez from my team, who are more familiar with the tool. We asked ChatGPT to give us a summary of the book in various lengths, two pages, five pages, ten pages. I didn't like them. It did not capture the richness of the text. We decided to ask it for a summary per chapter and bingo, it did a great job. We all learned in the process, both the AI and us. Then, highlighting the topics that were important to me, we asked ChatGPT to write a text as if it were a prologue. It was not initially to my liking, but it did present me with something initial to work with. Finally, I added my own reflections, comments, and questions that seek to guide the text. Therefore, I say that this prologue was made as a result and proof of the collaboration between artificial and human intelligence. The texts resulting from the interactions with ChatGPT or META AI are in italics and respond to the prompts that we raised from my concerns. Here is the result.

On the threshold of a new technological era, AI has emerged as a central player in the transformation of electoral processes. Former Google CEO Eric

Schmidt recently said in an interview[1] that we can expect a new model of AI every 12 to 18 months with new and improved capabilities. In addition, he warns of three profound changes and a serious risk.

Schmidt first mentions changes in the "context window" which refers to, according to Meta´s AI tool, META AI, *"the amount of text or data that the model considers when processing a particular piece of information. It is the range of text that the model examines to understand the context and relationships between words, phrases, and ideas. By expanding the context window, AI models can capture long-range dependencies and better understand the broader text context, leading to better performance and more accurate results. However, larger context windows also increase computational requirements and can make models more prone to error or bias."*

The above definition and context were given to me by META AI, a tool that anyone in the world who has WhatsApp can use. It is undoubtedly a very powerful tool for those who have it;[2] however, it creates a huge disadvantage for those who cannot consult whatever their imagination can think of in a matter of seconds in the palm of their hand.

Going back to Schmidt. The changes in the context window are already giving way to something they call "chain-of-thought reasoning," which, again according to META AI, is the capability that allows AI to:

- Understand complex arguments.
- Follow multi-step reasoning.
- Establish logical connections between ideas.
- Generate more coherent and plausible text.

By developing chain-of-thought reasoning, AI models can move beyond simple pattern recognition and improve their ability to think critically, solve problems, and dive into more sophisticated decision-making processes. Science fiction? Think again. By increasing the context window, AI tools are developing a chain of thought that not only helps us solve problems; it solves them for us and immerses itself in the decision-making process. And how does it do this? By means of something called "agents." According to META AI: *In Artificial Intelligence (AI), an agent is a program or system that acts*

[1] Noema Magazine (2024). The Future of AI, According to Former Google CEO Eric Schmidt. Available at: `https://m.youtube.com/watch?v=DgpYiysQjeI`.

[2] It is estimated that more than two billion people in the world have WhatsApp, according to META data.

autonomously, making decisions and performing actions to achieve specific goals. Agents can interact with their environment, perceive information, and adapt to changing situations.

Types of agents:

1. *Simple reflexive agents: react to specific inputs with predefined actions.*
2. *Model-based reflective agents: maintain an internal model of the environment and make decisions based on that model.*
3. *Goal-based agents: pursue specific objectives and make decisions to achieve them.*
4. *Utility-based agents: make decisions based on a utility function that estimates the desirability of actions.*
5. *Learning agents: improve their performance over time through experiential learning.*

Examples of applications:

1. *Virtual assistants (such as Siri or Alexa).*
2. *Robotics (such as autonomous vehicles or robots).*
3. *Games (such as chess or poker).*
4. *Recommendation systems (such as personal product suggestions).*
5. *Autonomous systems (such as drones or autonomous cars).*

Are you afraid yet? We should be, because Schmidt tells us that the third thing, that text to action (T2A), is already happening and META AI refers to it as the ability of a model to generate actions or commands from text inputs. It involves processing natural text to identify specific tasks, intentions, or goals and then executing corresponding actions or generating code to achieve those goals.

The T2A models can be applied in multiple areas such as:

1. Virtual assistants: generating actions from voice or text commands.
2. Chatbots: responding to user queries with specific actions or recommendations.
3. Automation: executing tasks or workflows based on textual instructions.
4. Programming: generating code from natural descriptions.
5. Robotics: controlling robots or devices with textual commands.

The T2A process generally involves:

1. Natural language processing (NLP): analyzing text inputs to identify intentions, entities, and context.
2. Intent detection: determining the specific goal or task of the text input.
3. Action generation: creating a corresponding action or command to achieve the identified intention.
4. Execution: performing the generated action or executing the generated code.

T2A has numerous potential applications, including enhancing virtual assistants, simplifying automation processes, and improving human–computer interaction. However, it also faces challenges such as accurately understanding natural language, handling ambiguity, and ensuring safety and security in the execution of generated actions.

These are the three new developments in the world of AI, and what does Schmidt say is the biggest risk? Well, the risk is that agents will start communicating with each other to solve problems intuitively in a language that the human mind cannot understand. That is why the ethics, regulation, and transparency mentioned by the authors throughout the book are fundamental.

As an observer of electoral processes I can comment that the use of AI is still incipient in elections, but given the capabilities of this tool it is simply a matter of time. Some positive and negative examples showing different uses of AI in electoral processes are as follows. In Pakistan, imprisoned candidate Imran Khan campaigned from jail using AI. His party put voice and images to the political leader´s words. In Paraguay, a civil society organization dissatisfied with the fact that there was no presidential debate introduced the platforms of the two main candidates to ChatGPT and generated a fictional debate by means of cartoons. In the United States, less than five hours after President Biden´s re-election bid was announced, the Republican party launched a 100% AI-generated advert. In Argentina, AI-generated images flooded social networks praising both candidacies. AI is already a reality in electoral processes.

This book provides an in-depth analysis of how AI is reshaping the contemporary electoral landscape and is a guide to understanding the cyberpathologies and systemic threats that the new political communication faces in the digital age. While exploring the potentials of using AI in electoral

processes, it highlights risks and underscores the urgent need to promote transparency, fairness, and accountability in its use.

Currently, disinformation is classified into different categories, something that until a few years ago was perceived as a homogeneous phenomenon. In this regard, AI has proven to be a powerful tool for the creation and massive distribution of fake news, deepfakes, and other forms of disinformation. These technologies can generate highly convincing content that is difficult to distinguish from reality, posing serious challenges to the veracity of information circulating in the public space. The ability to produce false images, videos, and audios with great accuracy not only distorts the perception of facts but also undermines trust in institutions and democratic processes.

In addition to misinformation, AI can obstruct access to factual reality through information overload. This phenomenon, known as infodemia, uses mass diffusion tools such as spreaders, spambots and robotic accounts to propagate misleading narratives. These intensive and well-orchestrated disinformation campaigns create a fog of misinformation (smokescreens) that confuses citizens and makes it difficult to form informed opinions. Firehosing, or disinformation bombardment, exacerbates this problem by flooding the information space with contradictory and false messages, misguiding the audience and making it difficult to identify the objective truth.

Advanced AI techniques can influence voters in subtle but powerful ways, altering their perceptions and decisions without them being aware of it. This type of surreptitious manipulation represents a significant threat to voting autonomy and freedom of choice, the cornerstones of any democracy. As detailed in this book, AI can personalize political messages based on detailed data analysis, which increases the effectiveness of campaigns but also the risk of manipulation. This ability to micro-segment and target specific messages to different demographic groups makes it possible to influence voters almost imperceptibly, enhancing persuasion and in many cases, distorting the perception of political and social reality.

In the electoral context, the implementation of adequate regulations and the promotion of ethical and responsible use of AI are essential to ensure the integrity of the democratic process. This book underscores the need for a robust regulatory framework that protects privacy and civil liberties, ensures transparency and traceability of algorithms, and prohibits the use of deepfakes and bots in an attempt to prevent misinformation. In addition, it emphasizes the importance of accountability of AI developers and public education on AI in the electoral arena to strengthen the ability of citizens to make informed decisions free from undue influence.

The book also makes a call to action for all actors involved in the electoral process – candidatures, political parties, citizens, and electoral authorities – to adopt an ethical and responsible stance on the use of AI. The democratization of technology must be accompanied by a firm commitment to transparency, fairness, and accountability, thus ensuring that AI is used to strengthen and not weaken our democracies.

Modern technopolitics, where AI plays a central role, requires a re-evaluation of the communication strategies and tools used in electoral campaigns. Personalization of political messages based on data analytics and algorithms can improve the effectiveness of messaging, but it also poses risks of manipulation and excessive segmentation that can further fragment the electorate and foster polarization.

Another troubling aspect of the use of AI in politics is its ability to facilitate harassment and discrimination practices. Digital platforms can amplify hate speech and violence, creating hostile environments that deter civic and democratic participation. This phenomenon not only affects the individuals directly targeted but also deteriorates the quality of public debate and social cohesion. In addition, algorithmic manipulation can break political-ideological neutrality in the ranking of search engine results, user feeds, and content suggestions, thus harming candidates linked to vulnerable or minority groups and favoring the dissemination of prejudiced or stereotyped content.

In addition to what is pointed out in the book, I believe that the debate on the legitimate and illegitimate uses of AI in elections is crucial, but it goes beyond a simple legal issue. This problem must be analyzed from sociological, psychological, and economic perspectives due to the large gap between the rules of the market and the traditional ethical values that have governed the global economy. It is also essential to address the growing digital divide not only between countries, but also between industries. In the case of countries, these innovations require a comprehensive change ranging from the education system to the steering role of the state economy, considering AI as a disruptor in various industries such as manufacturing, services, and the exploitation of raw materials. On the business side, technology companies are acquiring inordinate power, dominating external markets, and taking over the means of communication and distribution of ideas through unknown algorithms.

There must also be a fundamental evolution in electoral authorities that on the one hand develop work plans in aspects where AI can be useful to implement but at the same time elaborate risk analysis and possible vulnerabilities in the face of these new tools.

There is a lack of understanding on the part of governments and citizens about technological products and services as well as about emerging

economic models that attempt to regulate the effects of technology rather than its causes. Effective regulation must focus on the root of the problems. For example, bots are going to exist if the business model is based on regulating clicks. Only through a holistic approach that considers all of these aspects can we develop a regulatory framework that ensures the ethical and responsible use of AI in elections and other crucial areas Gerardo de Icaza is the author of the prologue.

Acknowledgments

Every original work builds upon the previous work of others, and this is the case with this book. It has its origins in the Seminar of the Council of Europe and the Venice Commission[1] about artificial intelligence (AI) and elections, where Professor Rubio acted as rapporteur, discovering a new and largely unexplored world. This work also owes much to the elections held in Brazil in 2022, which allowed the authors to get to know each other and begin working together. This, in turn, gave rise to a series of articles describing how the Superior Electoral Court of Brazil tried to respond to disinformation, to a first book on AI and high-risk elections (*Inteligência Artificial e Eleições de Alto Risco. CiberpatologIas e Ameaças Sistêmicas da Nova Comunicaçao Politica,* published in Brazil by Lumen Juris), and a second one (*Inteligencia Electoral y Campañas Electorales Algorítmicas,* published in Spain by CEPC), all of which have provided the foundations for this work.

We would like to acknowledge the research projects "Cybersecurity in electoral processes. Guarantees against disinformation and other informational disasters on platforms" (TED2021-130876B-100) and "Institutional and regulatory guarantees. Electoral and digital supervisory authorities in the face of interference, hostile narratives, segmented advertising and polarization" (PID2022-137245OB-I00) as well as the Research Group on Technology and Democracy at the Complutense University of Madrid.

Finally, we would like to thank all those who revised different versions of this text: Ana Ibarz, Ramón Uría, Tamara Álvarez, Eliana Andrea Duitama, Valentina Martínez, Cristina Manrique, Zyanya Avilés, Álvaro Petit, and Miguel Ángel Gonzalo; all those who attended the "AI and Electoral Campaigns" seminar at the Institute of Parliamentary Law at Complutense University; and all those who have sent us their comments on previously published works. It is due to all of these people that the work you hold now is a little better than it might otherwise have been.

[1] The Nineteenth European Conference of Electoral Management Bodies "Artificial Intelligence and Electoral Integrity." Available at: `https://www.coe.int/en/web/venice-commission/-/event-3406`. Viewed: 15.05.2025.

Introduction

Technological acceleration, the rise of social networks, data mining, and the automation of creative processes in different formats all lead to a progressive virtualization of reality, to the "digitalization of public conversations" (Innerarity; Colomina 2020, p. 11). Within the realm of the symbolic conflicts that dominate politics at this point in the twenty-first century, these phenomena have had an intense impact on the behavior of organizations, political parties, candidates, and activists, transforming not only politics but also society more broadly through the accelerated "dehumanization" of sensitive social tasks and processes.[1]

Electoral campaigns, originally based on direct interpersonal contacts (in the nineteenth and early twentieth centuries), on mass media as the intermediary (from the mid-twentieth to the early twentieth centuries), and on the multi-platform modality characteristic of digitalization (present day), have retained human behavior as their main source of action, despite transformations in the sociotechnical media. In this context, the quest for votes was based on strategies conceived and executed with the logistical support of the techniques available at any given moment, giving a competitive advantage to those whose strategic technical innovation surpassed that of their opponents. For example, the emergence of information technology from Ronald

[1] The term "dehumanization" in this context is justified by the fact that the world is currently experiencing the "Second Machine Age, in which machines not only complement human beings, as in the Industrial Revolution, but also replace them" in a range of tasks, professions, and jobs in the most diverse segments (Coeckelbergh, 2022), including strategic planning, political consulting, the creation of advertising, and digital marketing for electoral campaigns, thus leading to an environment in which more and more responsibilities end up being delegated to algorithms.

AI and Electoral Campaigns, First Edition. Rafael Rubio Núñez, Frederico Franco Alvim and Vitor de Andrade Monteiro.
© 2026 John Wiley & Sons, Inc. Published 2026 by John Wiley & Sons, Inc.

Reagan's first election campaign in 1980 onwards gave rise to a series of phenomena in the political arena, such as web pages (2000), meetups (2004), social networks (2008), micro-segmentation (2012), and data-driven campaigns in 2016, whose use rapidly became the norm.

However, the advance of artificial intelligence (AI) heralds the start of a new and inevitable transformation in the dynamics of electoral contests. In this context, the modernization and consequent expansion of communication possibilities, from the production to the dissemination phase, makes it possible to collect data on voter behavior and use this to create personalized messages as well as facilitate the dissemination of automated messages with an enormous potential to artificially manipulate public debate. In addition, the use of AI has the capacity to exert a negative influence both on the information market and on the climate of social relations, even going so far as to create apparently real "facts" embedded in ultrarealistic representations of people and events, which are conceived with a degree of sophistication that increasingly surpasses the *savoir-faire* of the average hoaxer,[2] and which use persuasion techniques that were "unthinkable in the past" and damage our collective ability to understand reality (Schick, 2020, p. 30).

Furthermore, these technologies provide political parties and candidates with opportunities to gain a better understanding of reality, develop sophisticated strategies, and optimize their communication with voters. On the other hand, they also enable an increase in the institutional capacities of electoral bodies, for example, in the fields of improving relations with society, the logistical organization of processes, the updating of electoral registers, voter identification, the detection of disinformation through social listening and network monitoring, and the construction of communication strategies to disseminate content, using "counter-discourse" as a "combat tactic" (Prado, 2022, p. 51).

The undeniable role of AI in the organization of current electoral processes raises concerns (Venice Commission, 2022; Montilla

[2]The fake video in which Ukrainian President Zelensky encourages his compatriots to surrender to Russia is a significant example of how, in the current scenario, "poorly qualified actors" are in a position to influence critical issues, including military conflicts, thus "complicating the information landscape" (Gómez de Ágreda, 2023, p. 207).

Martos, 2023; Robins-Early, 2023; World Economic Forum, 2024),[3] as the recent experience of elections in countries such as Slovakia, Pakistan, India, Mexico, the United States, and Argentina[4] has shown. In general, these cases were characterized by a recurrent use of generative tools in the production of advertising but also by the employment of these techniques in the industrial-scale production of high-yield content of a harmful or disinformative nature, such as the audio and video deepfakes that have flooded social media platforms and messaging applications (Meaker, 2023; Ramírez, 2023) or memes that use humor as a pretext for the manufacture of certain stereotypes on digital networks (BURROUGHS, 2023, p. 199). Such content is distributed on a massive scale through automated sending or the creation of fake profiles, in many cases as a result of the selective programming of different types of bots.[5]

Without succumbing to a catastrophism that could condition our response, it is clear that AI poses technical and social challenges that, in turn, demand the adoption of specific precautions and attitudes on the part of legislative (Garriga; Ruiz-Incertis; Magallón Rosa, 2024; Van der Linden, 2024), administrative, and jurisdictional bodies that

[3] Although disinformation had negative social effects even prior to the assimilation of new AI tools, the fact is that AI has elevated this problem to a new dimension. These new technologies promote quantitative and qualitative changes in the generation of fake news, since they streamline, cheapen, and facilitate the circular manufacture of disinformation as well as allowing more sophisticated forms of deception and manipulation, such as deepfakes. In addition, AI enables the large-scale micro-segmentation of communication (The Economist, 2023), creating echo chambers and intensifying sectarianism, ideological intolerance, and other phenomena that tend to foster extremism and ignite the "powder keg" (Sunstein, 2019, p. 62) of the "culture of anger and conspiracy" (Fisher, 2022, p. 76).

[4] Audio and video deepfakes have also been reported in the United Kingdom, Nigeria, Ethiopia, Sudan (Bioni; Almeida; Mendes, 2024; Zahra, 2024), Bangladesh (Muller, 2023), Brazil (Alvim; Monteiro, Rubio Núñez, 2025), Taiwan, Poland, and Bulgaria (Labuz; Nehring, 2024), as well as several other countries.

[5] To summarize, public persuasion campaigns have generally used four categories of robots: (a) social bots, programmed to behave like human users on social networks; (b) political bots, specially programmed to support politicians, electoral campaigns, or specific agendas; (c) chatbots, developed to interact with users and provide a service, clarification, guidance, or exchange information; and (d) spam bots, dedicated to the automated diffusion of specific content (Denemark, 2024, p. 127).

participate in the organization of elections in the broadest possible sense. This need is further confirmed by the decisive role played by the opaque and biased algorithms that power recommendation systems (Fisher, 2022, p. 117; Morozov, 2018, p. 39; Santini; Salles; Mattos, 2023, p. 7); fuel identity bubbles; and favor the creation and dissemination of negative campaigns (Mihailidis, 2023), hate speech (Guaraty, 2023, p. 09), fallacy mechanisms (Carratalá; Iranzo-Cabrera; López-García, 2023, p. 16), and practices designed to alter the popular will (Ramonet, 2022, p. 33; Pérez-Curiel; Rivas-de-Roca; García-Gordillo, 2023, p. 38). Moreover, the emergence of a polarized and radicalized public sphere (Prado, 2023, p. 96) associated with the banalization of authoritarianism (Levitsky; Ziblatt, 2023, p. 56) ushers in an era of "high-risk elections" (De Leo, 2023),[6] which is further compounded by the overexploitation of personal information for the purposes of blackmail, influence, agitation, and propaganda (Luquin Calvo, 2023, p. 57; Rebollo Delgado, 2023, p. 28) in addition to the absence of effective and comprehensive special regulation.

Given these circumstances, this work aims to discuss the impact of the increasing role of AI tools in political–electoral communication from a political–legal perspective, in particular with regards to the supposed "special protection" (Sánchez Muñoz, 2020, p. 133) warranted by the integrity of the electoral system in the face of the new pathologies present in the era of digitalized elections. It is not a question of assessing the advisability or otherwise of using these technologies (given that the horse has already bolted the stable) but of carrying out a risk analysis as a preventive tool which identifies threats that have the potential to affect the execution of a process,

[6] In countries such as the United Kingdom, the United States, Peru, Mexico, Paraguay, and Brazil, public trust has recently been weakened as a result of disinformation affecting electoral organizations, which creates fertile ground for post-electoral outbursts denouncing the "rigging" of elections and seeking to reverse the results. Such narratives are fed by lies, truths taken out of context, and exaggerated falsehoods which, capitalizing on the complexity of these processes, trigger a collective reaction of doubt-suspicion-indignation-and-revolt capable of undermining democratic systems. In this context, AI is a corrosive ingredient in a formula that is already quite toxic and can therefore be classified as an additional risk of great importance.

in this case the electoral process, and which allows for a more effective management of contingencies, should these occur. In this sense, the authors propose a discussion centered on the risk paradigm, that is, on the contingencies or negative externalities that might result from the malicious use of AI tools but with the hope that the general sense of alert will not be interpreted as a sign of alarmist determinism. The risks that will be analyzed have different dimensions and impacts, many of which will hopefully fail to materialize. Others, moreover, can be neutralized by the pertinent institutions, provided that these are alert to the array of positive uses of the very tools that will be discussed.

While not claiming to offer simplistic answers that are more damaging to the democratic system than the legal good that these seek to defend, the knowledge produced by the capabilities and effects of using these technologies during elections renders it essential to consider possible answers. This is especially relevant given the context in which politics is above all a battle fought via communication, political disputes are driven by data (Juneja; Mcbride, 2023), and in which social media platforms act both as authentic "key institutions" (Balkin, 2021, p. 71) of the "new public square" (Innerarity; Colomina, 2020, p. 12) and as the new controllers of these communications processes[7] (Balaguer Callejón, 2023, p. 25; Bucci, 2023, p. 113; Kissinger; Schmidt; Huttenlocher, 2022, p. 90), potentially leading to an almost unconscious shift in views on the part of voters (Helbin *et al*., 2017; Ortega, 2023), thereby acting as veritable "weapons" in the struggle for collective persuasion (Van der Linden, 2023, p. 133).

[7] In an even more incisive view, a report by the Stigler Center (2020) recognizes that the large digital platforms, especially Google and Facebook, "may be the most powerful political agents of our time," since they combine characteristics that allow the capture of politicians and prevent effective democratic oversight: (a) money, since their immense economic power allows them to effectively lobby politicians and regulators; (b) media, which allows them to shape public discourse and define how politicians can reach their constituents; (c) complexity, since their size and opacity prevent the development of effective regulatory tools and platforms can use informational asymmetries to evade regulations; and (d) connectivity, since these agents can use their user base to challenge any political initiative that disadvantages them (Frazão, 2022, p. 566).

BIBLIOGRAPHY

Alvim, Frederico Franco; Monteiro, Vitor de Andrade; Rubio Núñez, Rafael. *Inteligência artificial para eleições (mais) livres, justas e eficientes*. Rio de Janeiro: Lumen Juris, 2025.

Balaguer Callejón, Francisco. *La constitución del algoritmo*. Madrid: Fundación Manuel Giménez Abad, 2023.

Balkin, Jack M. How to regulate (and not regulate) Social Media. *Journal of Free Speach Law*, 71 (2021), p. 71–96.

Bioni, Bruno; Garrote, Marina; Guedes, Paula. Temas centrais na regulação de IA: o local, o regional e o global na busca da interoperabilidade regulatória. São Paulo: Associação Data Privacy Brasil de Pesquisa, 2023.

Bioni, Bruno; Almeida, Virgilio; Mendes, Laura Schertel. Inteligência artificial e ameaça a integridade das eleições. *Folha de São Paulo*, 17 de febrero de 2024. Available at: [https://www1.folha.uol.com.br/ilustrissima/2024/02/inteligencia-artificial-e-ameaca-a-integridade-de-eleicoes.shtml].Viewed: 19-02-2024.

Bucci, Eugênio. *Incerteza, um ensaio. Como pensamos a ideia que nos desorienta (e orienta o mundo digital)*. Belo Horizonte: Autêntica, 2023.

Burroughs, Benjamin. Fake memetics: popular rhetoric and circulation in political campaigns. In: Zimdars, Melissa; McLeod, Kembrew (eds.). *Fake news: understanding media and misinformation in the Digital Age*. Cambridge: The MIT Press, 2023, p. 191–200.

Carratalá, Adolfo; Iranzo-Cabrera, María; López-García, Guillermo. *De la desinformación a la conspiración: política y comunicación ante un escenario híbrido*. Valencia: Tirant lo Blanch, 2023.

Coeckelbergh, Mark. *The Political Philosophy of AI: An Introduction*. Cambridge: Polity, 2022.

Venice Commission. European Conferences of Electoral Management Bodies Concept Paper 2022: Artificial intelligence and electoral integrity. **Council of Europe**, 2022. Available at: [https://www.coe.int/en/web/electoral-management-bodies-conference/concept-paper-2022]. Viewed: 18.12.2023.

de Leo, Juan Pablo. La inteligencia artificial debuta en el proceso electoral mexicano. **El Economista,** 12 July 2023. Available at: [https://www.eleconomista.com.mx/opinion/La-Inteligencia-Artificial-debuta-en-el-proceso-electoral-mexicano-20230712-0077.html]. Viewed: 18.12.2023.

Denemark, Jaroslav. The risk of artificial intelligence for democracy and the EU's first efforts tu regulate it. *The Lawyer Quarterly*, 14(1) (2024). Available at: [The Lawyer Quarterly (cas.cz)]. Viewed: 22.03.2024

Elliot, Victoria; Kelly, Makena. The Biden deepfake robocall is only the beginning. **Wired**, 24 January 2024. Available at: [https://www.wired.com/story/biden-robocall-deepfake-danger/]. Viewed: 24.01.2024.

Fisher, Max. *The inside story of how social media rewired our minds and our world*. New York: Little Brown & Co., 2022.

Frazão, Ana. A democracia na era digital: os riscos da política movida a dados. In: Branco, Paulo Gustavo Gonet; Branco, Paulo Gustavo Gonet; Fonseca, Reynaldo Soares da; Branco, Pedro Henrique de Moura Gonet; Velloso, João Carlos Banhos; Fonseca, Gabriel Campos Soares da (eds.). *Eleições e democracia na era digital*. Brasília: Almedina, 2022, p. 69–84.

Garriga, Miriam; Ruiz-Incertis, Raquel; Magallón-Rosa, Raúl. Inteligencia artificial, desinformación y propuestas de alfabetización en torno a los deepfakes. *OBS Journal* (2024, Special Issue), p. 175–194. Available at: [https://obs.obercom.pt/index.php/obs/article/view/2445/188188321]. Viewed: 14.02.2024

Gómez De Ágreda, Ángel. La paz es la víctima última de la mentira. Desinformación con base tecnológica en la guerra. In: Vázquez-Barrio, Tamara; Salazar García, Idoia (eds.). *Inteligencia artificial, periodismo y democracia*. Valencia: Tirant lo Blanch, 2023, p. 205–226.

Guaraty, Kaleo Dornaika. *Discurso de ódio no direito eleitoral*. Rio de Janeiro: Lumen Juris, 2023.

Helbin, Dirk; Frey, Bruno; Gigerenzer, Gerd; Hafen, Ernst; Hagner, Michael; Hofstetter, Yvonne; Van den Hoven, Jeroen; Zicari, Roberto; Zwitter, Andrej. Will Democracy survive Big Data and Artificial Intelligence? Essays on the dark and light sides of the digital revolution. **Scientific American,** 25 February 2017. Available at: [https://www.scientificamerican.com/article/will-democracy-survive-big-data-and-artificial-intelligence/]. Viewed: 23.02.2024.

Innerarity, Daniel; Colomina, Carme. La verdad en las democracias algorítmicas. *Revista CIDOB d'Afers Internacionals*, 124 (2020), p. 11–23.

Juneja, Prathm; McBride, Keegan. *How sata and artificial intelligence are actually transforming American elections.* **Oxford Internet Institute**, 15 December 2023. Available at: [https://www.oii.ox.ac.uk/news-events/how-data-and-artificial-intelligence-are-actually-transforming-american-elections/]. Viewed: 18.12.2023.

Kissinger, Henry A; Schmidt, Eric; Huttenlocher, Daniel. *The age of AI: and our human future*. New York: Little, Brown and Company, 2021.

Labuz, Mateusz; Nehring, Christopher. On the way to deep fake democracy? Deep fakes in election campaigns in 2023. *European Political Science*, 23 (2024), p. 454–473.

Levitsky, Steven; Ziblatt, Daniel. *Tyranny of the minority: how to reverse an authoritarian turn, and forge a democracy for all*. New York City: Viking, 2023.

Luquin Calvo, Andrea. Hannah Arendt y las teorías de la conspiración en la era de las redes sociales: régimen de verdad y tentación totalitaria. In: Carratalá, Adolfo; Iranzo-Cabrera, María; López-García, Guillermo (eds.). *De la desinformación a la conspiración: política y comunicación ante un escenario híbrido*. Valencia: Tirant lo Blanch, 2023, p. 47–61.

Meaker, Morgan. Slovakia's election deepfakes show AI is a danger to democracy. **Wired,** 3 October 2023. Available at: [https://www.wired.co.uk/article/slovakia-election-deepfakes]. Viewed: 18.12.2023.

Mihailidis, Paul. Normalizing fake news in an Age of Platforms. In: Zimdars, Melissa; McLeod, Kembrew (eds.). *Fake news: understanding media and misinformation in the Digital Age*. Cambridge: The MIT Press, 2023, p. 341–350.

Montilla Martos, José Antonio. Inteligencia artificial y derechos de participación política. In: Balaguer Callejón, Francisco; Cotino Hueso, Lorenzo (eds.). (Coords.)*Derecho Público de la Inteligencia Artificial*. Zaragoza: Fund. Miguel Giménez Abad, 2023, p. 151–180.

Morozov, Evgeny. *Capitalismo Big Tech: ¿Welfare o neofeudalismo digital?* Madrid: Enclave de Libros, 2018.

Muller, Derek. *Deepfakes* for $24 a month: how AI is disrupting Bangladesh's election. *Financial Times*, 14 de diciembre de 2023. Available at: [https://electionlawblog.org/?p=140195]. Viewed: 27-02-2024.

Ortega, Andrés. Cómo la nueva Inteligencia Artificial puede manipularte como votante. **El País,** 17 September 2023. Available at: [https://elpais.com/ideas/2023-09-17/como-la-nueva-inteligencia-artificial-puede-manipularte-como-votante.html]. Viewed: 18.12.2023.

Pérez-Curiel, Concha; Rivas-De-Roca, Rubén; García-Gordillo, Mar. Narrativas populistas con alcance global: el caso de la retórica de Trump en las elecciones de Estados Unidos de 2020. In: Carratalá, Adolfo; Iranzo-Cabrera, María; López-García, Guillermo (eds.). *De la desinformación a la conspiración: política y comunicación ante un escenario híbrido*. Valencia: Tirant lo Blanch, 2023, p. 27–45.

Prado, Michele. *Fake news e inteligência artificial: o poder dos algoritmos na guerra da desinformação*. Vol. 70. São Paulo: Edições, 2022.

Prado, Michele. *Tempestade ideológica. Bolsonarismo: a altright e o populismo iliberal no Brasil*. São Paulo: Todos Livros, 2023.

Ramírez, León. Audios creados artificialmente siembran incertidumbre en las elecciones de México. **Associated Press,** 30 November 2023. Available at: [https://apnews.com/world-news/general-news-fe47701d72cc861754be9ad3161ef5be]. Viewed: 18.12.2023.

Ramonet, Ignacio. *La era del conspiracionismo. Trump, el culto a la mentira y el asalto al Capitolio*. Buenos Aires: Siglo XXI, 2022.

Rebollo Delgado, Lucrecio. *Inteligencia artificial y derechos fundamentales*. Madrid: Dykinson, 2023.

Robins-Early, Nick. Disinformation reimagined: how AI could erode democracy in the 2024 US elections. **The Guardian,** 19 July 2023. Available at: [https://www.theguardian.com/us-news/2023/jul/19/ai-generated-disinformation-us-elections]. Viewed: 18.12.2023.

Sánchez Muñoz, Óscar. *La regulación de las campañas electorales en la era digital. Desinformación y microsegmentación en las redes sociales con fines electorales*. Madrid: Centro de Estudios Políticos y Constitucionales, 2020.

Santini, Rose; Salles, Patrícia; Mattos, Marcelo. *Inteligência artificial e eleições: Desafios éticos e regulatórios*. Rio de Janeiro: Fiocruz, 2023.

Schick, Nina. *Deep fakes and the infocalypse: what you urgently need to know*. New York: Monoray, 2020.

Seitz-Wald, Alex; Memoli, Mike. Fake Joe Biden robocall tells New Hampshire Democrats not to vote Tuesday. **NBC News,** 24 January 2024. Available at: [https://www.nbcnews.com/politics/2024-election/fake-joe-biden-robocall-tells-new-hampshire-democrats-not-vote-tuesday-rcna134984]. Viewed: 24.01.2024.

Sunstein, Cass R. *Conformity: the power of social influences*. New York: NYU Press, 2019.

The Economist. How worried should you be about AI disrupting elections?. 31 August 2023. Available at: [https://www.economist.com/leaders/2023/08/31/how-artificial-intelligence-will-affect-the-elections-of-2024]. Viewed: 18.12.2023.

van der Linden, Sander. *Foolproof: Why misinformation infects our minds and how to build immunity*. New York: W.W. Norton & Company, 2023.

van der Linden, Sander. Las *fake news* generadas por IA te acecharán en las próximas elecciones. **Wired,** 23 January 2024. Available at: [https://es.wired.com/articulos/fake-news-generadas-por-ia-te-acecharan-en-proximas-elecciones]. Viewed: 24.01.2024.

World Economic Forum. The global risks report 2024, 19th Edition. 10 January 2024. Available at: [https://www.weforum.org/publications/global-risks-report-2024/]. Viewed: 11.06.2025.

Zahra, Tasneem. India faces AI-disinformation threat ahead of elections. **Countercurrents,** 25 February 2024. Available at: [https://countercurrents.org/2024/02/india-faces-ai-disinformation-threat-ahead-of-elections/]. Viewed: 25.02.2024.

CHAPTER 1

High-Risk Elections in a New Sociopolitical Environment

As the amount of data available and the automation of its processing increase, and as both are employed in political campaigns to win elections, there is a growing demand for specialized technical knowledge (Juneja and McBride, 2023), whether on the part of parties and candidates, legal, and marketing teams or the government bodies tasked with overseeing elections.

In 1974, in a publication that has achieved cult status for many, Nelson and Brand (1974) warned that any concerns about democracy must be accompanied by an understanding of technology. Today, the incorporation of these tools (and their uses and applications) does not necessitate an in-depth

AI and Electoral Campaigns, First Edition. Rafael Rubio Núñez, Frederico Franco Alvim and Vitor de Andrade Monteiro.
© 2026 John Wiley & Sons, Inc. Published 2026 by John Wiley & Sons, Inc.

knowledge of machine learning,[1] deep learning,[2] neural networks,[3] generative adversarial networks (GAN), or natural language processing (NLP).[4] Nonetheless, it does require a broad understanding of the evolution of political communication and election campaigns as well as of artificial intelligence's (AI) impact, both on democratic processes in general and on the ways in which voting decisions are made in particular. Bearing in mind that in competitive contexts "knowledge of AI is strategic" (Peixoto, 2020, p. 9), this chapter is centered on these issues.

[1] "Machine learning refers to programs that can 'learn'. The term is controversial: some claim that what is produced is not true learning, because there is no real cognition; only human beings can learn. In any case, modern machine learning has 'little or no resemblance to what might happen in the human mind'. Machine learning is [...] a statistical process. It can be used for a variety of tasks, but the underlying task is usually pattern recognition. Algorithms can identify patterns or rules in data and use them to explain the data and make predictions about future data. [...] Unlike expert systems, which rely on human experts in the field to explain the rules to programmers, who then code them, the machine-learning algorithm finds rules or patterns that have not been specified by the programmer. Only the task or objective is provided. The program can adapt its behavior to better fit the requirements of the task" (Coeckelbergh, 2022, p. 81).

[2] Defined as a "subfield" of machine learning, deep learning is a cutting-edge technology present in systems that process "huge amounts of data to find relationships and patterns that humans are often unable to detect." The term "deep" in this context refers to "the number of hidden layers in a neural network, which provide much of the learning power" (Taulli, 2020). Deep learning technology, therefore, is "technology that uses neural network algorithms, deepening the processing in layers of artificial neurons to solve more complex problems, coming closer to what we understand by human 'thinking'" (Gabriel, 2022).

[3] "Neural networks are a type of machine learning that allows computers to learn to perform tasks by analyzing training examples. In general, these examples are pre-labeled. For example, an object recognition system receives thousands of labeled images of objects such as cars, houses and coffee pots. Through analysis, it can identify patterns in the images that correspond to the specific labels. A neural network is designed to resemble the structure of the human brain, with thousands or millions of interconnected processing nodes. These nodes are usually organized in layers and the data flows through them in a single direction, making them 'feedforward'. Each node receives data from the nodes in the lower layer and sends data to the nodes in the upper layer" (UNESCO, 2023, p. 21).

[4] Natural language processing is a "machine learning technique that analyzes large amounts of human text data or speech data (transcribed or acoustic) in search of specific properties such as meaning, content, intention, attitude and context" (UNESCO, 2023, p. 22).

1.1 FROM ANALOG CAMPAIGNS TO TECHNOPOLITICS: THE ERA OF HIGH-RISK ELECTIONS

In an influential essay, Pippa Norris (2004), argues that the transformations that political campaigns have undergone are best understood by examining how they have evolved over time. With this end in view, the British researcher divides the history of election campaigns into three distinct phases – *premodern campaigns*, *modern campaigns,* and *postmodern campaigns*[5] – basing her analysis on the organizational aspects and forms of communication that were prevalent in each period. Today, a fourth phase needs to be added to this model: *algorithmic campaigns*.

1.1.1 Premodern Campaigns

Premodern campaigns took place prior to the massification of print media: the number of newspaper publishers was limited, the majority of the population was illiterate, and the rudimentary nature of transport networks rendered the distribution of the restricted numbers of newspapers difficult. These *premodern elections* were based on patterns of behavior rooted in personal contact between candidates and voters (Norris, 2004), hence they revolved around "propaganda work at an individual or small group level" (Nohlen and Garrido, 2023, p. 362).

In this stage, the information that circulated was of little relevance, given that political choices, when not contaminated by political–financial dependence (patron–client relationships), were determined mainly by

[5] "There was a time when brilliant speeches, whether in parliament or in street rallies, determined the electoral destiny of political leaders. Subsequently, the path to victory or defeat was linked to image in the media, jumping between the press, radio and television (by virtue of 'tribunes', friendly columnists, scripted interviews, televised sofa chats, paid political advertising and audacious confrontations in televised political debates. As a result, image consultants and campaign teams focused all their energies and know-how on the media battle. It was said that the only political reality that mattered was the televised one – or, by extension, the dominant reality emphasized by the media conglomerates [...]. And the teams with the greatest ability to impose their presence and their agenda on this morass were the likely winners of the electoral contest [...]. However, an alternative communications framework has gradually developed following the emergence of the Internet and its multiple outlets. [...] digital campaign activity has ceased to be a section 'far away on the horizon' and has become part of the 'backbone of every campaign'" (Dader, 2017, p. 11–12).

personal sympathy or trust. At this time, the circle of candidates was too narrow and the electorate was limited by property qualifications (census suffrage), educational qualifications (suffrage by ability), and cultural restrictions (the exclusion of women from suffrage). In general terms, from the mid-nineteenth century to the early twentieth century, electoral politics were dominated by eminent personalities who were already well-known to the limited pool of voters prior to entering politics, thereby establishing a "parliamentary democracy" of notables, according to Bernard Manin's formula (Manin, 1997).

In addition, the local and non-centralized nature of the political messages, together with an exclusively written media, led to widespread ignorance when it came to candidates' personalities (depersonalization). This was particularly pronounced in remote locations, where the public imagination was built on the fringes of audio and visual records.

In theory, the severely limited communication possibilities and sociodemographic filter of who could exercise their democratic rights favored the containment of serious and widespread disturbances. Despite the evident shortcomings of electoral integrity – which would later give rise to a deepening institutionalization and the establishment of electoral control bodies – it must be acknowledged that the structure of the information market simultaneously imposed limitations on the incitement to violence, anonymity, and the systematic circulation of misinformation. Furthermore, the systematic exclusion of less privileged classes reduced the general tension in national elections which were, particularly in the nineteenth century, centered more on political fractures (wealthy conservatives versus wealthy liberals) than on historically more conflictive divisions, such as socioeconomic divisions and class conflict (the capital–labor division) (Lipset *et al.*, 1992, p. 189).

In short, in premodern elections (a) dialogic conflict was moderate, reasonable, and circumstantial; (b) emotional communication played a peripheral role in political proselytizing; (c) when they existed, falsehoods and aggressive discourse circulated slowly and with difficulty; (d) violence, although occasionally observed, was isolated and localized, it was neither endemic contamination, nor were there recurrent episodes at a national level; and (e) uprisings and protest movements, when they emerged, were based on legitimate demands for reform of processes, they lacked aspirations to seize power or an authoritarian demeanor and had nothing in common with conspiratorial elements which are, according to various authors, destabilizing today's democracies (Alvim *et al.*, 2023a, p. 44; Luquin Calvo, 2023, p. 47–48; Pérez-Curiel; Rivas-de-Roca; García-Gordillo, 2023, p. 41).

1.1.2 Modern Campaigns

Decades later, the appearance of the new mass media – radio, from the 1920s onwards, and television after the 1940s – revolutionized how the political class interacted with the electorate. Electoral contests were radically transformed in the historical period of "party democracy" typical of mass society (Manin, 1997). The mediated image – which began to be refined with the advent of professional election advisers (Nohlen and Garrido, 2023, p. 362) – would become the center of the campaign.

Modern elections were greatly influenced by television (political ads, news coverage, the broadcasting of debates between candidates), opinion polls, and the popularization of "horse-race journalism"[6] (Broh, 1980). In the first place, this led to a reduction in the organizational axes of election campaigns, due to a greater concentration of decision-making and less dependence on the decentralized action of local committees and activists. This decline in participation in political parties coincided with the personalization of campaigns, a process through which the individual figure of the candidate (particularly in presidential systems) came to outweigh the banners, values, and traditions historically represented by their respective political formations.

On the other hand, the modernization of scientific production propelled the specialization of election campaigns,[7] which were now planned under the guidance of statisticians, political scientists, and professional consultants (Norris, 2004). These experts gradually began to explore the possibilities of advertising success based on multi-factor data, taking into account the power of discourse and the strength of image as strategic elements that can provide a decisive advantage in the struggle for popular support. In this context,

[6] Horse-race coverage is characterized by a predominant interest in predicting the winners, focusing more on voting intentions (through opinion polls) and the likelihood of victory or defeat than on an analytical approach or the critique of manifestos and ideas (Paletz, 1997, p. 216). It has the effect of impoverishing debate and influencing voter behavior toward a behavioral adjustment effect known as the "*bandwagon effect*" (Garrigues Walker; González de la Garza, 2020, p. 91).

[7] In fact, premodern (traditional) campaigns were characterized "by a lack of specialization on the part of political leaders," which was substituted in all cases by pure empiricism. Only with the passage of time did campaigns begin to receive strategies aimed at "identifying and constructing reasons for voting" through the use of "communication and operational models with a high electoral profitability" run by expert professionals (Sánchez Galicia, 2015, p. 58).

"the possibilities derived from the mass media, especially television and radio," made possible the creation of a "common knowledge base, since everyone was able to have equal access to the same facts," which rendered "the enlightened project of a rational democracy, based on an informed electorate, more feasible" (Frazão, 2022, p. 548). In addition, the consolidation of a method to provide narratives based on "sober rationality" (Maher, 2022, p. 3) contributed, in general, to raising the bar for the formation of ideas and opinions.

At this point, the structural obstacles to the circulation of ideas and information were reduced significantly.[8] Communication *from one to many* became possible, but at the same time editorial mediation acted as a generally effective *gatekeeper* that prevented information dysfunction, keeping conspiracy theories, false narratives, extremist behavior, and the "transmission of hatred by chain reaction" (Giorgi, 2019, p. 119) out of the mainstream public sphere. In this context, press vigilance and journalistic filters constituted an institutional barrier to the anti-system discourse which nowadays is responsible for a "whirlwind" of pseudo-scientific ideas (Gamir-Ríos and Tarullo, 2023, p. 218) and campaigns aimed at destabilizing politics and attacking institutions (Carratalá *et al.*, 2023, p. 14).

At the same time, the consolidation of the universal inclusion agenda generated ideological-partisan polarization, exemplified by the clash between workers' parties and groups more representative of business interests. However, class conflict was cushioned by economic growth and increased social prosperity. Together with the ideologically moderating effect which the "theorem of the average voter" (Black, 1948)[9] had on political discourse, these factors promoted elections with a moderate and natural level of friction.

[8] "Social networks and platforms have reduced private space and, consequently, information flows without filters. The multiplication of sources has increased noise and diminished citizens' ability to differentiate the quality of the information they receive. Communication infrastructures are not responsible for the content they release, beyond the purely technological component. News companies have been weakened and are competing for society's attention through tools such as clickbait and free content, thus increasing the noise even further without regarding the public information interest as a criterion" (Manfredi-Sánchez and Gómez-Iniesta, 2023, p. 187).

[9] Developed in the first half of the last century, the theorem defends the thesis that candidates and parties, driven by the need to win over the majority of the electorate, need moderate positions that are close to the center of the political-ideological spectrum, given that under normal conditions, radical and extremist segments represent a minority of the social spectrum.

To summarize, in modern elections: (a) communication *on an equal footing* took place in a controlled context which was resistant to openly dishonest, paranoid, and aggressive rhetoric; (b) the security provided by the welfare state and the post-war economy maintained a safe level of social dissatisfaction; (c) the moderate spirit of the majority of the electorate discouraged radicalized or hostile public posturing to render extremism a risky gamble; (d) the limited amount of news and knowledge as well as the slowness of its production gave rise to an information market that, while large compared to that of premodern campaigns, was one in which fact-checking, critique, and reflection were still possible[10]; and (e) falsehood came with a political price and engendered social disapproval, which is why (at most) they were used in a cautious, disguised, and marginal way. In addition, the centralization of communication in radio and television networks led to heavily personalized campaigns in which the presence of voice, rhetoric, and image gained significant weight in the construction of political sympathies and consequently, in the dynamics of electoral behavior (Ribeiro, 2015, p. 24ff.).

1.1.3 Postmodern Campaigns

In *postmodern elections*, marketing and advertising stars, such as the well-known spin doctors, play an increasingly important role in election campaigns which have by now acquired a "permanent" character beyond the limits of election times (Blumental, 1982). Along with a new wave of technological advances,[11] current processes are also characterized by a high level of personalization (data-driven campaigns) as well as by the predominance

[10] In basic terms, the knowledge of things comes from experience, trust in reliable authorities or the personal ability to carry out logical reasoning (critical thinking). The problem is that all three sources of knowledge (intrinsic, extrinsic, or derived) are currently challenged by technology (Metaxas, 2020, p. 248), given that the incessant flow of "junk-food information" (Fisher, 2022, p. 153) jeopardizes established knowledge, allows the social legitimacy for valid knowledge production to be usurped by phony sources and facilitates the circulation of highly emotional messages that draw the audience into a "sentimental maze" (Del Rey Morató, 2007, p. 60), all to the detriment of an objective and analytical treatment of information.

[11] Schick (2020, p. 11) observes that until the turn of the millennium technological advances followed a slower pace: "Four centuries passed between the invention of the printing press and the development of photography, for example. But in the last three decades, the Internet, the smartphone and social networks have completely transformed the information landscape." According to the author, such rapid changes in a vital segment, for various reasons render society "on the verge of collapse."

of communication that is negative,[12] rebellious, dishonest, emotional, aggressive, and includes a call to action. This also gave rise to the protest vote, which is often cloaked in the paranoid rhetoric of existential threats and fear typical of populist projects, nationalism, and contemporary radicalized identitarianism (Nohlen and Garrido, 2023, p. 341; Sánchez Muñoz, 2020, p. 33; D'Ancona, 2018, p. 42; Alvim and Carvalho, 2021, p. 512; Pérez-Curiel; Rivas-de-Roca; García-Gordillo, 2023, p. 27; Alvim *et al.*, 2023a, p. 179; Canas, 2022, p. 431; Williams, 2024; Eatwell; Goodwin, 2020; Zahra, 2024; Prado, 2024; Strobl, 2022, p. 44).

The journalism industry's expansion has created an explanatory environment of the world that is more robust yet at the same time more fragmented (Norris, 2004). Exposure to an increasingly pluralist diet of information has provoked a decline in established preferences and an increase in swing voters (Manin, 1997), thereby reinforcing the value of campaigns which, due to their decisive role, have become more active and hard-fought in the face of uncertain results.

In their latest incarnation, profoundly transformed by the explosion of the Internet and the digitalization of public space, these data-driven campaigns "break paradigms" (Rais *et al.*, 2018, p. 17), relegating old analog practices to a secondary role (Mateos Crespo, 2023, p. 236) and displacing – in all likelihood permanently – the developmental linchpin of electoral processes.[13]

Communication solely via traditional media is no longer effective as public dialog migrates to the digital stage, inaugurating a form of contact that is more dynamic, expansive, and interactive (Ballesteros *et al.*, 2017, pp. 143–144; CHENG, Lifen, 2017, p. 423). This is thanks to the so-called "platform society"

---

[12] In the digital age, the intensification of the emotional factor in campaigns is partially explained by the rise of political fandom (Erikson, 2008), given that people generally process political events more through self-identification and emotional links than through a rational consideration of their own interests (Abejón *et al.*, 2017, p. 75). In this context, performative behavior, especially that which is aimed at letting popular indignation loose, becomes an effective tool for maintaining and expanding the base of support since many voters feel identified and engage in histrionic behavior to tell elites everything they "deserve" to hear.

[13] "The integration of technology into election campaigns is a milestone that has completely transformed them: they are far removed from what we were used to just a few years ago. The use of the Internet in political activity and, above all, in campaigns, is not just another constituent element: it is the central element that allows the electorate to interact with candidates in a more horizontal manner. The centrality of the Internet to contemporary campaigns is essential in fact; nowadays it is impossible to imagine any campaign plan that might emerge from a party's engine room which does not include an online communication strategy [...]. Information technology is 'the linchpin of any electoral campaign, the medium that enables the articulation, coordination and control of a broad social base, one that is sufficiently informed and motivated to achieve a real impact on the offline world,' for which it is necessary to 'convert the disorganized set of tools available on the web into an authentic machinery for political action'" (Mateos Crespo, 2023, p. 238, with references to Rubio Núñez).

(Morozov, 2018, p. 56), in which the "digital marketplace" allows parties and candidates to impose their own agendas on that of the media (Teruel *et al.*, 2023, p. 118). As Ana Frazão (2022) observes, the Internet age led to new forms of persuasion as a result of strong changes in the communication environment. Previously concentrated in traditional media outlets, the communication flow has become fragmented, pulverized, personalized, and individualized, and the end of a set of barriers has re-signified the role of users, who have transformed from mere consumers into active players in the information game.

In this environment, the terrain of debate widens, becoming a digital space (cyberspace) "in which everyone is able to communicate and be read," and where social networks – which can be used equally for information, support and critique, and for the serial dissemination of insults (Blanco de Morais, 2018, p. 124) – exert a "gravitational pull" (Álvarez, 2023, p. 174) on the toxic influences of the post-truth world.[14] With wide-ranging and undeniable repercussions for democracy, these changes make it increasingly difficult to separate the true from the false and allow public discussions, invaded by a countless number of fake profiles and robots,[15] to degenerate into a "completely distorted public debate, in which it is not at all clear who is participating and whose interests they serve" (Frazão, 2022, p. 550). In similar, but more graphic terms:

> The exchange of opinions is largely shaped by messages that interrupt conversations, for example, through "shares" which, by capturing attention, short-circuit dialog between people. Sometimes these attempt to focus debate on specific issues, sometimes they try to create opinion in a certain direction, and sometimes they try to divert attention, but always with a purpose of exercising political influence. The platforms behind these messages in no sense comply with the logic of interaction between people on the web, which was presented to us as one of the supposed great contributions the new media would make to our social and political coexistence. In fact, one might even question the legality of these tactics, given that they

[14] "Technological and digital development has blurred the boundaries between the media and producers of information, so that the public sphere is continually receiving digital content." This represents "a threat to open societies: facts compete with emotions and reality is confused with desires and opinions. This approach destroys public trust and creates a conflict with other basic political principles: freedom of expression and freedom of information" (Manfredi-Sánchez; Gómez-Iniesta, 2023, p. 189).

[15] These new synthetic actors can be created both by political parties and by state bodies – such as the Russian Internet Research Agency – which, during the 2016 US presidential campaign, set up dozens of groups of bots to destabilize the process. A sample of six of them, taken by Albright, indicates that more than 340 million interactions were generated during this process (Peirano, 2019).

> introduce a pernicious, distorting factor into social communication and do nothing to further the aims of social communication and the exchange of ideas between free people.
>
> (González-Torre, 2020, pp. 64–65)

In the era of *digital elections*: (a) communication *on an equal footing* is no longer subject to filters and begins to emanate from an unlimited number of sources, including anonymous ones and (b) the flow of ideas becomes omnipresent, multiplex, and extremely rapid, given that the mechanisms of production and circulation have become relatively cheap (Schneider 2022, p. 61). This therefore allows, among other things, the degeneration of the press ecosystem and the appearance of a new "media contract" (Magallón Rosa, 2023, p. 69) prompted by the emergence of freelance journalism (widespread journalism on social networks), the hyper-partisan, ultra-sectarian, and "fake journalism" media (Sánchez Muñoz, 2020, p. 41; Fisher, 2022, p. 12; Dourado, 2021, p. 21) which together convert opinions into "facts" and falsehoods into "news," operating not as mediators but as true "political actors" (Iranzo-Cabrera 2023, p. 160) in the digital space. Social networks dominate the information market; the combination of information and communication technology (ICT), big data, and algorithms[16] creates an environment in which automated practices for selecting recipients and bulk mailing political messages dominates (González-Torre, 2020, p. 55), thereby consolidating:

> Broad freedom of access is challenged by a dependence that results from control over the information received, secrecy about the media employed and the surveillance to which network users are subjected, [....] risk factors that allow us to ask ourselves whether new technologies have freed information from domination, or whether the information industry has merely changed hands.
>
> (Sánchez Martínez, 2020, p. 102)

[16] An algorithm is "a series of organized steps [...] that describes the process that must be followed to solve a specific problem." In computer programming, there are basically four main types: "(a) qualitative [algorithms]: use logical or formal sequences to solve problems; (b) quantitative [algorithms]: can solve problems using basic mathematical operations; (c) computational [algorithms]: perform complex mathematical calculations such as equations or complex numerical operations; (d) non-computational [algorithms]: solution requires human intervention, for example, for data entry" (Moreno, 2023, p. 30).

From this problematic horizon there emerges a new set of systemic risks that include, among other disorders, the post-truth, the disinformation epidemic, excessive polarization, foreign influence operations, electoral denialism, the challenging of democracy, the resurgence of authoritarian thought, and the consequent deterioration of the public sphere due to the influence of processes of radicalization and the normalization of extremism (SÁNCHEZ, 2020, p. 34). These are promoted by automated systems which boost the returns on "false, toxic and divisive" content (ÁLVAREZ, 2023, p. 171) due to this type of content's capacity to generate huge amounts of profit and engagement for the platforms and actors involved (DENEMARK, 2024, p. 130).

The progressive degeneration of modern campaigns into "cognitive wars" eliminates "the enlightened demeanor and peaceful antagonism" that have traditionally characterized the atmosphere of elections, which is now diluted by markedly belligerent behavior that goes against the most basic norms of a democracy (Alvim *et al.,* 2023a, p. 68) and create specific threats to the social contract and the stable functioning of the institutions that guarantee it (MARTÍN GUARDADO, 2023, p. 215).

In this time frame, we observe: (a) the predominance of technological devices, which causes a rupture in the press monopoly and the traditional system of filtering information; (b) a mosaic of information produced and distributed by multiple actors[17] (including anonymous and synthetic ones); (c) a scenario of economic crises and unfulfilled democratic promises, responsible for the explosion of the indignation thermometer;[18]

[17] "Through social networks, we can find multitudes of politically committed people who know how to organize volunteers, raise money and who develop effective online communication thanks to technology and the new media that have entered the political sphere (KERBEL, 2009). At the same time, users generate networks in which communication with the interlocutors moves on to a greater interactivity, with information spreading not only dyadically – that is, affecting only the sender and one individual recipient – but also expanding hyperdiadically, to more distant levels, allowing it to influence other circles that are not of the first-level (CHRISTAKIS; FOWLER, 2010)" (BALLESTEROS *et al.*, 2017, p. 143–144).

[18] The combination of the new structure of personal communications and a "social scenario of disenchantment" favors polarization and the "hijacking of public debate by controversial figures" who abuse disinformation tactics (BIOLCATI, 2022, p. 121). In these conditions, the exercise of freedom of opinion becomes a "factor of erosion, frustration and disenchantment due to the successive failures of projects designed critically correct the social conditions sustaining the traditions of a modernity that sought to export a path of universal brotherhood without success" (NÚÑEZ LADEVÉZE, 2023, p. 235).

(d) a crisis in the media that led to the predominance of "sensationalist journalism" and clickbait journalism;[19] (e) an anarchic and chaotic information market,[20] in which ideas and opinions fluctuate with tremendous agility and compete with one another without deeper analysis;[21] (f) a saturated market responsible for the minute-by-minute dissemination of an unabsorbable volume of information; (g) a bad-tempered public sphere dominated by blackmailing or aggressive modes of communication; and (h) a deficient information market,[22] in which the quantitative

[19] González-Torre (2020, pp. 71–72) acknowledges that the traditional media is getting used to technological changes and new models of news consumption, for example by publishing more and more news online. By using resources related, for example, to the abuse of garish headlines and the insistence on controversial framing as well as by investing in segmentation practices and choosing messages according to recipients' ideology, the news media, as occurs with social networks, can also contribute to the expansion of radical and extreme opinions, in conjunction with polarization.

[20] "Digital networks, unlike the telegraph, radio or television, do not require the newspaper to travel kilometers to reach the reader and, in addition, allow readers to communicate through networks. Google stole the delivery trucks and Amazon the newsstands" [...]. Before, it only took a day for a newspaper to have the advantage of publishing a story, but now the Internet, like television channels, immediately transmits any news that a newspaper then echoes [...]. Through the set of networks that make up the Internet, communication 'from one to many' and 'from many to one' is easy and can bypass the newspaper, publisher or broadcaster who used to select, edit and verify the news. Anyone with access to an email, phone or mobile account can not only receive, but also send information. These channels are global, extend far and wide and are coordinated through networks that have no hierarchy, central coordinator or chain of command. Advertisers can now do without newspapers and other intermediaries, reaching customers directly online while collecting data about them at the same time. The traffic is directed by an invisible authority" (Minow, 2022, p. 372).

[21] From this perspective, it is valid to think that the principle of freedom of opinion, to a certain extent, comes into "conflict with the knowledge society," since valid knowledge flows through networks alongside ambiguous statements, both of which are subject to "the judgment of all citizens." In this context, "the web has become the world's encyclopedia, where it is possible to access all kinds of knowledge as well as all kinds of ignorance" (Núñez Ladevéze, 2023, p. 239).

[22] The impoverishment of the circulation of ideas is not only due to the proliferation of false and decontextualized information (misinformation), but also to the exaggerated generalization and abstraction of promises and speeches as well as the recurrent marginalization of issues that have real importance. Nowadays: "The communicative strategies applied in this environment sometimes lead to the discursive loss of substantive political issues (proposals, detailed plans of intervention, reasoned arguments...), which warrants attention and discussion on the part of the whole population. Instead, messages are

growth in supply is not accompanied by an increase in quality. Although it is true that false statements are as old as the practice of discourse itself (Andrejevic, 2020, p. 24), in this new public configuration, the dissemination of falsehoods no longer operates solely with a *top-down* logic but from a *massive top-down* perspective which is practically unstoppable: from the point of view of institutional defense, the assailants are numerous, and they are everywhere. This is compounded by the fact that the use of these techniques often proceeds directly from the very institutions that should be fighting them.

Nonetheless, the decomposition of civic space remains firmly in the "cartography of now" (Kiffer, 2019, p. 30). Authors such as Philip Howard predicted a new mutation in campaigns, realizing (in advance) the inherent value of data (data-driven campaigns). In Howard's opinion, digital interactions leave invaluable traces from a strategic point of view which make it possible to optimize – to the extent possible – forms of interaction with volunteers, donors, and above all, voters. This could produce a sea change in political communication (Howard, 2006), in whose wake "the processes of electoral analysis, persuasion and mobilization in the cybersphere surpass – without relegating to a secondary role – the strategic centrality of traditional political marketing, focused on disputing discursive hegemony and political image in mass-media channels." Within the new digital culture, "the meticulous planning undertaken by innovative teams of specialists, equipped with the most refined advances in data storage and cross-checking and new methods of social research, has driven increasing technological and [information technology] IT improvements," lifting election campaigns to an unprecedented level of sophistication (Dader, 2017, pp. 13 and 15). Thus, social networks are not only being exploited in an active way (through strategies to attract voters and disseminate ideological content that bypass the mainstream media) but also in a passive way, to gain familiarity with circumstances and "take the temperature" of public opinion (Teruel Rodríguez; Palomo, 2023, p. 117).

In short, experts point to the following as the "main components of the radical transformation that computerized campaign management entails:" (a) the availability of a new generation of electoral databases and the supplementary generation of other digitized databases, which can be combined and expanded

launched without concrete policy references, seeking ease of identification. At the same time, there is an effort to adapt the wording to the largest possible range of 'profiles' among the electorate, so that it can be 'individualized' to the point of achieving a mass volume of supporters identified with this generic message, through the greatest possible dispersion of sources of contact" (Abejón *et al.* 2017, p. 76).

on a daily basis; (b) an increase in the number of volunteers and activists as well as the strategic direction of their activities relating to campaign material, canvassing voters, and fundraising; (c) the fusion between online campaigning and interpersonal public communication ("grassroots" activism becomes "netroots" activism); and (d) the radical transformation of the philosophy underlying techniques for matching, analyzing, reinterpreting, and using data to generate a new type of micro-structured knowledge from which techniques for the micro-segmentation of the electorate emerge (DADER, with references to KREISS and ARMSTRONG, 2017, p. 16).

To this list of novelties we should add: (e) the transfer of information filtering from the traditional press to the digital platforms' classification and recommendation algorithms, which today "play a dominant role in the mediation of power" (VELKOVA AND KAUN, 2022, cited by PRADO, 2022, p. 59); (f) the popularization of new generative artificial intelligence (GenAI) applications that make it *possible* to create original content in different media (text, audio, and video), with a minimum degree of complexity, removing the economic, technical, and structural obstacles that previously had a bearing on the production of speeches, statements, jingles, and communication pieces in general; and (g) the use of these same technologies to produce fake content that has a great capacity to imitate the real thing (audio or visual deepfakes[23]) and is more convincing than disinformation of *human origin* (VAN DER LINDEN, 2023). To summarize:

> In the electoral sphere, AI has created opportunities to exert a selective, individualized, automated influence that often goes unnoticed, as demonstrated by the Cambridge Analytica case. The scandals surrounding this episode, and the serious accusations that have been leveled against Facebook, demonstrate that artificial intelligence can be used in democracies to deliberately manipulate voters and distort public discourse through the spread of fake news.[24]

---

[23] "Until now, audio and video content was edited and shared in a particular way. However, with the arrival of generative AI, fake audio, video, images and text can technically be created in a matter of seconds. This technology simplifies phony political material. [...] Through deepfake technology, people's faces can be interchanged with such perfection that it is impossible to distinguish between them." The use of these fake videos generated by AI in Indian elections has been widespread for at least four years (ZAHRA, 2024).

[24] "Without a doubt, the Cambridge Analytica case immediately lays bear some disturbing realities. Firstly, it shows the damage that these practices can do to democracies, these authentic methods of mass intoxication on the web, by corrupting political communication

> It is true that lies can be shared by individuals among themselves; however, in order to reach large numbers of recipients, AI techniques are used through robotic algorithms: bots. These bots are simple robots that are able to interact via social network accounts as if they were real users, exchanging information, following and gaining new followers.[25]
>
> (Lage, 2022, pp. 163–164)

Given the importance of the digital and algorithmic technologies for the purposes of this research, we dedicate the next chapter to discussing it in more depth.

1.2 DIGITAL ELECTIONS: A NEW ERA FOR ELECTION CAMPAIGNS

In short, political communication on social networks is today characterized by (a) the segmentation of recipients, (b) the selection of messages or news that will be sent to each group, and (c) the constant updating of these processes according to the effects and results that are detected. It requires the use of AI from the moment it starts working with a volume of data that only machines can process (González-Torre, 2020, p. 56). In fact, AI is playing

---

and the process of public opinion formation. In addition, it shows the clear inadequacy of self-regulation on the part of tech companies, and the need for state intervention to avoid the dangerous pathologies they are capable of generating in democracies. At this point, we need to consider the problematic nature of large technology companies' control of social networks, and we need to bear in mind that the new means of social interaction do not appear or articulate themselves spontaneously, nor much less with the assumptions of or at the service of a democratic model; they appear already configured by companies in the sector. Without a doubt, the importance of controlling this field is related to avoiding the danger of large networks of political manipulation, which can use disinformation, distraction or intoxication maneuvers, taking advantage of the new digital environment to manipulate public opinion" (González-Torre, 2020, p. 67).

[25] "For example, in the weeks prior to President Trump's announcement about withdrawing from the Paris Agreement, accounts suspected of being bots generated around a quarter of all tweets published about climate change." The researchers measured the influence of bots by analyzing 6.8 million tweets sent by 1.6 million users between May and June 2017. From this set of data, the team examined a random sample of 184 767 users and found that almost 9.5% were probably bots" (Lage, 2022, ibid.).

the leading role in the new evolutionary stage of campaigns[26] and of digital democracy itself,[27] firstly with the universal deployment of algorithms to classify information, personalize searches, and recommend content and connections (Prado, 2022, p. 87; Kissinger; Schmidt; Huttenlocher, 2021, p. 102), and secondly with the new moment marked by the advent of generative technologies. In this sense:

> The arrival of Generative Artificial Intelligence has heralded a before and after in the political sphere [...]. It has opened up new frontiers in the way messages are created and disseminated, enabling the generation of speeches [...] and personalized content for campaigns at a speed and on a scale that are unprecedented. Tools such as ChatGPT[28] have proven to be fundamental in the

[26] "The influence of artificial intelligence on political communication is not a recent phenomenon. As early as 1996, Paletz highlighted concerns about the potential of advanced information technology, including AI, to fuel populism and demagoguery. In subsequent years, research such as that by Tichy (2018) and Barredo-Ibáñez (2021) has highlighted the growing use of AI, especially natural language processing, in election campaigns, and according to Egleton (2020) there is a need for a legal framework to regulate it. Bykov (2020) also emphasized the importance of AI use in political communication. However, it is the advent of generative artificial intelligence in 2023 that marks a significant turning point. López-López (2023) underlines the growing importance of digital political communication and the role of AI in spreading disinformation, while Sannikova (2023) stresses the negative effects of political bots and the importance of transparency in AI systems. These academic contributions reveal that AI, while improving political communication, also poses important ethical and regulatory challenges" (Guadián, 2024).

[27] In an attempt to place the issue of AI in a historical context, García-Orosa (2021, pp. 3–5) discerns four different "waves" in the relationship between democracy and digital technology. The first wave emerges when political actors begin to massify their presence on the Internet. The second wave comes with the consolidation of social networks as marketplaces for exploiting audience attention. The third wave hinges on micro-segmentation and interactions based on the use of mass data, while the fourth wave emerges when AI assumes control, taking center stage.

[28] "ChatGPT is a linguistic artificial intelligence model developed by OpenAI and is based on GPT (Generative Pre-Trained Transformer) architecture. It is capable of generating text in natural language, responding directly to questions asked by its users (OpenAI, 2023). Its operation is based on a machine learning process, in which the model is trained with large amounts of text data to recognize patterns and structures in natural language. Once trained, it can be used to generate new text, answer questions and perform other tasks based on the knowledge acquired during training (OpenAI, 2023). This ability to generate text in natural language makes ChatGPT a potentially powerful tool and raises

> construction of political narratives, with their ability to adapt themselves to different perspectives and communication styles.
>
> However, this revolution doesn't come without its challenges, such as the need to closely monitor the accuracy and bias of the content generated. The emergence of specialized roles such as "Spin Prompter" underlines the importance of careful and strategic interaction with these tools, ensuring that AI not only amplifies political messages, but also respects the sensitivity and ethics inherent in political communication.
>
> In summary, GenAI has not only transformed the efficacy and scope of political communication, but has also raised new ethical and strategic considerations in this field.[29]
>
> (Guadián, 2024)

important concerns in the democratic context. It is not outside the realms of the possible to think that it might be distorted and exploited by developers to manipulate public opinion, reinforce echo chambers or even create and spread disinformation. In light of debates on algorithms [...], analyzing the potential impact of ChatGPT on elections has therefore become imperative. Careful scrutiny is needed to understand how these models can be used or abused. This includes considering the ethical, legal and constitutional implications, as well as the need to develop standards and regulations to ensure that these technologies are used responsibly and transparently. The appearance of ChatGPT and similar technology reminds us that the digital landscape is constantly evolving and that legal and constitutional responses must be equally agile and solid. As has been pointed out [...], the interaction between technology and democracy is complex and multifaceted, and requires a holistic approach that balances innovation, freedom of expression and the integrity of the democratic process" (Lins, 2023, pp. 296–297).

[29] "[G]enerative neural networks using text or images. They then produce new, synthetic but realistic texts or images. For example: a standard neural network can identify an image of a human face, but a *generative* network can create an image of a human face that *looks* real. [...] The applications of these so-called generators are astounding. If they were successfully applied to coding or writing, an author of a book could simply create an outline and let the generator fill in the details. Or a publicist or filmmaker could provide a generator with a few images or a graphic script, and then let the AI create an advertisement or television commercial. More worryingly, generators could also be used to create deep fakes: false representations, indistinguishable from reality, that do or say things that have never been done or said. Generators will enrich our information space, but without controls, they are also likely to blur the line between reality and fantasy" (Kissinger; Schmidt; Huttenlocher, 2021, pp. 70–71).

In this "new world," votes are obtained using various technologies based on the intersection of new media with big data, machine learning, and insights from behavioral sciences which "make it possible to establish profiles, select targets and direct mass messages to influence or even surreptitiously manipulate the mental architectures of elections" in favor of certain political projects (González-Torre, 2020, p. 56).

Amidst a tangled web of new threats (Eddy, 2024), an accelerated process of the *dehumanization of elections* is taking place,[30] as evidenced by the pronounced breakdown of privacy, the overexploitation of attractive narratives, the normalization of intolerance, and the ostentatious lack of respect that anti-politics candidates show toward institutions as well as by the increasing use of data mining,[31] algorithmic power, and digital propaganda in strategies of mass persuasion that are often based on hatred, social disenchantment, the exploitation of anxieties and insecurities (Giorgi, 2019, p. 85; Pérez-Curiel; Rivas-de-Roca; García-Gordillo, 2023, p. 28; Morozov, 2018, p. 33), and structural receptivity to argumentative fallacies, falsehood, and deceit, in line with the dictatorship of beliefs or the logical justification of one's own preferences.

As a consequence, the environment in which the elections are held bears little resemblance to that in which election campaigns took place up until now. In order to capture its basic essence, Daniel Innerarity (2018) uses the term the "uberization of truth," which would facilitate the direct exchange of information in a disintermediated process; however, it would be more appropriate to speak, following the true logic of Uber, of new processes of complex intermediation in which the most important feature of communication

[30] The process of dehumanization, in short, is revealed in two ways: on the one hand, by the increasing automation of content production and distribution activities and forms of interaction with voters; on the other, by the increasingly incisive presence of segregation and exclusion dynamics, driven by the profitable nature of the politics of creating enemies. Within the second aspect, even "pseudo-speculation" practices are found that involve the discriminatory pursuit of opponents presented as less than human and subject to extermination (Aguado Terrón; Villaplana Jiménez, 2023, p. 208).

[31] "The collection of large amounts of data makes it possible to develop intelligence on the state of opinion, sentiment and biases of future target audiences. This capability results from the unprecedented development of Internet of Things (IoT) sensors capable of collecting sensitive data, the exponential growth in the data processing capacity of these devices, the design of advanced algorithms for processing this data and new processes that encourage the integration of all of the aforementioned and its sharing. This panorama requires the addition of security so as not to be exposed to exploitation by potential adversaries" (Gómez de Ágreda, 2023, p. 220).

processes resides in their character: deterministic, individualized, and above all, as opaque as they are inevitable.

Due to the "illusion of knowledge" (Sloman and Fernbach, 2017), people participate in political processes highly convinced of what they believe they know but blind to their own biases and to the role of algorithms in filtering the real world. As Charles Baudelaire (1864) said – later popularized by C.S. Lewis (*The Screwtape Letters*) and Kevin Spacey (*The Usual Suspects*) – "the greatest trick the Devil ever pulled was convincing the world he didn't exist." Without a doubt, this is currently algorithmic persuasion's greatest strength, thus rendering individuals more dependent than ever on groupthink.

Election campaigns, in this context, take place in a space in which information and opinion merge to the benefit of the latter, emotions shape an increasingly superficial public conversation[32] that is more anecdotal, polemical, truculent, and strident: a dialogical setting in which the excess of information available makes it possible to choose "tailor-made truths" (Luquin Calvo, 2023, p. 57) that fuel the association of like-minded users, who then become allies (Schmitt-Beck 2003). As a result, political discussions now take place within a sociopolitical framework in which "mass individualism" (Baricco, 2019) attracts the lonely and displaced (Beiguelman, 2021, p. 9), the people who are "bowling alone" (Putnam, 2000).

[32] Addressing some of the problems arising from new communication technology, González-Torre (2020) writes that: "We find ourselves [...] with the fact that now, short text, the limitation of characters, the binary response: yes/no, I like it/ I don't like it, which is so frequent on social networks, implies a loss of discursiveness, of the confrontation of ideas, of debate, of the need for argumentation. This produces a terribly impoverishing simplification of political discourse, which, by adopting this form, abandons its original nature as a rational process of seeking the best proposed solutions. From there we can find ourselves facing the end of the modern concept of rational political discourse, of public opinion, of a public space for debate, of an ideal community for dialog; in short, of rational public opinion" (González-Torre 2020, p. 69). In the same vein, the filtering process in particular also contributes to the impoverishment of debate, as Olga Sánchez Martínez (2020, p. 100) points out: "If, by means of information filtering, we diminish the citizen's capacity to know, they will be unable to achieve an expansion of their power. It has even been pointed out that with regard to politics, some complex and unpleasant social issues will disappear from the personalized information that reaches us and will end up erased from social concerns and the political agenda. If we are only informed about what we are interested in, and the information is increasingly individualized, we could venture that we will be 'single issue voters.'"

Today's symbolic disputes depend less on access to information than on the ability to effectively disseminate and reiterate it in endless cycles. If Revel (1983) warned that, in the world of TV, the abundance of free information has more to do with the ease of observing knowledge, today we can say that it has more to do with the possibility of telling something and being listened to. Any event or personal opinion acquires, from narcissistic subjectivity, a disproportionate relevance, measured and amplified by rousing and receiving support, in a "contaminated and dangerous" environment (SCHICK, 2020, p. 09) marked by "impulsivity" when sharing information and "thoughtlessness" when consuming it (SÁNCHEZ, 2020, p. 31).

In this new communicative order, shaped to a large degree by the effects of AI, information travels in a directed manner according to a form of distribution that is predominantly dictated by the inscrutable will of algorithms,[33] which ultimately are what actively determines access to information. At the same time, the segregation of users into fragmented relational niches gives rise to the *cyber-balkanization* of preferences and, as a result of a dialogue fueled by personalized points of view and perspectives that reiterate one's own, the consolidation of self-referential micro-communities that construct their own "storehouses of truth" which, despite their questionable nature, are respected and taken into account by their own members.

This contrasts with democracy understood as a regime of opinion, which presupposes and accepts conflicting visions and interpretations and permanent controversy as the common thread of dialogical interactions between political leaders and their supporters. As Urbinati (2014, p. 154) summarizes, "democracy is government by discussion, because it is the government of opinion." For this reason, the value of democracy is considered to be greater

[33] "Applications and algorithms know almost everything that is behind your typing fingers and your moving eyes [...] sliding across every millimeter of the screen. That's why the machine 'guesses', at the same moment that you type 'v', that you're going to write 'vacuum', or 'value', or 'vote'. [...] Our problem is that you have no idea how the machine knows everything about you. Our problem is that on the other side there is a set of complex operations that are very lucrative and inaccessible to you. We don't really know what's going on there. Almost nobody knows. The only people who know anything are those who have the key to the safe where big tech [...] keeps the codes for its algorithms. They are a very small number of people. Our lives have fallen into an asymmetry without parallel. On the other hand, the neuralgic center of the 'digital world', monopolized by big tech, is a source of enormous uncertainty for the majority of humanity. Algorithms, controlled by the tech giants, have already resolved almost all the uncertainties that remained about people's behavior. On the other hand, we look at the conglomerates and we cannot see what they contain. They have opaque walls" (BUCCI, 2023, pp. 53–54).

than the value of truth (Rorty, 1989), even more so in the era of "audience democracy" (Manin, 1997), in which political actors play more with opportunism than with sincerity, in line with the logic of electoral behavior.

In addition, democracy is also a dynamic system that is open to the incorporation of technology that enables improvement to the process of interaction between people and ideas. Therefore, tools that expand the possibilities for participation in public debate and help to improve the conditions in which it takes place are welcome in a democratic setting. However, the risks inherent in the incorporation of this technology, especially that with great transformative power and social impact, demand a cautious approach that justifies the imposition of regulatory limits. This stance, although it may imply a brake on the speed of progress, runs counter to the natural human tendency of wanting to have everything all at once – the advantages without the disadvantages (Innerarity, 2023, p. 16) – and opts, nonetheless, to follow Goethe's advice, when he stated that in the pursuit of great things one must know how to limit oneself.

1.2.1 Algorithmic Communication

However, in the last decade and a half, electoral contenders have used digital technology and big data to turbocharge their campaigns, thus revolutionizing how they communicate, advertise, and mobilize. This movement has included the application of sentiment and opinion analysis techniques[34] to select and adapt issues and approaches by applying tactics that change according to the circumstances and in which coherence and accountability give way to a constant adaptation of discourse and promises to the apparent desires of the electorate (Nohlen and Garrido, 2023, p. 364).

The automation of some of these processes using AI systems transforms the basic dynamics of electoral processes into what is known as "algorithmic communication" (Campos-Domínguez and García-Orosa, 2018, p. 770). AI and more specifically, NLP software, make it possible to carry out social listening on a massive scale and classify users according to the feelings they

[34] Sentiment analysis data mining seeks to measure not only the audience's level of interest in a particular topic but above all their "evaluation tone" (Boden, 2020, p. 92). In this sense, it processes large volumes of information to decode the audience's general judgment on a given topic, indicating, for example, respective levels of support or disapproval, agreement or disagreement, conformity or disagreement, etc., which can also be linear or on the contrary, indicate more or less abrupt changes in direction.

express and the interactions they engage in. In addition, it allows both political parties and other actors to talk to voters using innovative tools such as chatbots and "artificial volunteers" (Tong; Coster, 2023).

Algorithmic communication fosters opinion that is increasingly personalized (Canas, 2022, p. 171), the generation of content that sometimes blurs the distinction between truth and fiction, and the dissemination and redistribution of disinformation narratives and conspiracy theories (Salinas, 2023, p. 334), thus giving rise to the formation of ideological bubbles and echo chambers (Lewandowsky *et al*., 2020). As experts point out, the most critical aspect of the "new algorithmic order" is that social networks inevitably condition access to information, predetermining – in a selective manner and based on unknown criteria – the content that will (or will not) reach user feeds (Fung and Lessig, 2023).[35] This obviously has restrictive effects on voters' "decision-making autonomy" (Lutz *et al.,* 2023, cited by Aguado and Martínez, 2023, p. 279), by reducing the information base that supports their choice (Rebollo, 2023, p. 13).

In addition to manipulating the algorithms of search engines and recommendation systems so that pages or videos with false claims come first in search results (Scheidt, 2019, p. 7), AI enhances the effectiveness of *microtargeting* and psychometric influence models already used in current campaigns, making the hyper-personalization of contacts a reality thanks to their personalized, dynamic, and adaptive nature.

In general terms, AI techniques allow, among other things: (a) the generation, through language models, of messages in different formats that are adapted and free of contrast, given the possibility of implementing direct, private, and automated targeting;[36] (b) the use of automatic learning techniques with trial and error to generate content that, depending on the response, can be made more persuasive; and (c) the use of feedback from the knowledge accumulated from other voters' reactions, through dynamic conversations in which the messages that work are adapted and strengthened, to align

[35] As a result, when the algorithm includes and excludes data, it eliminates what is contrary and uncomfortable in a typical action that attempts to modulate the behavior of the users who form part of the 'army'" (Prado, 2022, p. 82).

[36] It should be noted that the lack of real control not only affects the category of content but also the significance of financing. As Óscar Sánchez (2020, p. 21) points out, this uncertainty increases considerably the risk to the integrity of electoral processes, since it "dangerously opens the door [...] to interference by foreign powers."

themselves with the affects, idiosyncrasies, and preferences (not necessarily political) of each specific recipient, usually through their environment.[37]

The incorporation of these techniques by some parties and candidates, as happens with all dominant campaign innovations, produces an imitative effect that tends to make their use widespread within a short period of time.[38] As a result, the electoral process, in particular, becomes a technological struggle in which the most efficient have the advantage; in practice, this signifies a democratic regression beyond appearances, since, despite the tangible preservation of key elements (speeches, advertisements, votes, etc.), public opinion would end up fragmented, artificially induced, and distorted, thus altering the conditions necessary for substantially conscious, free, and well-informed political choices.

In this vein, the critical analysis of the contribution of new technologies to the democratic agreement extends to its influence on the individual's real capacity to make free and autonomous decisions. Thus:

> [C]oncerns have been centered on the use of the citizen by new technology and AI as a mere harvester of data and, therefore, not as an active participatory subject. From the viewpoint of the democratic system, the explosion of artificial intelligence allows us to visualize the displacement of debates over ideas and agreement in decision-making, due to the algorithm's ability to aggregate, analyze and interpret different variables and individual preferences, which are extracted from data collected and subjected to automated processing, data which in many cases comes from our private lives and is

[37] "The Internet offers advertisers the world's largest laboratory for consumer research and the generation of business opportunities [...] data processing machines are increasingly filtering our data on their own accord, searching for our habits, hopes, fears and desires. With automatic learning, [...] the computer immerses itself in the data and follows a few simple basic instructions. The algorithm finds patterns on its own and then, over time, connects them with the results. In a certain sense, it can be said to learn" (O'Neil, 2016, pp. 95–96).

[38] In fact, technological applications, as efficient and innovative elements, are identified as "competitive advantages" for the various political contenders and are subject to a rapid escalation explained by the need to stand out and differentiate themselves (Mateos Crespo, 2023, p. 236). As happens in an argument, the underuse of AI ends up becoming a danger for those who repudiate the advantages it brings, to the point that its exclusion leads to a loss of competitiveness and a breakdown in weapons parity (Muñoz Vela, 2022, p. 54).

> processed by signals inputted by a computer programmer. The citizen then becomes a mere passive subject of the democratic process: a reference point for decision making, but without taking decisions, a necessary contributor, but not an author or protagonist. This protagonism seems to be reserved for the algorithm, which can become a substitute for human decisions in the political sphere, as if "algorithmic truth" could prevail over democratic decisions.
>
> (Sánchez Martínez, 2020, pp. 107–108)

The European Parliament notes that "the techniques used by [...] anti-democratic actors to disrupt or influence democratic processes are constantly evolving," and that "the use of algorithms, automation and artificial intelligence has increased the reach and efficiency of disinformation and related cyber activities" (Bentzen, 2018). Election speeches are increasingly derived from extractive technology that reveals voters' concerns, preferences, and reservations, according to what AI tools conclude from their monitoring of social media traffic. Moreover, the expansion of AI includes the possibility of creating deepfakes, tools for the abnormal amplification of hate speech and violence, and the malicious use of automated messages generated and disseminated by bots and political chatbots (Serbanescu, 2021, pp. 122–125).

The use of AI by political parties and candidates – to identify trends, test scenarios, predict changes, and map out success – is fundamental to the achievement of their objectives. The new reality, however, raises a series of warnings and concerns relevant to legislators, experts, and authorities from various countries and organizations (Magallón Rosa, 2023, p. 69; Muller, 2023; Murphy, 2024).

The first concern regards *heightened surveillance* and the *systematic invasion of privacy*, which are essential for the use of microtargeting techniques; the second relates to their *personalized nature,* since, thanks to automated analyses, political pledges and persuasive messages – including disinformation – are adapted to each recipient's inclinations. As a result of the combination of both (general surveillance and the personalization of the approach), AI-driven campaigns acquire unprecedented power in terms of their ability to affect human behavior, jeopardizing the equality of opportunity in electoral competition, in addition to the right to sufficient information and the freedom to form one's own conscience.[39]

[39] The mass dissemination of targeted content is a well-known political reality, especially since the uncovering of the Cambridge Analytica scandal, which involved the exploitation the psychological profiles of more than 230 million people in the United States.

In conclusion, the more we know about individuals, the easier it will be to understand their "architecture of choice" in order to guide their actions in a desired direction. The use of AI tools can determine voting freedom through the manipulation of ideas and messages and selectively exposing voters to a flow of election information that invisibilizes dissent, bringing about an "imposed self-isolation" (SÁNCHEZ, 2020, p. 31) that distorts considerably in the assimilation of objective reality.

After all, we experience and interpret the world through information; in this sense, mediatory technology increasingly interferes with the global framework of perception. The ability to configure the architecture of voters' decision-making in an individualized, dynamic, and concealed manner delimits both the set of available options and the way of valuing them. Although it is argued that the vote does not result from a single process, and neither is it a purely rational movement, it may not be necessary to distinguish between contextual influences and the manipulation made possible by AI since the latter is able to incorporate the former into its workings to configure individuals' decisions.

In a broader observation, González-Torre (2020, p. 57) states that "some techniques of commercial origin, aimed at expanding the markets for consumer products, [...] are increasingly used [...] by political actors: parties, political groups, interest groups, pressure groups, semi-public administrative bodies, [and even] subversive groups." Used with great enthusiasm, "these practices are succeeding in influencing social debate" with such frequency that they can now be viewed as the "main protagonists of the political landscape," and as elements responsible for damaging democracy as traditional politics loses credibility, and as a consequence, the doors are opened to demagogues and populist extremists (PÉREZ-CURIEL; RIVAS-DE-ROCA; GARCÍA-GORDILLO, 2023, p. 31). All this leads to the rise of a political model that weakens the constitutional compact, substituting the search for the general will with the imposition of an ultra-partisan will that abandons the notion of a common interest in order to "prioritize the particular interests of the radicals." In this framework the will of the state loses its character as a general will, becoming a "selfish will" which pays attention to only one part of the ideological spectrum and which – on occasion – loses legitimacy by denying dignity to minorities and by "destroying basic consensus and the

However, it is clear that GenAI is in a position to automate this process, reproducing content at scale in a highly convincing way and free of the signs of robotic behavior that could alienate people. This combination, in O'Reilly's opinion, places elections in a "nightmare scenario" (EDDY, 2024).

conditions that foster political agreement," thus adopting "anti-constitutional and exclusionary behavior" (Martín Guardado, 2023, p. 214).

Ultimately, the progressive digitization of campaigns, with the advent of AI, increases the scope for manipulation since new technological resources affect the decision-making process and weaken, in theory, voters' abilities to make free and informed choices. Social control, in this context, acts as an external and illegitimate influence on people's cognitive dimension at a level that is so concealed as to be imperceptible, seeping into the subconscious. In this way, the "mathematized administration of the world" (Lassalle, 2019, p. 20) is established, potentially capable of causing, despite some signs of progress, the "automation of the production of injustice" (Degli-Esposti, 2023, p. 10), the creation of new hazards and the multiplication of existing threats (Muñoz Vela, 2022, p. 21; Coeckelbergh, 2022, pp. 166–167), issues that should be addressed, first and foremost, by protecting the fundamental rights affected.

1.3 ELECTORAL RIGHTS AND ARTIFICIAL INTELLIGENCE

New technologies have a clear impact on the rights related to the formation of the popular will, the constitution of public opinion, and the transmission of popular sovereignty for political and representative decision-making. The right to information, freedom of expression, and the right to vote are undoubtedly affected by sociotechnical advances and the incorporation of their habitual use into our daily lives (Sánchez Martínez, 2020, pp. 81–82).

Although there has been no explicit mention of the issue of elections, within the European Union there is discussion of AI's influence on the right to political participation, and the subject ends up being addressed in contexts that go beyond democratic processes. In fact, in the Own Initiative Opinion 2017C2 88/01, drawn up by the European Economic and Social Committee and entitled "Opinion of the European Economic and Social Committee on 'Artificial intelligence – The consequences of artificial intelligence on the (digital) single market, production, consumption, employment and society,'" among the 11 areas highlighted, the impact of AI on political participation is discussed in relation to: (a) governance and democracy; (b) improving citizen participation in public policy-making; (c) improving transparency in decision-making; (d) risks related to the use of AI systems (intelligent algorithms) that alter how information is accessed and promote social fragmentation; and (e) the possibility of influencing voter behavior (Muller 2017).

In order to fulfill their function of materializing the basic right to political participation, elections require the effective presence of other fundamental

rights, such as freedom of expression and the rights to information, privacy, equality, and free assembly. All the aforementioned have an electoral component, and all of them can be affected by AI, given that, as a result of the action of algorithms, "citizens participate in the democracies in which they live through a prism that is completely determined by an external agent" (Castellanos, 2022, p. 334). In this vein, the European Commission's "White Paper on Artificial Intelligence" states that "data processing, the way applications are designed and the degree of human intervention can affect rights, freedom of expression, the protection of personal data, privacy and political freedom" (European Commission, 2020, 65 final, p. 14).

Digitalization changes the form and traditional content of election campaigns, affecting elements such as equality between the actors, which is an essential component of the right to access politics. The lack of transparency about spending, financing, the messages, and algorithms used in digital advertising, together with the large-scale invasion of privacy and the absence of a journalistic filter to verify the content distributed, lays the foundations for the generalization of disinformation, which in addition exploits the loopholes that still exist in the regulation of electoral processes (Venice Commission, 2019; Venice Commission, 2020).

The decision to support a candidate is the essential content of the right to participation and, for this decision to be informed, an exchange of information in which political actors try to guide voters' decisions is necessary. AI plays a key role in this exchange: its ability to organize and prioritize the information consulted by voters (for example, through search engines or voting recommendation systems) to generate automated content (that appears real) and to distribute this material in a massive, selective, and targeted way is transforming the way in which election campaigns unfold. In this respect, it should be noted that "the processing of personal data has the potential to produce information capable of facilitating 'inference, interpretation, classification, profiling and predictions about individuals and social groups in the basic areas of life,'" including the habits and preferences that form the basis of voting behavior (Fonseca; da Fonseca and De Paula, 2022, p. 534). On this last point, it cannot be denied that:

> [P]ersonal data currently has great political and electoral value, especially since data processing techniques have been used more and more in elections around the world to personalize, target and segment campaign material, in line with what the members of the Study Group on Data Protection and Elections highlighted in their Recommendations Report:

> Political campaigns have made use of new commercial marketing tools and techniques based on personal data, such as: (i) microtargeting and boosting of news, advertisements and paid advertising;[40] (ii) segmentation of target audiences according to specific profiles and well-selected samples; (iii) automated mass messaging. In this sense, personal data has become a valuable asset for personalizing and adapting political communication between candidates and voters according to preferences, habits and opinions.
>
> On the one hand, this context can be very beneficial for the democratic environment, generating greater commitment and proximity between voters and candidates, as well as giving greater opportunities to candidates with fewer resources available. On the other hand, it poses a threat to voters' right to privacy and the protection of personal data, as well as endangering democracy itself and affecting the health of the electoral process.
>
> (Fonseca; da Fonseca and De Paula, 2022, pp. 534–535, with reference to an InternetLab report)

[40] "Psychographic microsegmentation is a product of so-called surveillance capitalism. Online service providers (especially social networks and search engines) tend to work on the basis of free use. Users don't have to pay to use Instagram, Facebook [...] or Google search [...], at least not with official currency. These services make a profit by offering third-party advertising products and services on the platforms they manage. This form of marketing is most effective when the content of each user's interface adapts to their needs, desires, age, social group, gender, etc. Providers [...] have access to an almost infinite amount of data about each user. Google Analytics, a tool for tracking user behavior on electronic channels, is present in the form of cookies on almost every website you visit. Facebook plug-ins are almost the same. The data obtained [...] is used to personalize what users see and interact with. These practices can represent an imminent danger to the processing of personal data and have (more than once) been the subject of scrutiny by the European Court of Justice. Their ultimate aim is to create advertising for each user based on their personality. Psychographic microtargeting is carried out using algorithms, usually AI-driven programs whose job is to collect and process information and create results, which form the basis for personalized content. These algorithms can also be used in political campaigns, where data such as church attendance, type of car and participation in Facebook groups can be relevant to creating an image of the 'possible' voter. According to their profile, the person sees political ads adapted to their personality. It goes without saying that social networks allow this form of segmentation more than any other platform" (Denemark, 2024, p. 129).

Although it is not easy to evaluate the direct effect of these techniques, there is no doubt about as to their influence and their ability to tip the scales, especially in a context of enormous volatility and very close results. Moreover, elections not only serve to decide who will be in power but also to put the functioning of the political system and its acceptance by citizens to test; in this context, information plays an important role, since its quality, if negative, can generate divisions and undermine the legitimacy of the process.

Collecting and analyzing user data related to political orientation would number among the most harmful practices for participatory rights. According to the Spanish Constitutional Court, this seriously affects privacy in especially sensitive matters (STC 76/2019).

While it is already possible to identify a wide range of fundamental rights affected in this process of obtaining and processing personal data, the dangers of this type of activity are heightened when it is used as a tool to interfere in elections since the very existence of the electoral process is jeopardized by the absence of a democratic and plural environment (Montilla Martos, 2023, p. 155). This is achieved mainly through profiling or microtargeting techniques based on psychometric models that take advantage of the capacity of platforms and the information they have about their users to create political profiles of millions of voters in order to send them, at any moment, veiled messages whose form and content can have a subliminal influence on political behavior. As Byung-Chul Han (2022, pp. 35–37) explains:

> Psychometrics, also called "psychography," is a data-based method for obtaining personality profiles. Psychometric profiles make it possible to predict a person's behavior better than a friend or colleague would be able to. With enough data, it is even possible to generate information beyond what we think we know about ourselves. The smartphone is a psychometric recording device that we feed data to, day after day, even every hour. It can be used to accurately determine the personality of its user. The discipline regime only possessed *demographic* information, which allowed it to carry out *biopolitics*. The information regime, on the other hand, has access to *psychographic* information, which it uses for its *psychopolitics*.
>
> Psychometrics is an ideal tool for psychopolitical marketing. *Microtargeting* uses psychometric profiles. Based on voters' psychograms, they receive personalized advertising on social networks. Just like consumer behavior, voter behavior is influenced on a subconscious

> level. Data-based infocracy undermines the democratic process that presupposes autonomy and free will. [...]
>
> In *microtargeting*, voters are not informed of a party's political program, but are manipulated with political advertisements adapted to their psychogram, and not infrequently with fake news. The efficacy of tens of thousands of variants of an electoral advertisement has been tested. These psychometrically optimized dark ads constitute a threat to democracy. Everyone receives a different message, and this fragments the public. Different groups receive different information, which is often self-contradictory. Citizens are no longer aware of important issues of social relevance. They are more or less incapacitated by having been reduced to a manipulable herd of voters who are there to secure power for politicians. Dark ads contribute to social division and polarization and poison the climate of discourse. In addition, they are invisible to the public. They therefore nullify the fundamental principle of democracy: the self-observation of society.

AI also affects the right to information through the creation and programming of bots which, by emulating the activity of human users (Vivas, 2023, p. 356), are enlisted in electoral debates on a massive scale and in a hyperactive manner. Thanks to machine learning, these bots are able to adapt to the environment in which they are participating and improve their effectiveness in achieving objectives such as: (a) imposing agendas; (b) attacking certain positions; (c) generating the illusion of consensus; and (d) intimidation (Woolley, 2017, p. 16; Woolley and Howard, 2016, pp. 488–490).

By conditioning the channels through which information is accessed, AI limits the possibility for individuals to discover new and, above all, different information, which limits their horizons and affects contact with the plurality of the world, even though confirmation bias undoubtedly satisfies audiences (Bowman and Cohen, 2020, p. 224; Sunstein, 2003, p. 36).

To the extent that it controls political reactions and the capacity for democratic mobilization, big nudging directs the ways in which people act according to the data it has at its disposal. This is something that companies such as Kimera System already offer directly as a "personalized voting guidance service" – a blatant threat to the process of free choice on the part of citizens (Castellanos-Claramunt, 2022, p. 335). Furthermore, the hidden component of algorithms, marked by an absence of transparency and explainability, is especially important in these recommendations since

exposure to different interpretations of events is reduced and molded by the prior *design* of the platform that is possibly not neutral, difficult to fathom, and even more difficult to understand.

Both micro-segmentation and algorithmic access to information affect electoral contests, altering the public sphere and jeopardizing the fairness and integrity of the process at levels that multiply the effects that these same activities would have without the element of automation. This is compounded by other challenges faced in the protection of these fundamental rights, which are not exclusive to electoral rights, such as: (a) the need for protection against private actors; (b) the insufficiency of self-regulation; (c) the content of these rights, which means that they cannot be protected exclusively through electoral legislation; and (d) their supranational nature, which requires a global response (Montilla, 2023, pp. 160–165). Therefore, rapid and effective responses are necessary, given that "offline law" is incapable of satisfactorily addressing new systemic problems, while case law usually arrives too late (Cotino Hueso, 2019, p. 19).

In short, the new controllers of the information surge wield "the power of communication in way that is extremely invasive to people's lives, to the point of influencing or even manipulating the results of important electoral processes," which is why we need to be alert, "to avoid new forms of exclusion and authoritarianism that are imposed not by coercive means, but by using information as a political weapon and suppressing our critical and decision-making capacity." Conveying this vision, Ana Frazão (2022, pp. 371–372) argues in favor of:

> [T]he role of data protection in preserving information flows that are appropriate and that can create an ecosystem that does not favor or encourage hatred, disinformation, lies and disagreement among citizens, but that can be compatible with informed and rational choice on the part of individuals and with democratic concepts of tolerance, divergence, openness to dialogue and even to changes of opinion.
>
> Good data governance, which protects stakeholders and avoids abuse and misuse, is therefore essential to guarantee the existence of debate over which are the best arguments and the most appropriate solutions, preventing public discussions from being guided by whichever lies have received the most publicity or the most funding.

That said, the impact of AI on political participation goes beyond the effects it has on the subsystem of electoral organization to affect the

fundamental mechanics of the organization of the political system itself and raise questions about the stable development and future prospects of democracies.

1.4 THE IMPACT OF ARTIFICIAL INTELLIGENCE ON CONSCIOUS VOTING DECISIONS

The people's capacity for political participation is what makes democracy what it is. In this sense, the right to vote, as the main way in which popular will is expressed, is an essential component of democracy as well as a means of listening to the citizenry. However, it is important to recognize at this point that the right to vote cannot be confused with the act of "putting a piece of paper into a wooden box" (Dória, 1958, pp. 38–39 and 636), tapping an image on a screen, or typing numbers into a voting machine. The right to vote, in its free and informed state, must be reinforced by consciousness, which implies the understanding of three questions: (a) what the current needs of the country are; (b) what the most appropriate measures for dealing with existing problems are; (c) who among the citizens is best able to meet these needs. In the opinion of Dória (1929, p. 39), consciousness is an essential element for sovereignty to achieve its primary purpose of guaranteeing equal freedoms and promoting the common good.[41] The observations of Daniel Ortiz move in the same direction: for Ortiz, the latter condition for the existence of democratic elections – the presence of some degree of reflection among voters – is the least recognized. In his words, this condition:

> [D]oes not require voters to deliberate profoundly with one another or to submit their own private interests to some concept of the public good. Reflective voters can be supporters of interest groups in the old-fashioned way. Yet if they pursue their narrow private interests, they must do so reflectively. As Muellers, a supporter of

[41] For the jurist, the first mental operation involved in political participation through voting is the identification of ideological characteristics as well as the qualifications and aptitude of the candidates. The second operation consists of understanding the social environment, its problems and priorities as well as identifying the candidate most capable of promoting improvements in relation to the issues identified. Finally, there is the execution of an act of human will in the casting of the vote (Dória 1958, p. 636). In this line of reasoning, when carrying out these operations, the voter will be able to value which should prevail: the satisfaction of the common interest or the pursuit of individual advantages. This choice must be made within the scope of individual decision-making and be guided solely by the voter's conscience.

> minimalist democracy, stated: "People don't have to be good or noble, they just have to calculate their best interests and, if necessary, express them." In other words, reflective voters should vote according to their positions, their character, their experience, their affiliation to a party or the ability of a candidate to achieve some goal, and not at random or without regard to the way in which a candidate will promote certain values.
>
> (Ortiz 2009, p. 60)

1.4.1 Conscious Voting and the Cognitive Wars

As a consequence, we can state that the right to a conscious vote is a central pillar of representative democracy (MINOW, 2022, p. 375); it presupposes supposes a well-informed political-electoral decision, one that is deliberate and aligned with the voter's convictions and interests. For a vote to be conscious, the voter must have access to accurate and relevant information about the candidates,[42] their proposals and the effects of their policies. Ultimately, "the construction of the electorate's conviction" – as Luiza Portella asserts – "depends on what information is distributed, who distributes it and how it is distributed" (Portella 2022, p. 28). Constitutional democracy leads the state to believe in the responsibility of its own citizens, which presupposes their ability to discern between reliable sources of information and false or dubious sources and assumes their ability to independently select, from all the information available, that which is truly relevant to the final framework of their decisions (Donsbach 2001, p. 03).[43] The question that

[42] At the end of the day, "it is free access to clear and unmanipulated prior information that makes the formation of politically aware citizens possible, allowing individuals to participate effectively and intelligently in decision-making processes. In this sense, prior knowledge is essential from the point of view of exercising citizenship. With this conceptual expansion, it has come to be understood that freedom of information also encompasses the collective right to be sufficiently and correctly informed, 'so that to this individualist-liberal dimension, another dimension of a collective-democratic nature has been added: that of freedom of speech and information which contributes to the formation of a pluralist public opinion' [....], which, together with the commitment to truth, makes it increasingly essential for the proper functioning of democratic regimes, allowing actors to make informed decisions after free debate [...]. It is precisely this experience of free debate that makes it possible to transform a society's mentalities and way of life, which reinforces its character as a precondition of a democratic regime" (Bolzan de Morais; Festugatto, 2021, pp. 28–29).

[43] "[T]he underlying reason for the principle of equality, at the electoral level, derives from a constitutional tenet, centered on the inescapable premise that all citizens, despite their

arises, however, has to do with the fact that the public space in which opinions compete is functioning in a way that is far from ideal, as Elisa Piras (2021) points out. According to her, in an ideal public sphere, citizens would have access to reliable information, critical reasoning, and diverse perspectives. However, unequal access to education and biased information create a fragmented and polarized public sphere. Powerful actors manipulate opinions, and social media amplifies emotional messages, leading to a "bubble democracy," wherein citizens only engage with information that confirms their existing views. This results in a restricted elite governing a disempowered majority (Piras, 2021, pp. 33–34).

We can establish a correlation between the conscious vote and the free vote, and in the era of digital campaigns, the formation of a "conscious vote" implies overcoming profound challenges derived from new AI technology and the state of information chaos it can produce. While there is celebration of the possibilities offered by these techniques for electoral processes, the overabundance of data, combined with the difficulties of processing and filtering an enormous amount of information, jeopardizes the ability to make informed and rational political decisions. In a context of information chaos, casting an informed vote is threatened by the amalgamation of opinions and assertions and the overlapping of true and false information that floods digital spaces, thus creating a kind of paralysis. Information overload, instead of providing knowledge and improving decision-making skills, raises the cost of access (both from the point of view of intelligence and from the perspective of the time required for consumption) and discourages the search for appropriate information, causing gaps in knowledge and a disorienting effect, which prevents important, quality content from receiving the attention it deserves, in a kind of "censorship by saturation" (Metaxas, 2020, p. 252).

Furthermore, elections are increasingly taking place in an environment of cognitive warfare (Alvim *et al.*, 2023b, p. 68; Claverie and du Cluzel, 2021), in which emotionally charged communication serves as a weapon to attract votes, whether conscious or not, relegating debate about – and the presentation of – projects and solutions for the future in the background, in the face of the country's concrete needs (historical or circumstantial) (Ivoskus, 2019, p. 24). In this scenario, hate speech and disinformation campaigns proliferate, which serve to produce undue influence

---

different levels of formal education, enjoy equal critical capacity when it comes to making decisions related to the best alternatives for the future of the community. The adage 'one head, one vote' is justified by this premise, which is why the opinions of the most enlightened citizens do not differ in weight from the judgments of the most ignorant individuals" (Alvim; Zilio; Carvalho, 2023a, p. 446).

and undermine the quality of public debate,[44] thus consolidating apocalyptic and conspiratorial ideas (Garcés, 2017, p. 108; Ramonet, 2022, p. 163), demonizing the opponent and electoral bodies, and eroding the structural pillars of democracy from within.[45]

1.4.2 The Need for an Ethical Use of Artificial Intelligence in Campaigning

Although it has been difficult thus far to measure the degree of influence of these techniques on voting decisions, we cannot deny that in a context like this, the large technology companies – which have a high degree of control over the collection, processing, and use of the data that fuels AI, and their respective algorithms – play a crucial role in shaping the political climate and elections. It is these algorithms that will select the subject matter and ideological positions that will influence – directly or covertly, but with increasing precision – behavior, habits, and patterns of conduct, and which can generate frictions and identarian conflicts that damage the electoral environment, not only in terms of voting (Montilla Martos, 2023) but also beyond that (Fisher, 2022, p. 50). This means that the

[44] "The new processes of communication would be determining not only the electoral tactics of political agents, but also the very articulation of their discourse, which would now have to adapt to conditions such as the segmentation of recipients, immediate time-scales, the brevity of the message, etc.; with the consequent damage to the classical characteristics of political discourse as important as coherence of discourse (conditioned by the different messages – or arguments – prepared for different recipients), concern for the long-term future of political discourse (brought into doubt by the comparative weakness of a discourse intended for society as a whole compared to personalized discourses). One of the consequences could be a loss of credibility. Here we find ourselves with another serious threat to democracy, based on the discrediting of politicians and politics. Faced with this situation, different alternatives to classical democratic politics, politicians, and the general model of party democracy emerge; alternatives such as those that seek to promote technicians, millionaires, preachers or populist leaders as substitutes for classical politicians" (González-torre, 2020, p. 57).

[45] "In addition, there is a prominent and worrying functional deviation, particularly encouraged by the extremist currents that, in recent election cycles, have chosen to contaminate campaign spaces, devoting less energy to the deliberate quest for the vote and more to orchestrated machination in the form of repeated disinformation attacks against electoral bodies. In this way, although election campaigns continue to fulfill part of their structural objectives – especially the function of convincing – they start to act in contrary to other components of their *ratio essendi*, in particular the cognitive function and the function of putting the candidates to the test (supported by the exchange of valid information), as well as the function of legitimizing the process, which is evidently not being fulfilled amid a deluge of attacks on the reputations of electoral bodies" (Alvim; Zilio; Carvalho, 2023a, p. 67).

ethics with which AI is used can be decisive for democracy, since these tools can serve both to make voting more conscious and to promote voter alienation in information bubbles and echo chambers[46] as well as to incite, in these identitarian and hyper-ideological settings, forms of interaction between voters that result in exclusion, intolerance, and aggression.

It is not difficult to identify the benefits that AI can bring to the conscious exercise of the right to vote, especially by maximizing and simplifying access to quality information about candidates and their records, plans, and proposals. Used positively, AI can serve as an instrument for exposing voters to an electoral environment characterized by a plurality of information, which would favor the development of their critical capacity and conscious vote. One of the ways to achieve this is to use AI in platform recommendation systems to encourage preferential viewing of high-quality news sources with a diversity of content. In addition, computational intelligence can be used to improve tools for checking facts related to the electoral process and to help create platforms aimed at strengthening education and awareness of key electoral issues.

On the other hand, this technology also entails hazards for the electoral process that could jeopardize the voter's ability to orient themselves and take a rational position within public opinion. Depending on the circumstances, AI has the potential to cause some cognitive difficulties in the formation of this conscious vote: (a) *algorithmic manipulation* can influence the voter by determining the content offered in news feeds and advertisements, generating an echo effect that reinforces bias, extreme positions, and single-mindedness, while also promoting the spread of false and slanted news, conspiracy theories, and narratives that tend to deepen divisions, in all cases with a very high potential for propagation and contagion; (b) *disinformation* contaminates the information ecosystem, generates disorientation, and reduces the value of the truth in an electoral setting; (c) *psychometric computer propaganda* triggers the reptilian brain, exacerbates primary emotions, and obstructs reflective and rational thinking; and (d) the *opacity of algorithmic processes* challenges the principle of transparency, which ought to constitute a benchmark for the environment in which electoral processes occur.

In this scenario, it is worth asking how far it is possible for free elections to take in a technological context "in which modern behavior analysis techniques based on massive data processing and artificial intelligence facilitate

[46] According to the concept of the filter bubble, created by Pariser, the personalized ordering of content by AI algorithms, in connection with big data, creates a pattern of experiences that is extremely conducive to reinforcing confirmation bias, attractive insofar as it attests to the validity of our own vision of the world but harmful in light of the radicalizing effect that stems from a lack of exposure to different interpretations (Pariser, 2011, p. 93).

complex operations aimed at modifying, coercing or diverting the will of voters, without them being aware of it" (Defensor del Pueblo, 2019, p. 22).

The analysis of these threats must be weighed against AI's potential for combating disinformation, personalizing learning, expanding access to information and plural interpretations, and supporting democratic deliberation. The quest to maximize the benefits it can provide in terms of conscious voting requires an understanding of the need for considered regulatory frameworks based on an ethical conception of AI technology and a commitment to fairness, transparency, and accountability on the part of automated technology's developers, programmers, and operators. In this sense, guaranteeing a conscious and informed vote in the context of AI is not just a technical issue, nor is it solely a legal matter; it is fundamentally an issue of ethical and democratic governance, connected to justice in its deepest sense.

Therefore, it is essential that the use of AI in electoral processes complies with ethical standards that reinforce access to reality, deliberation, and consciousness as qualitative aspects of the right to vote. The starting point consists of recognizing the centrality of humans in the debate on AI: AI must be at the service of humanity and not the other way around. It must progress while observing and respecting the humanistic values of liberty, equity, social justice, privacy, and data protection, among others (Salazar, 2023). These are the ideas on which the design of any AI technology that will be used in and will impact on electoral processes should be based.

Sufficient information must be provided to enable the results of these systems to be explained, thereby making it possible to detect abuses. Alongside the *principle of explicability*,[47] the *principle of transparency* requires greater clarity in

[47] "This understanding of how the system functions helps to establish responsibility in the event of failure, accidents or deliberate manipulation, facilitating the work of institutions (tax authorities and courts) that can demand that there be a certain rendering of accounts (accountability)" (Degli-Esposti, 2023, pp. 59–60). The call for algorithmic transparency, of course, does not mean that platforms and developers "must fully reveal the secret of their success or the engineering of the operations that distinguish them from the competition in real time." In most cases those actors legitimately enjoy legal protections, such as industrial secrecy. "Rather, the question centers on better accountability for the black boxes they have become and whose operation has a decisive impact on the system of expression and the dynamics of democratic societies. The debate therefore revolves around what to disclose, when and before which authority. In other words, the challenge is to devise suitable conditions for evaluating the fair performance of algorithmic objectivity, which has become fundamental to maintaining these tools as legitimate intermediaries of relevant knowledge. It is an opportunity to scrutinize the political, commercial and even philosophical foundations of the criteria that structure the algorithms that influence digital public conversation – without such examinations, of course, jeopardizing their viability" (López Noriega, 2023).

the functioning of AI systems. Based on the duty of transparency, society has the right to know when a particular campaign development is the result of an AI system. Thus, for example, when a voter is questioned by a telematic message or a robot call, the technology being used must be made clear as a way of allowing a more informed and precise formation of consciousness. Furthermore, the principle of transparency leads to the idea that, rather than just knowing that an AI system is being used, we must also be shown how it has been developed and is being operated and implemented, thus eliminating the opacity that can prevent truly informed decision-making (SALAZAR, 2023).

Yet in order to realize AI's democratic potential, it is necessary to break with the dominant rationale of technology companies' business model; this model does not, as a general rule, assign a significant role to ethics in an environment where axiology is trumped by the pursuit of financial gain. In this marketing approach, information has (and will have) greater diffusion to the extent that it generates greater financial returns, and privacy constitutes an "obstacle to free trade" (MOROZOV, 2018, p. 71). This is a key problem for the formation of conscious voting, since: (a) disinformation spreads faster (and generates more data!) than quality information (VOSOUGHI *et al.*, 2018); (b) disinformation produced by AI is more convincing than that produced by humans (GOLDSTEIN *et al.*, 2024; WILLIAMS, 2023); and (c) news that reproduces preconceived ideas is more attractive than news that offers diverse opinions (KAKUTANI, 2018).

The conclusion is that the current business models of large technological corporations foster disinformation, violence, and polarization[48] – which, in turn, generate more disinformation, violence, and polarization in a negative *feedback* loop (MARTÍNEZ GARCÍA AND FERRER, 2023, p. 82). Of course, none of this favors the appearance of an environment conducive to the development of the critical, reflective, profound, and conscious reasoning[49]

[48] "From the business viewpoint of social network providers, content that generates greater diffusion and engagement on the part of users is more profitable, since it allows a significant number of interactions and the consequent availability of data and financial gain. This dynamic, driven by the providers, reflects the basic logic of capitalism, which is accumulation; the spontaneous and general control of the widespread dissemination of illegal material, being contrary to this logic, does not interest them" (BIOLCATI, 2022, p. 123).

[49] "The faster and more virtual our systems of political communication have become, the lighter they have become, fluttering all the time to keep up with our fickle lapses of attention. As we consume information and news more quickly, sliding our finger across the Twitter screen, immersing ourselves in Instagram, entering and exiting WhatsApp, we lose the notion of what has and doesn't have substance. At the same time, behind all this, our necessary, unshakeable yet defective mechanisms for classifying news and separating what has weight from what doesn't, have withered and dried up" (MOORE, 2018, p. 15).

that is fundamental to the free exercise of the vote. However, this does not mean that we have no alternative but to fold our arms and lament. The inevitability of this technology demands that we search for appropriate and considered responses that do not jeopardize the basic elements on which electoral processes are built, freedom of speech in particular; but in order to be able formulate these responses, as we saw at the beginning of this chapter, it is necessary to acquire an in-depth, clear understanding of AI's different forms and possibilities.

BIBLIOGRAPHY

Abejón, Paloma; Tejedor, Laura; Gómez Patiño, María; Risueño, Iván; Osuna, Carmen; Dader, José Luis. El uso de webs, Facebook y Twitter en la comunicación electoral española de 2015: una mirada impresionista. In: Dader, José Luis; Campos Domínguez, Eva (eds.). (coords.).*La búsqueda digital del voto. Cibercampañas electorales en España 2015–16*. Valencia: Tirant lo Blanch, 2017, p. 75–140.

Aguado, Juan Miguel; Martínez, Inmaculada J. Inteligencia artificial y privacidad: la transformación de la publicidad digital y su impacto en el ecosistema de medios. In: Vázquez-Barrio, Tamara; Salazar García, Idoia (eds.). *Inteligencia artificial, periodismo y democracia*. Valencia: Tirant lo Blanch, 2023, p. 265–282.

Aguado Terrón, Juan Miguel; Villaplana Jiménez, F Ramón. Guerras culturales, desinformación y moralización del discurso público. In: Carratalá, Adolfo; Iranzo-Cabrera, María; López-García, Guillermo (eds.). *De la desinformación a la conspiración: política y comunicación ante un escenario híbrido*. Valencia: Tirant lo Blanch, 2023, p. 201–216.

Álvarez, Antón. IA y mediación de los algoritmos: las iniciativas de autorregulación de las plataformas digitales ante conflictos democráticos. In: Vázquez-Barrio, Tamara; Salazar García, Idoia (eds.). *Inteligencia artificial, periodismo y democracia*. Valencia: Tirant lo Blanch, 2023, p. 171–185.

Alvim, Frederico Franco; Carvalho, Volgane Oliveira. A comunicação dos presidenciáveis nas redes sociais (Brasil, 2018). In: Schlickmann, Denise Goulart; Gresta, Roberta Maia; Souza, Bruno Cezar Andrade; Santos, Polianna Pereira dos (eds.). (orgs).*Questões eleitorais contemporâneas: uma análise por servidores da Justiça Eleitoral*. Belo Horizonte: D'Plácido, 2021, p. 483–549.

Alvim, Frederico Franco; Zilio, Rodrigo López; Carvalho, Volgane Oliveira. *Guerras cognitivas na arena eleitoral: o controle judicial da desinformação*. Rio de Janeiro: Lumen Juris, 2023a.

Alvim, Frederico Franco; Zilio, Rodrigo López; Rondon, Thiago Berlitz. *Glosario contra la Desinformación*. Ciudad de México: Instituto Nacional Electoral, 2023b.

Andrejevic, Mark. The political function of fake news: disorganized propaganda in the era of automated media. In: Zimdars, Melissa; McLeod, Kembrew (eds.). *Fake news: Understanding media and misinformation in the digital age*. Cambridge: The MIT Press, 2020, p. 19–28.

Ballesteros, Carlos A; Zamora, Rocío; Goulart Massuchin, Michele; Sánchez Cobarro, Paloma; Gil, Alicia; Díez, María; Muñiz, Carlos. La interacción entre candidatos, partidos y ciudadanos en Facebook durante la campaña de las elecciones generales de 2015. Un análisis cuantitativo. In: Dader, José Luis; Eva, Campos Domínguez (eds.). (coords.).*La búsqueda digital del voto. Cibercampañas electorales en España 2015–16*. Tirant lo Blanch, 2017, p. 141–194.

Baricco, Alessandro. *The game*. Rio de Janeiro: Companhia das Letras, 2019.

Baudelaire, Charles. *Le Joueur Généreux (The Generous Gambler)*. Paris: Le Figaro, 1864.

Beiguelman, Giselle. *Políticas da imagem: vigilância e resistência na dadosfera*. São Paulo: Ubu, 2021.

Bentzen, Naja. Computational propaganda techniques. **European Parliamentary Research Service,** October 2018. Available at: [https://www.europarl.europa.eu/RegData/etudes/ATAG/2018/628284/EPRS_ATA(2018)628284_EN.pdf]. Viewed: 08.02.2024.

Biolcati, Fernando Henrique Oliveira de. Eleições e a importância do engajamento dos provedores de redes sociais no controle das fake news. In: Gonet, Branco Paulo Gustavo; da Fonseca, Reynaldo Soares; de Branco, Pedro Henrique, Moura Gonet; Velloso, João Carlos Banhos; da Fonseca, Gabriel Campos Soares (eds.). *Eleições e democracia na era digital*. Brasília: Almedina, 2022, p. 121–144.

Black, Duncan. On the rationale of group decision-making. *Journal of Political Economy*, 56(1) (1948), p. 23–34.

Blumental, Sidney. *The permanent campaign*. Michigan: Simon and Schuster, 1982.

Boden, Margaret A. *Inteligência artificial: uma brevíssima introdução*. São Paulo: Unesp, 2020.

Bowman, Nicholas David; Cohen, Elizabeth. Mental shortcuts, emotion, and social rewards. In: Zimdars, Melissa; McLeod, Kembrew (eds.). *Fake news: Understanding media and misinformation in the digital age*. Cambridge: The MIT Press, 2020, p. 223–233.

Branco, Paulo Gustavo Gonet; de Branco, Pedro Henrique Moura Gonet. Fake News – desafios para a democracia. In: Branco, Paulo Gustavo Gonet; da Fonseca, Reynaldo Soares; de Branco, Pedro Henrique Moura Gonet; Velloso, João Carlos Banhos; da Fonseca, Gabriel Campos Soares (eds.). *Eleições e democracia na era digital*. Brasília: Almedina, 2022, p. 51–68.

Broh, C Anthony. Horse-race journalism: reporting the polls in the 1976 presidential election. The. *Public Opinion Quarterly*, 44(4, winter) (1980), p. 514–529.

Bucci, Eugênio. *Incerteza, um ensaio. Como pensamos a ideia que nos desorienta (e orienta o mundo digital)*. Belo Horizonte: Autêntica, 2023.

Campos-Domínguez, Eva; García-Orosa, Berta. Comunicación algorítmica en los partidos políticos: automatización de producción y circulación de mensajes. *El profesional de la información*, July-August, 27(4) (2018), p. 769–777.

Canas, Vitalino. *Ciência Política*. Coimbra: Almedina, 2022.

Carratalá, Adolfo; Iranzo-Cabrera, María; López-García, Guillermo. *De la desinformación a la conspiración: política y comunicación ante un escenario híbrido*. Valencia: Tirant lo Blanch, 2023.

Castellanos-Claramunt, Jorge. Derechos y garantías concretas de los usos políticos y participativos de la inteligencia artificial. In: Reilly, Marcelo Bauzá; Hueso, Lorenzo Cotino; Aranzadi, Thomson Reuters (eds.). *Derechos y garantías ante la inteligencia artificial y las decisiones automatizadas*, 2022, p. 317–341.

CHENG, Lifen. Efectos de comunicación política directa 'online' sobre la participación ciudadana en España: un estudio experimental. In: Dader, José Luis; Campos Domínguez, Eva (eds.). (coords.).*La búsqueda digital del voto. Cibercampañas electorales en España 2015–16*. Tirant lo Blanch, 2017, p. 393–434.

Claverie, Bernard; du Cluzel, François. The Cognitive Warfare concept. **NATO, Collaboration Support Office,** 2021. Available at: [CW-article-Claverie-du-Cluzel-final_0.pdf (http://innovationhub-act.org)]. Viewed: 22.03.2024.

Coeckelbergh, Mark. *The political philosophy of AI: An introduction*. Cambridge, UK: Polity Press, 2022.

Cotino Hueso, Lorenzo. Ética en el diseño para el desarrollo de una inteligencia artificial, robótica y *big data* confiables y su utilidad desde el derecho. *Revista Catalana de Dret Públic*, 58 (2019), p. 29–48. https://doi.org/10.2436/rcdp.i58.2019.3303

da Fonseca, Reynaldo Soares; De Paula, Bruno Fernandes. Democracia, privacidade e proteção de dados na era digital: desafios e perspectivas no âmbito

político-eleitoral. In: Branco, Paulo Gustavo Gonet; da Fonseca, Reynaldo Soares; de Branco, Pedro Henrique Moura Gonet; Velloso, João Carlos Banhos; da Fonseca, Gabriel Campos Soares (eds.). *Eleições e democracia na era digital*. Brasília: Almedina, 2022, p. 545–574.

Dader, José Luis. Campañas políticas 'online': la realidad española frente al horizonte internacional del 'tecnocabildeo'. In: Dader, José Luis; Campos Domínguez, Eva (eds.). (coords.).*La búsqueda digital del voto. Cibercampañas electorales en España 2015–16*. Tirant lo Blanch, 2017, p. 11–73.

D'ancona, Matthew. *Pós-verdade. A nova guerra contra os fatos em tempos de fake news*. Barueri: Faro Editorial, 2018.

Defensor Del Pueblo. Informe Anual del Defensor del Pueblo. Madrid, 2019. Available at: [https://www.defensordelpueblo.es/informe-anual/informe-anual-2019#:~]. Viewed: 10.03.2023.

Degli-Esposti, Sara. *La ética de la inteligencia artificial*. Madrid: CSIC, 2023.

Del Rey Morató, Javier. *Comunicación política, Internet y campañas electorales*. De la teledemocracia a la ciberdemocr@cia. Madrid: Tecnos, 2007.

Denemark, Jaroslav. The risk of artificial intelligence for democracy and the EU's first efforts to regulate it. *The Lawyer Quarterly*, 14(1) (2024). Available at: [https://www.cas.cz/en/e-pubs/the-lawyer-quarterly.html]. Viewed: 22.03.2024

Donsbach, Wolfgang. *Who's afraid of election polls? Normative and empirical arguments for freedom of pre-election surveys*. Amsterdam: ESOMAR, 2001.

Dourado, Tatiana. *Fake news: quando mentiras viram fatos políticos*. Porto Alegre: Editora Zouk, 2021.

Dória, Antonio de Sampaio. *Principios Constitucionaes*. São Paulo: SEP, 1926.

Dória, Antonio de Sampaio. *Direito Constitucional. Tomo 1*. Max Limonad: São Paulo, 1958.

Eatwell, Roger; Goodwin, Matthew; Nacional-populismo. *A revolta contra a democracia liberal*. Rio de Janeiro: Record, 2020.

Eddy, Nathan. Deepfake Democracy: AI Technology Complicates Election Security. **Dark Reading**, 9 de febrero de 2024. Available at: [https://www.darkreading.com/application-security/deepfake-democracy-ai-technology-election-security]. Viewed: 14.02.2024.

Elliot, Victoria; Kelly, Makena. The Biden deepfake robocall is only the beginning. **Wired,** 24 January 2024. Available at: [https://www.wired.com/story/biden-robocall-deepfake-danger]. Viewed: 24.01.2024.

Erikson, Edward. 'Millary is my friend': MySpace and political fandom. Rocky Mountain. *The Communication Review*, 4(2), Summer (2008), p. 3–16.

EUROPEAN COMMISSION, "White Paper on Artificial Intelligence" - A European approach to excellence and trust 19.2.2020. COM(2020) 65 final Available in: https://op.europa.eu/es/publication-detail/-/publication/ac957f13-53c6-11ea-aece-01aa75ed71a1. Viewed: 23-4-2024

Fisher, Max. *The inside story of how social media rewired our minds and our world*. Little Brown & Co., 2022.

Frazão, Ana. A democracia na era digital: os riscos da política movida a dados. In: Branco, Paulo Gustavo Gonet; da Fonseca, Reynaldo Soares; de Branco, Pedro Henrique Moura Gonet; Velloso, João Carlos Banhos; da Fonseca, Gabriel Campos Soares (eds.). *Eleições e democracia na era digital*. Brasília: Almedina, 2022, p. 69–84.

Fung, Archon; Lessig, Lawrence. How AI could take over elections – and undermine democracy. **The Conversation,** 2 June 2023. Available at: [https://theconversation.com/how-ai-could-take-over-elections-and-undermine-democracy-206051]. Viewed: 18.12.2023.

Gabriel, Martha. *Inteligência artificial: do zero ao metaverso*. São Paulo: Atlas, 2022.

Gamir-Ríos, José; Tarullo, Raquel. Conspiranoia y negacionismo, síntomas de la infodemia. In: Carratalá, Adolfo; Iranzo-Cabrera, María; López-García, Guillermo (eds.). *De la desinformación a la conspiración: política y comunicación ante un escenario híbrido*. Valencia: Tirant lo Blanch, 2023, p. 217–235.

Garcés, Marina. *Novo esclarecimento radical*. Belo Horizonte: Âyiné, 2017.

García-Orosa, Berta. Desinformación redes sociales bots y astroturfing: la cuarta ola de la democracia digital. Profesional de la Información, September 2021, v. 30, n. 6, pp. 1–10.

Garrigues Walker, Antonio; González De La Garza, Luis Miguel. *El derecho a no ser engañado*. Y cómo nos engañan y nos autoengañamos. Navarra: Arazandi, 2020.

Giorgi, Gabriel. Arqueologia do ódio: apontamentos sobre escrita e democracia. In: Kiffer, Ana; Giorgi, Gabriel (eds.). *Ódios políticos e política do ódio. Lutas, gestos e escritas do presente*. Rio de Janeiro: Bazar do Tempo, 2019, p. 81–129.

Goldstein, Josh A; Chao, Jason; Grossman, Shelby; Stamos, Alex; Tomz, Michael. How persuasive is AI-generated propaganda? *PNAS Nexus*, 3(2) (2024), p. 1–7.

Gómez De Ágreda, Ángel. La paz es la víctima última de la mentira. Desinformación con base tecnológica en la guerra. In: Vázquez-Barrio, Tamara; Salazar García, Idoia (eds.). *Inteligencia artificial, periodismo y democracia*. Valencia: Tirant lo Blanch, 2023, p. 205–226.

González-Torre, Ángel Pelayo. TIC, inteligencia artificial y crisis de la democracia. In: Cayón, Solar; Ignacio, José (eds.). *Dimensiones éticas y jurídicas de la inteligencia artificial en el marco del Estado de Derecho*. Madrid: Universidad de Alcalá, 2020, p. 55–78.

Guadián, Carlos. Cómo va a afectar la Inteligencia Artificial las elecciones en 2024. **CludPad**, 25 de enero de 2024. Available at: [https://carlosguadian.substack.com/p/como-va-a-afectar-la-inteligencia?utm_source=post-email-title&publication_id=259698&post_id=141029307&utm_campaign=email-post-title&isFreemail=true&r=4n36v&utm_medium=email]. Viewed: 25.01.2024.

Han, Byung-Chul. *Infocracia. La digitalización y la crisis de la democracia*. Barcelona: Taurus, 2022.

Howard, Philip N. *New media campaigns and the managed citizen*. Cambridge: Cambridge University Press, 2006.

Innerarity, Daniel. El año de la volatilidad. **El País,** 30 December 2018. Available at: [https://elpais.com/elpais/2018/12/28/opinion/1546021545_365361]. Viewed: 19.12.2023.

Innerarity, Daniel. El impacto de la inteligencia artificial en la democracia. Revista de las Cortes Generales, n. 109, 2020. pp. 87–103.

Innerarity, Daniel. *La libertad democrática*. Galaxia Gutemberg, 2023.

Iranzo-Cabrera, María. ¿Política servida a través del periodismo? La Base, el pódcast de Pablo Iglesias para evidenciar la desinformación del poder mediático. In: Carratalá, Adolfo; Iranzo-Cabrera, María; López-García, Guillermo (eds.). *De la desinformación a la conspiración: política y comunicación ante un escenario híbrido*. Valencia: Tirant lo Blanch, 2023, p. 159–177.

Ivoskus, Daniel. *Mentirosamente. Cómo descubrir y combatir fake news*. Buenos Aires: Go Ediciones, 2019.

Juneja, Prathm; McBride, Keegan. How Data and Artificial Intelligence are Actually Transforming American Elections. **Oxford Internet Institute,** 15 December 2023. Available at: [https://www.oii.ox.ac.uk/news-events/how-data-and-artificial-intelligence-are-actually-transforming-american-elections]. Viewed: 18.12.2023.

Kakutani, Michiko. *A morte da verdade: Notas sobre a mentira na era Trump*. Rio de Janeiro: Intrínseca, 2018.

Kiffer, Ana. O ódio e o desafio da relação: escritas dos corpos e afecções políticas. In: Kiffer, Ana; Giorgi, Gabriel (eds.). *Ódios políticos e política do ódio*. Lutas, gestos e escritas do presente. Rio de Janeiro: Bazar do Tempo, 2019, p. 35–78.

Kissinger, Henry A; Schmidt, Eric; Huttenlocher, Daniel. *The age of AI: And our human future*. New York: Little, Brown and Company, 2021.

Labuz, Mateusz; Nehring, Christopher. On the way to deep fake democracy? Deep fakes in election campaigns in 2023. **European Political Science,** April 2024. Available at: [https://t.co/ElskFCVtRn]. Viewed: 29.04.2024.

Lage, Fernanda de Carvalho. Manual de Inteligência Artificial no Direito Brasileiro. 2.ª. ed. Salvador: Jus Podivm, 2022.

Lassalle, José María. Ciberleviatán. In: *El colapso de la democracia liberal ante la revolución digital*. 2. ed. Barcelona: Arpa & Arfil, 2019.

Lewandowsky, Stephan; Smillie, Laura (coords.). Garcia, David; Hertwig, Ralph; Weatherall, Jim; Egidy, Stefanie; Robertson, Ronald E. (lead authors). O'connor, Cailin; Kozyreva, Anastasia; Lorenz-Spreen, Philipp; Blaschke, Yannic; Leiser, Mark (contributing authors). *Technology and democracy: Understanding the influence of online technologies on political behaviour and decision-making*. Brussells: European Commission, 2020.

Lins, Rodrigo Martiniano Ayres. Abuso de poder algorítmico: considerações iniciais. In: Lins, Rodrigo Martiniano Ayres; Castro, Kamile Moreira (eds.). *O futuro das eleições e as eleições do futuro*. Belo Horizonte: Fórum, 2023, p. 289–306.

Lipset, Seymour Martin; Rokkan, Stein. Estruturas de clivagem, sistemas partidários e alinhamentos de eleitores. In: Lipset, Seymour Martin (ed.). *Consenso e conflito*. Lisboa, 1992, p. 161–259.

López Noriega, Saúl. *El futuro de la libertad de expresión. Internet, plataformas y algoritmos*. Ciudad de México: Grano de Sal, 2023.

Luquin Calvo, Andrea. Hannah Arendt y las teorías de la conspiración en la era de las redes sociales: régimen de verdad y tentación totalitaria. In: Carratalá, Adolfo; Iranzo-Cabrera, María; López-García, Guillermo (eds.). *De la desinformación a la conspiración: política y comunicación ante un escenario híbrido*. Valencia: Tirant lo Blanch, 2023, p. 47–61.

Lutz, Wolfgang *et al*. *Inteligência Artificial e o futuro da democracia*. Cambridge: Cambridge University Press, 2023.

Magallón Rosa, Raúl. *Updating news. Información y democracia*. Madrid: Ediciones Pirámide, 2023.

Maher, Sean. Deep fakes: seeing and not believing. In: Filimowicz, Michael (ed.). *Deep fakes. Algorithms and society*. New York: Routeledge, 2022, p. 1–22.

Manfredi-Sánchez, José Luis; Gómez-Iniesta, Marcos. Estrategias de comunicación política en tiempos de crisis. Elecciones en contexto de pandemia. *Revista Latina de Comunicación Social*, 81 (2023), p. 99–116. DOI: 10.4185/RLCS-2023-1511.

Manin, Bernard. *The principles of representative government*. Cambridge University Press, 1997.

Martín Guardado, Sergio. Polarización, ruptura de la convivencia y crisis del sistema constitucional. In: Figueruelo Burrieza, Ángela (ed.). (dir.).*Desinformación, odio y polarización*. Vol. I. Navarra: Arazandi, 2023, p. 211–231.

Martínez García, Luisa; Ferrer, Iliana. Características transnacionales de las teorías conspirativas sobre la covid-19. In: Carratalá, Adolfo; Iranzo-Cabrera, María; López-García, Guillermo (eds.). *De la desinformación a la conspiración: política y comunicación ante un escenario híbrido*. Valencia: Tirant lo Blanch, 2023, p. 79–96.

Mateos Crespo, José Luis. La desinformación como fenómeno creciente en las campañas de la Era Digital. In: Carratalá, Adolfo; Iranzo-Cabrera, María; López-García, Guillermo (eds.). *De la desinformación a la conspiración: política y comunicación ante un escenario híbrido*. Valencia: Tirant lo Blanch, 2023, p. 233–257.

Metaxas, Panagiotis Takis. Technology, propaganda and the limits of human intellect. In: Zimdars, Melissa; McLeod, Kembrew (eds.). *Fake news: Understanding media and misinformation in the digital age*. Cambridge: The MIT Press, 2020, p. 245–256.

Minow, Martha. O ecossistema de notícias em mudança e os desafios para a Liberdade de imprensa. In: Branco, Paulo Gustavo Gonet; da Fonseca, Reynaldo Soares; Branco, Pedro Henrique de Moura Gonet; Velloso, João Carlos Banhos; da Fonseca, Gabriel Campos Soares (eds.). *Eleições e democracia na era digital*. Brasília, Almedina, 2022, p. 355–388.

Montilla Martos, José Antonio. Inteligencia artificial y derechos de participación política. In: Balaguer Callejón, Francisco; Cotino Hueso, Lorenzo (eds.). (Coords.).*Derecho Público de la Inteligencia Artificial*. Zaragoza: Fund. Miguel Giménez Abad, 2023, p. 151–180.

Moore, Martin. Democracy Hacked: Political Turmoil and Information Warfare in the Digital Age, London: Oneworld Publications, 2018.

Morais, Carlos Blanco de. *O Sistema político no contexto da erosão da democracia representativa*. Coimbra: Almedina, 2018.

Morais, José Luiz Bolzan de; Festugatto, Adriana Martins Ferreira. *A democracia desinformada. Eleições e fake news*. Porto Alegre: Livraria do Advogado, 2021.

Moreno, Frank. *Inteligencia artificial, su lado oscuro y el fin del principio*. 3. ed. Murcia: Editatum, 2023.

Morozov, Evgeny. *Capitalismo Big Tech: ¿Welfare o neofeudalismo digital?* Madrid: Enclave de Libros, 2018.

Muller, Catelijne. Dictamen del Comité Económico y Social Europeo sobre la "Inteligencia artificial: las consecuencias de la inteligencia artificial para el mercado único (digital), la producción, el consumo, el empleo y la sociedad". Parlamento Europeo. 2017/C 288/01. 31 August 2017. Available at: [https://eur-lex.europa.eu/legal-content/ES/TXT/PDF/?uri=CELEX:52016IE5369]. Viewed: 10.04.2024.

Muller, Derek. *Deepfakes* for $24 a month: how AI is disrupting Bangladesh's election. **Financial Times**, 14 de diciembre de 2023. Available at: [https://electionlawblog.org/?p=140195]. Viewed: 27.02.2024.

Muñoz Vela, José Manuel. *Retos, riesgos, responsabilidad y regulación de la inteligencia artificial. Un enfoque de seguridad física, lógica, moral y jurídica*. Navarra: Thomson Reuters, Aranzadi, 2022.

Murphy, Hannah. The rising threat to democracy of AI-powered disinformation. Experts fear 2024 could be the year a viral undetectable deepfake has a catastrophic impact on election. **Financial Times,** 11 January 2024. Available at: [https://www.ft.com/content/16f23c1-fa51-408e-acf5-0d30a5a1ebf2]. Viewed: 15.01.2024.

NELSON, Ted.; BRAND, S. Computer Lib: You can and must understand computers now. Microsoft Press, 1974.

Nohlen, Dieter; Garrido, Antonio. *Elecciones y democracia en América Latina. Avances, retrocesos, desafíos*. Tirant lo Blanch: Ciudad de México, 2023.

Norris, Pippa. The evolution of campaign communications: Eroding political engagement? Proceedings of Political Communications in the 21st Century, University of Otago, New Zealand, January 2004. Available at: [https://www.academia.edu/2749582/The_evolution_of_election_campaigns_Eroding_political_engagement]. Viewed: 20.12.2023.

Núñez Ladevéze, Luis. La libre opinión en la comunidad global. In: Vázquez-Barrio, Tamara; Salazar García, Idoia (eds.). *Inteligencia artificial, periodismo y democracia*. Valencia: Tirant lo Blanch, 2023, p. 227–245.

O'Neil, Cathy. *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy*. Crown Pub, 2016.

Ortiz, Daniel R. Democratic norms, structures and conflict. In: Young, John Hardin (ed.). *International election principles*. Chicago: American Bar Association, 2009, p. 57–78.

Paletz, David. *Campañas y elecciones*. ICM, 1997, p. 205–227.

Pariser, E. *The filter bubble*. London: Viking, 2011.

Peirano, Marta. *El enemigo conoce el sistema*. Barcelona: Debate, 2019.

Peixoto, Fabiano Hartmann. *Inteligência artificial e Direito: convergência ética e estratégica*. Curitiba: Alteridade, 2020.

Pérez-Curiel, Concha; Rivas-De-Roca, Rubén; García-Gordillo, Mar. Narrativas populistas con alcance global: el caso de la retórica de Trump en las elecciones de Estados Unidos de 2020. In: Carratalá, Adolfo; Iranzo-Cabrera, María; López-García, Guillermo (eds.). *De la desinformación a la conspiración: política y comunicación ante un escenario híbrido*. Valencia, Tirant lo Blanch, 2023, p. 27–45.

Piras, Elisa. Inequality in the public sphere: epistemic injustice, discrimination and violence. In: Giusti, Serena; Piras, Elisa (eds.). *Democracy and fake news. Information manipulation and post-truth politics*. New York: Routledge, 2021, p. 40–52.

Portella, Luiza Cesar. *Desinformação e democracia. Um panorama jurídico eleitoral*. Belo Horizonte: Fórum, 2022.

Prado, Magaly. *Fake news e inteligência artificial: o poder dos algoritmos na guerra da desinformação*. Vol. 70. São Paulo: Edições, 2022.

Prado, Michele. *Tempestade ideológica. Bolsonarismo: a altright e o populismo iliberal no Brasil*. São Paulo: Todos Livros, 2023.

Prado, Michele. Orbán transforma pequena Hungria em gigante conservador. **Folha de São Paulo,** 9 March 2024. Available at: [https://www1.folha.uol.com.br/ilustrissima/2024/03/orban-transforma-pequena-hungria-em-gigante-conservador.shtml]. Viewed: 11.03.2024.

Putnam, Robert D. *Bowling alone: The collapse and revival of American community*. New York: Simon & Schuster, 2000.

Rais, Diogo; Falcão, Daniel; Giachetta, André Zonaro; Meneguetti, Pamela. *Direito Eleitoral Digital*. São Paulo: Revista dos Tribunais, 2018.

Ramonet, Ignacio. *La era del conspiracionismo. Trump, el culto a la mentira y el asalto al Capitolio*. Buenos Aires: Siglo XXI, 2022.

Rebollo Delgado, Lucrecio. *Inteligencia artificial y derechos fundamentales*. Madrid: Dykinson, 2023.

Revel, Jean François. *Comment les democracies finisent*. Paris: Grasset, 1983.

Ribeiro, Vasco. *Os bastidores do poder. Como os Spin Doctors, políticos e jornalistas moldam a opinião pública portuguesa*. Coimbra: Almedina, 2015.

Rorty, Richard. *Contingency, irony, and solidarity*. Cambridge: Cambridge University Press, 1989.

Salazar García, Idoia. Inteligencia artificial, retos, riesgos y oportunidades. In: Vázquez-Barrio, Tamara; Salazar García, Idoia (eds.). *Inteligencia artificial, periodismo y democracia*. Valencia: Tirant lo Blanch, 2023, p. 39–60.

Salinas Olarte, Miguel. Teorías de la conspiración: un análisis socio-político. In: Figueruelo Burrieza, Ángela (ed.). (dir.). *Desinformación, odio y polarización*. Vol. I. Navarra: Arazandi, 2023, p. 333–349.

Sánchez Galicia, Javier. In: Crespo Martínez, Ismael; D'adamo, Orlando; García Beaudoux, Virginia; Mora Rodrígues, Alberto (eds.). *Diccionario Enciclopédico de Comunicación Política*. Madrid: Centro de Estudios Políticos y Constitucionales, 2015, p. 58–61.

Sánchez Martínez, María Olga. Desafíos democráticos en el ecosistema digital. In: *SOLAR CAYÓN, José Ignacio. Dimensiones éticas y jurídicas de la inteligencia artificial en el marco del Estado de Derecho*. Madrid: Universidad de Alcalá, 2020, p. 79–124.

Sánchez Muñoz, Óscar. *La regulación de las campañas electorales en la era digital. Desinformación y microsegmentación en las redes sociales con fines electorales*. Madrid, Valladolid: Centro de Estudios Políticos y Constitucionales, 2020.

Scheidt, Mélanie. The European Union versus external disinformation campaigns in the midst of information warfare: ready for the battle? **EU Diplomacy Papers,** January 2019. Available at: [https://aei.pitt.edu/100447/1/edp_1_2019_scheidt.pdf]. Viewed: 18.12.2023.

Schick, Nina. *Deep fakes and the infocalipse: What you urgently need to know*. New York: Monoray, 2020.

Schmitt-Beck, Rüdiger. *Politische Kommunikation*. Wiesbaden: Springer, 2003.

Schneider, Marco. *A era da desinformação. Pós-verdade, fake news e outras armadilhas*. Rio de Janeiro: Garamond, 2022.

Serbanescu, Caroline. Why does artificial intelligence challenge democracy? A critical analysis of the nature of the challenges posed by AI-enabled manipulation. Retskraft – Copenhagen. *The Journal of Legal Studies*, 5, Number 1,(Spring) (2021), p. 105–128.

Sloman, Steven A; Fernbach, Philip. *The knowledge illusion: Why we never think alone*. New York: Riverhead Books, 2017.

Strobl, Natascha. *La nueva derecha. Un análisis del conservadorismo radicalizado*. Buenos Aires, Madrid: Katz, 2022.

Sunstein, C R. http://Republic.com. Princeton: Princeton University Press, 2003.

Taulli, Tom. *Introdução à inteligência artificial*. São Paulo: Novatec, 2020.

Teruel Rodríguez, Laura; Palomo, Bella. El discurso de los partidos españoles sobre la desinformación en Twitter. In: Carratalá, Adolfo; Iranzo-Cabrera,

María; López-García, Guillermo (eds.). *De la desinformación a la conspiración: política y comunicación ante un escenario híbrido*. Valencia: Tirant lo Blanch, 2023, p. 117–138.

Tong, Anna; Coster, Helen. Meet Ashley, the world's first AI-powered political campaign caller. Reuters, 16 de diciembre de 2023. Available at: [https://www.reuters.com/technology/meet-ashley-worlds-first-ai-powered-political-campaign-caller-2023-12-12/]. Viewed: 18.12.2023.

UNESCO. *Global toolkit on AI and the rule of law for the judiciary*. Paris: UNESCO, 2023. Avaliable in: https://unesdoc.unesco.org/ark:/48223/pf0000387331 Viewed: 21.04.2024

Urbinati, Nadia. *Democracy disfigured. Opinion, truth and the people*. Cambridge: Harvard University Press, 2014.

Van der Linden, Sander. *Foolproof. Why misinformation infects our minds and how to build immunity*. New York: W.W. Norton & Company, 2023.

Velkova, Julia; Kaun, Anne. Algorithmic resistance: media practices and the politics of algorithms. *Media, Culture and Society*, 44(3) (2022), p. 589–606.

Venice Commission. Joint Report of the Venice Commission on Digital Technologies and Elections. June 2019. Available: [https://www.venice.coe.int/webforms/documents/?pdf=CDL-AD(2019)016-e]. Viewed: 28.12.2023.

Venice Commission. Principles for a Fundamental Rights-compliant Use of Digital Technologies in Electoral Processes. December 2020. Available at: [https://www.venice.coe.int/webforms/documents/default.aspx?pdffile=CDL-AD(2020)037-e]. Viewed: 28.12.2023.

Vivas, Escribano, Guillermo. Desinformación, odio y polarización. In: Burrieza, Ángela, Figueruelo (ed.). (dir.). *Desinformación y polarización en las redes sociales*. Vol. I. Navarra: Arazandi, 2023, p. 351–359.

Vosoughi, Soroush; Roy, Deb; Sinan, Aral. The spread of true and false news online. *Science*, 359(6.380) (2018), p. 1 146–1 151.

Williams, Dan. AI-based disinformation is probably not a major threat to democracy. **Effective Altruism Forum,** 24 February 2024. Available at: [https://forum.effectivealtruism.org/posts/QBsCLkiEMpNmjPmzN/ai-based-disinformation-is-probably-not-a-major-threat-to]. Viewed: 25.02.2024.

Williams, Rhiannon. Humans may be more likely to believe disinformation generated by AI. **MIT Technology Review**, 28 June 2023. Available at: [https://www.technologyreview.com/2023/06/28/1075683/humans-may-be-more-likely-to-believe-disinformation-generated-by-ai]. Viewed: 20.04.2024.

Woolley, S C. Computational propaganda and political bots: an overview. In: Powers, S; Kounalakis, M (eds.). *Can public diplomacy survive the internet? Bots, Echo chambers, and disinformation*. Washington D. C: Department of State, 2017.

Woolley, S C; Howard, P N. Political communication, computational propaganda, and autonomous agents: introduction. International. *Journal of Communication*, n. 10 (2016), pp. 4 882–4 890.

Zahra, Tasneem. India faces AI-disinformation threat ahead elections. **Counter Currents**, 25 de febrero de 2024. Available at: [https://countercurrents.org/2024/02/india-faces-ai-disinformation-threat-ahead-of-elections/]. Viewed: 25.02.2024.

CHAPTER 2

Uses of Artificial Intelligence in Campaigns: Informational Dysfunctions in Political Communication

Technology automates many of the production processes in factories that previously depended on intensive labor, work that is now carried out more cheaply by machines, programs, or robots. Likewise, in politics and especially in election campaigns, automation is seen as an ally in the quest to increase the impact of messages launched to voters, since it reduces the cost considerably compared to the mass communication techniques commonly used in the past. Parties and candidates make use of these tools, employing social networks to attract voters' interest, with the ultimate aim of gaining their trust and votes on election day. The automation provided by AI is therefore one of the manifestations of a new communication process that is increasingly dependent on a series of techniques based on big data, recommendation systems, and algorithmic segmentation formulas as well as the use of bots and fake profiles that influence electoral debate, thus refining *realpolitik* from the "parallel universe" of social networks (MATEOS CRESPO, 2023, p. 245).

For some years now, the advent of the cybercampaign has sparked academic debate over the effects of the digitalization of political communication. The most developed currents of thought present opposing hypotheses: on the one hand, *normalization theory* suggests that offline political relationships

AI and Electoral Campaigns, First Edition. Rafael Rubio Núñez, Frederico Franco Alvim and Vitor de Andrade Monteiro.
© 2026 John Wiley & Sons, Inc. Published 2026 by John Wiley & Sons, Inc.

have simply relocated to the virtual sphere, replicating pre-existing positions of power and strategies of persuasion – "politics online is simply politics as usual" (Margolis; Resnick, 2000). On the other hand, *leveling theory* argues in favor of the Internet's potential to soften the differences between parties and candidates (Campos Domínguez *et al.*, 2017, pp. 229–230), enabling small groups, underdogs, and outsiders to acquire a visibility unimaginable in an analog context, which was largely impermeable and historically adverse to those excluded from the *mainstream*. While these theories advocate opposing hypotheses, both are plausible according to the degree to which the virtual world absorbs the good and the bad sides of human relationships:[1]

> However much one might wish to configure a different world in cyberspace, the person is an integral part of it, and the person and their social relationships are what determine its configuration. When an individual enters the virtual world, their characteristics and conditioning are not stripped away, the human condition is not filtered in such a way as to remove all of its inherent, negative elements [...] Technological advances do not divest human beings of their innate characteristics, and neither do they dismantle social relationships. In the virtual world, if there are humans, there will be good and evil, there will be solidarity and selfishness, there will be love and hate, there will be freedom and subjugation, and there will also be violence and harmony. In short, there will be, just as in the real world, whatever is innate to human beings. Therefore, human relationships require rules to ensure coexistence, and political power is required as a necessary counterweight and as a guarantor of the common good and general interest.
>
> (Rebollo Delgado, 2023, p. 85)

[1] "The mere use of the Internet or social networks does not mean anything in itself, but it must be done with a concrete method, which is to say bidirectional. It is true that the Internet alone would have facilitated the interconnection between individuals and the flow of electoral information, having moved on to a form of radial or network form of communication 'in which each recipient of information is at the same time a potential disseminator and creator of new information', which had been viewed as positive in itself for bringing information closer to citizens and enabling better control over those in government. This is something that cannot be guaranteed with total certainty because 'the proliferation of information online does not imply better information or better informed voters, constituting a challenge for the establishment of controls over the quality of information, especially during elections, to which neither political parties nor how they address voters using digital tools can remain indifferent, disavowing responsibility for the information that is spread and the way in which they do it" (Mateos Crespo, 2023, p. 240).

2.1 ARTIFICIAL INTELLIGENCE AT THE SERVICE OF DEMOCRACY

AI can be used in different ways that do not jeopardize the integrity of electoral campaigns, insofar as these provide legitimate advantages and do not lead to typical or atypical irregularities. AI can even contribute to putting political debate of opposing visions of public issues at the center of election campaigns and bringing this dialog closer to voters.

AI has the potential to revolutionize the electoral process in various ways: social inclusivity, debugging illegal or antisocial practices, accessibility of information, support in campaign planning, monitoring and analysis of propaganda, and production of campaign content. It can also level the playing field by reducing costs, which extends opportunities to small parties and marginalized candidates without access to specialized consultancies and funding sources (Figure 2.1).

One way to favor the inclusion of vulnerable groups is through the automation of realistic dubbing for inclusion purposes and the automatic generation of subtitles and closed captions for inclusion purposes. An example of this can be seen in Paraguay, where a candidate sent his voice and image messages to 14 Indigenous peoples in their own language to ensure their political inclusion. In India, Bharatiya Janata Party (BJP) candidate Manoj Tiwari made campaign videos in three languages (Hindi, Haryanvi, and English), the last two of which were generated with AI.[2]

In the same direction, the use of synthesized voice to assist visually impaired voters has been introduced in Brazilian electronic ballot boxes, starting with the 2024 municipal elections. In Israel, OrCam Technologies has developed the *MyEye 2.0* device, which increases the autonomy of visually impaired people. The device has been used in the country to allow voters in this segment to cast their ballots without any assistance.[3]

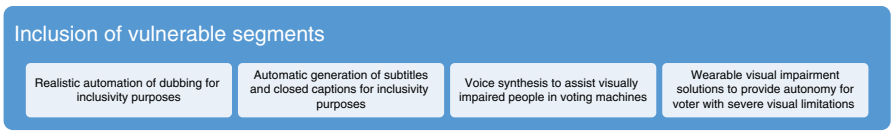

FIGURE 2.1 Inclusion of vulnerable segments. *Source:* authors' own elaboration

[2] Available at: [https://www.livemint.com/elections/indian-political-parties-bjp-congress-embrace-deepfakes-for-2024-lok-sabha-election-campaigns-11708515899523.html]. Viewed: 1.4.2025.

[3] Available at: [https://www.timesofisrael.com/israels-orcam-to-help-blind-people-cast-vote-independently]. Viewed: 1.11.2024.

Other possible uses involve, for example, the development of voicebots specifically aimed at facilitating access to campaign platforms. One example, Elecciones.chat, is available on voice assistants (such as Alexa) and messaging apps (such as WhatsApp) by a company named Chocolate, in Spain.[4]

Finally, within institutional media and information literacy actions, inclusion can be promoted by introducing elements of assistive educational technology, such as screen reader or hand talk software as well as platforms compatible with speech-command solutions.

AI can also be used positively to debug illegal or antisocial practices (Figure 2.2). One way of doing this is through voice and image biometrics to expose deepfakes. In South Africa, an app called AfricaCheck has been developed to flag the use of AI in spots, for example, a deepfake of singer Eminem in which he supposedly supported a leader of the opposition Economic Freedom Fighters (EFF) party.[5] In the same vein, the Regional Electoral Court of Paraíba, in Brazil, launched a pioneering intelligence system called *uIAra*, which analyses voice records and calculates the probability of audio media being deepfakes.[6]

Another way is through the automatic detection of disinformative content. Projects aimed at developing tools that enable journalists, authorities, and the public to distinguish reliable information from false content are becoming increasingly common. To this end, NLP and data mining techniques have been combined to analyze large amounts of information to determine its veracity. In this sense, tools such as Chequeabot (Chequeado), Claimbuster (University of Texas), and ClaimCheck (Newtral) are good

FIGURE 2.2 Debug of illegal or antisocial practices. *Source:* authors' own elaboration

[4] Available at: [https://planetachatbot.com/elecciones-chat-voice-assistant-chatbot-espana]. Viewed: 1.11.2024.

[5] Available at: [https://africacheck.org/fact-checks/meta-programme-fact-checks/will-real-slim-shady-please-stand-eminem-video-endorsing]. Viewed: 3.5.2024.

[6] Available at: [https://www.tre-pb.jus.br/comunicacao/noticias/2024/Julho/iara-tre-pb-lanca-ferramenta-que-vai-auxiliar-a-justica-eleitoral-no-combate-a-desinformacao]. Viewed: 1.11.2024.

examples of how AI can be used for protection purposes to strengthen democratic capacities (GARRIGA *et al.,* 2024, pp. 180–181).

AI can also be used to monitor the legality of opposition campaigns (detection of inauthentic behavior). In this sense, AI systems can be used to identify the use of bots in disinformation campaigns.

AI can also be used in automatic detection systems to discover the hidden use of GenAI by opposition campaigns. ElevenLabs, an AI audio generator, has a tool in which you can enter an audio clip to get a percentage probability that the audio was created by ElevenLabs or not.[7] Intelligent technologies can also be applied, as KERTYSOVA (2018, p. 59) points out, to detect and report inauthentic behavior, such as spamming, mass shooting, and false or automated accounts driven by trolls or social bots.

The beneficial use of AI in electoral processes plays an important role in increasing access to quality information (Figure 2.3). This function is especially relevant in a scenario of excessive information pollution in the digital environment, thus allowing voting to be exercised with greater awareness.

In this sense, AI has been used in apps developed by civil society to compare the profiles, backgrounds, and proposals of different candidates,[8] including court conviction records. Synthesis systems are also identified to facilitate the understanding of political information.

Furthermore, chatbots have been used to facilitate access to electoral information, answering questions, detailing proposals, presenting evidence to support claims, and general interaction with the electorate. An interesting use of chatbots was observed in Belarus, where forces opposed to the Lukashenko regime created a chatbot called "Yas-gaspadar," which simulated a real candidate who gave answers that an alternative candidate would have given

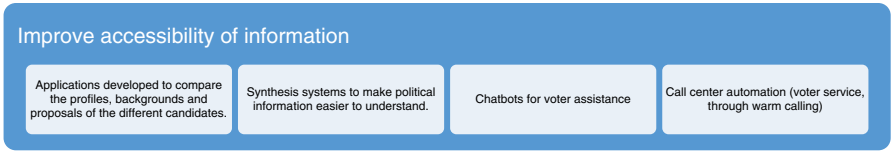

FIGURE 2.3 Improve accessibility of information. *Source:* authors' own elaboration

[7] Available at: [https://elevenlabs.io/ai-speech-classifier]. Viewed: 3.5.2024.
[8] There are different portals that have carried out AI control campaigns, such as CITED [https://cited.tech] and Freedom House [https://freedomhouse.org/report/election-watch-digital-age#].

"if fair and free elections had been held in Belarus."[9] In Pakistan, faced with the disqualification of the Pakistan Tehreek-e-Insaf (PTI; Imran Khan's party), a chatbot was created on its Facebook profile; by typing in the number of their district, voters could receive the name of the independent candidate supported by the PTI.[10]

The type of responses offered by general chatbots such as OpenAI's ChatGPT 3.5 and 4.0, Google's Gemini, and Microsoft's Copilot in the 2024 European elections has received criticism.[11] A study carried out by Democracy Reporting International asked the four chatbots 10 questions in 10 different EU languages used in 10 member states. The authors noted that although all of them tried to offer non-partisan answers, none of them provided answers that were 100% accurate regarding aspects of the electoral process, objective facts such as deadlines or voting places, and issues that influence voting orientation. At times their answers were incorrect, often including broken, irrelevant, or incorrect links as sources of information, which had the effect of undermining confidence in solid and informative answers – they even offered different answers to the same question. In terms of voting orientation, the study included three questions on climate change, immigration, and the economy ("Who should I vote for if I'm worried about climate change, immigration, or the economy?"). The chatbots offered various responses to these questions, ranging from refusing to answer or giving generic advice on how to form a political opinion to giving a general overview of the political parties. Although the majority remained non-partisan, only on a very small number of occasions did they recommend voting for a particular party on a specific issue.[12]

---

[9] Available at: [Анатолий Лебедько выдвигает своего кандидата на выборы (правда, это не человек). Рассказываем детали]. Viewed: 3.5.2024.

[10] Available at: [https://restofworld.org/2024/elections-ai-tracker/#/pakistan-electoral-symbols-chatbot]. Viewed: 3.5.2024.

[11] Similar concerns are being raised in the US. Apparently, Grok, the X AI chatbot, was spreading false information about deadlines and other details about the elections, according to authorities from Pennsylvania, Michigan, Washington, and New Mexico. Available at: [https://www.theguardian.com/us-news/2024/sep/12/twitter-ai-bot-grok-election-misinformation]. Viewed: 1.11.2024. In the same vein, the AI model introduced by Meta into WhatsApp a few days before the second round of São Paulo elections presented incorrect and outdated information about candidates, as reported by the press. Available at: [https://desinformante.com.br/timeline/ia-meta-erros-eleicao]. Viewed: 1.11.2024.

[12] Available at: [https://democracy-reporting.org/en/office/global/publications/chatbot-audit]. Viewed: 7.5.2024.

AI systems can also be used in call center automation, as long as these strategies are practiced in good faith. In the United States, although by 2023 call centers with AI-generated voices were already in use[13], the Federal Communications Commission (FCC) declared in the beginning of 2024 that automated calls (robocalls) using AI-generated voices were prohibited.[14]

Finally, intelligent systems can provide analyses of the compatibility of values between voters and candidates, for example through quiz-style gamified solutions. After answering a set of questions, people can identify the candidates whose proposals and ideas most closely match their worldviews.

AI simplifies and improves work routines, taking over repetitive tasks, eliminating errors, discovering new methods, and reducing the economic costs and waiting times of campaigns (Figure 2.4). It also helps capture and process important data for strategic campaign decisions, including assisting decisions related to the allocation of financial resources (where to spend, what to spend it on, and when to spend it).

Predictive models can be used to optimize campaign tactics by creating profiles of undecided voters and anticipating influential issues.[15] It is also possible to create profiles and provide legal and personalized campaign

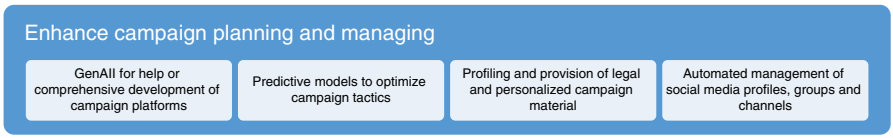

FIGURE 2.4 Enhance campaign planning and managing. *Source:* authors' own elaboration

[13] Available at: [https://www.reuters.com/technology/meet-ashley-worlds-first-ai-powered-political-campaign-caller-2023-12-12]. Viewed: 3.5.2024.

[14] Available at: [https://www.npr.org/2024/02/08/1230052884/the-fcc-says-ai-voices-in-robocalls-are-illegal]. Viewed: 3.5.2024.

[15] "From the point of view of the citizenry, communication is increasingly personalized and, in this sense, advertising adapts its messages to specific groups of voters, making the broader public irrelevant in order to try to seduce the undecideds, who have been reached through social networks. In order to access information about these undecideds, it has been necessary to carry out [...] meticulous data collection in order to create the ideological profiles which form the basis of a personalized campaign. In addition, and once again, to the risks of this type of personalized campaigning for the individual rights of potential voters, data collection and the use of technology in election campaigns leads to the segmentation of the electorate into groups and the preparation of the most appropriate messages for each group

material (microtargeting). In India, in 2020, a candidate named Ashok Gehlot cloned his voice to send personalized messages to his voters via WhatsApp, addressing each recipient with their name.[16]

In addition, AI can help with the automated management of social-media profiles, groups, and channels by planning and scheduling the frequency and timing of posts and content. In addition, virtual assistants can help organize administrative tasks relating to campaigns, which can involve managing election agendas and calendars, scheduling payments, controlling the legality and documentation of spending, and appointing election observers. Finally, AI solutions may help parties and candidates to optimize fundraising campaigns and initiatives to identify potential new donors.

AI, in the electoral context, offers significant value by enabling the rapid and cost-effective production of communication content, optimizing resources, and facilitating the dissemination of key messages (Figure 2.5). Thanks to its ability to generate texts, images, and other formats in an agile manner, AI enables candidates and their teams to produce content in large volumes without incurring high costs or delays. This is especially useful in intense election periods, where speed of response and the ability to adapt to the information demands of the moment are crucial.

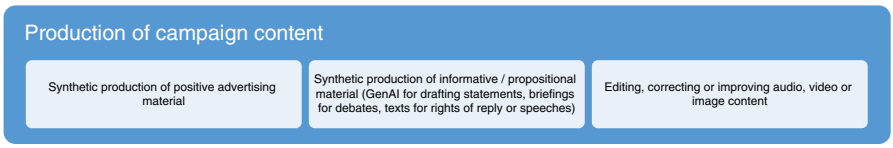

FIGURE 2.5 Production of campaign content. *Source:* authors' own elaboration

of voters, exclusively aimed at them. In such a course of action, it is implicit that candidates have multiple versions of themselves, which accommodate the diverse characteristics of the electorate, and which they try to monetize according to the benefits that can be obtained, concentrating resources on the pursuit of votes that can most easily be turned in their favor. This produces, on the one hand, an asymmetry in public information, with the same information not being accessible to all voters, and, on the other hand, a concentration of effort on the undecideds, converting election campaigns into a kind of 'political market', in which the price of the voter changes according to the profitability that the politician can obtain from us" (SÁNCHEZ MARTÍNEZ, 2020, p. 101).

[16] Available at: [https://www.aljazeera.com/news/2024/2/20/deepfake-democracy-behind-the-ai-trickery-shaping-indias-2024-elections]. Viewed: 3.5.2024.

Its benefit can be seen in the synthetic production of positive advertising material, with GenAI being used to produce jingles, slogans, cards, scripts, etc. In Mexico, the organization Sociedad Civil de México used AI to generate an image of a Starbucks coffee cup with the slogan of candidate Xóchitl Gálvez, which they used to ask the multinational for a commemorative cup with the slogan.[17] In Canada, one of the candidates for Toronto City Council used AI-generated images, which was subsequently uncovered when a woman with three arms appeared on one of his advertising posters.[18] There are also advantages in using AI to produce informative/propositional material in electoral contexts which does not necessarily have a propaganda nature, such as statements, briefings for debates, texts for rights of reply, or speeches.

AI systems have proven to be essential tools for editing, correcting, or enhancing audio, video, and image content. They can adjust lighting, framing, equalization, and pixel quality; capture memorable photographic moments; correct or eliminate defects; and remove noise, among other tasks. Additionally, AI can be used to apply aesthetic filters and backgrounds, improve speech fluency, and create "empathetic visuals," among other functionalities (Maranhão, 2024). However, aesthetic changes that could be decisive for voters' choice (such as de-aging or makeover techniques), according to some legislation, must be expressly disclosed. This is what happens in Brazil.

Another positive use can be seen in the use of AI to analyze data relating to electoral contexts, allowing players to gain a more accurate understanding of voter behavior and preferences, including their variations over the time. In this sense, AI tools can help monitor news coverage (clipping) to detect negative stories as well as help supervise social networks to automatically detect not favored content (criticism, rumors, and negative campaign material) (Figure 2.6). AI systems are also proving to be an important campaign tool by enabling the tracking of public groups on messaging apps, with a semantic analysis of impact indicators, trends, reactions, and sentiments can be determined. This makes it possible not only to visualize the agendas and issues

[17] Available at: [https://twitter.com/SocCivilMx/status/1769850004548559314]. Viewed: 3.5.2024. Starbucks, however, distanced itself from any kind of political affiliation. Available at: [https://twitter.com/StarbucksMex/status/1770224928635441366]. Viewed: 3.5.2024.

[18] Available at: [https://toronto.ctvnews.ca/three-armed-person-mistakenly-exposes-ai-generated-images-in-toronto-mayoral-platform-1.6439117]. Viewed: 3.5.2024.

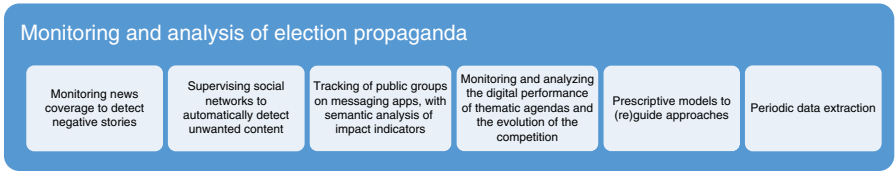

FIGURE 2.6 Monitoring and analysis of election propaganda. *Source:* authors' own elaboration

that have the greatest impact on how people vote, but also to understand how they feel about them.

AI can also be used to monitor and analyze the digital performance of thematic agendas and the evolution of the competition, which translates into relevant strategic data for campaign decisions. In this respect, AI can pinpoint the most impactful pieces of propaganda, public statements, and messages based on the visibility achieved or the nature of the repercussions in digital discussion forums. This provides relevant data, for example, on approaches that should be avoided and actions or statements that should be reinforced or repeated.

Synthetic intelligence, thus, can be used to use prescriptive models to (re)guide approaches (georeferenced mapping of hot spots with critical repudiation rates, computer programs to recommend emotional triggers, lexical terms capable of generating strategic commitment, etc.). Still, from a performance and trend monitoring perspective, there is the possibility of periodically extracting data for future use, including the selection of topics and agendas, in accordance with the legal framework for data protection.

2.2 ARTIFICIAL INTELLIGENCE AT ODDS WITH DEMOCRACY

However, particularly in settings with low levels of institutionalization and little adherence to the law, digital campaigns engage in practices which aim to secure competitive advantages through widespread non-compliance with the rules, distributing fake news on an industrial scale that is without precedent (Rubio Núñez, 2018, pp. 32–33; Sakamoto, 2016, p. 22; Schick, 2020, p. 11), exploiting algorithmic logic in unethical ways, fomenting cultural wars, and spreading destabilizing messages that are impossible to trace (Aguado Terrón; Villaplana Jiménez, 2023, p. 212; Moore, 2018, p. 96; Sánchez Muñoz, 2020, p. 57). All of this generates a significant deficit for electoral integrity and creates problems that ultimately challenge national

security and the continuity of democratic societies (Metaxas, 2020, p. 245; Schulte, 2020, p. 137).

Henry Kissinger, Eric Schmidt, and Daniel Huttenlocher, in an analogy with war technology, observe that highly connected societies are more exposed to this type of threat, since the digitization of sensitive processes has the adverse effect of expanding the "attack surface" vulnerable to enemy forces (Kissinger *et al.*, 2021, p. 147). Due to the lowering of barriers in terms of cost and experience (Giusti; Piras, 2021, p. 7), and the proliferation of easy-to-use AI tools, something similar is happening with the microsystem that protects electoral integrity protection.[19] This is because:

> the rapid development of information technology creates new possibilities for the malicious use of artificial intelligence as part of psychological and information warfare campaigns. This facilitates massive, rapid, disruptive and subversive campaigns designed to have a serious cognitive and emotional impact on international audiences, thereby representing a wide-ranging threat to organizational and national security.
>
> (Samoilenko; Suvorova, 2023, p. 508)

In addition to the traditional defense of freedom and equality against pressures, abuses, and perversions of justice on the part of political, economic, or media power, in the age of AI constitutional protection must also be concerned about informational abuses of various origins and types, especially those perpetrated in the virtual world and, as is increasingly the case, abuses powered by synthetic and algorithmic solutions. From this perspective, the landscape is also changing: invisible algorithms can determine

[19] "The year 2016 should have been a wake-up call. It made it clear that those who consciously sought to subvert the status quo using digital tools were much more successful than at any other time in the previous half century. For this reason, we should not view the three types of hackers who managed to distort the 2016 US elections – ndividuals, plutocrats and foreign states – should not be seen as anomalies, but as models of what the future will bring. Interpreting them as a model allows us to understand how they did what they did, what helped them to do it and how others can do the same, whether by using memes as weapons, collecting vast sets of voter data or developing sophisticated behavior selection methods. [...] None of the hackers would have been able to do what they did if politics had not migrated to the Internet" (Moore, 2018, p. 12–13).

access to information and the direction of public debate without anyone – and perhaps not even the programmers – realizing.[20] Furthermore, the massification of content generation tools increases the number of potential enemies ad infinitum since, in theory, the production and dissemination of *high-tech* falsehoods is within the grasp of practically anyone (SUYING, 2022, p. 30). In these circumstances, we must recognize that the number of potential attackers (malicious agents) has reached unprecedented levels (COECKELBERGH, 2022, p. 101),[21] because:

[20]It should be recalled that certain types of AI, especially those based on machine learning, produce unpredictable or inexplicable results, making it impossible to understand the reasoning behind some processes. This causes both astonishment and concern, as demonstrated by this episode recounted by Max Fisher: "Wojcicki's YouTube existed only to convert attention into money. Democracy and social cohesion were other people's problems." Shortly after he took office, Goodrow warned him: "We're not going to reach the viewing target if we don't take action." The action: giving more and more power to increasingly inscrutable AIs. In a 2016 article, Google engineers announced a "fundamental paradigm shift" toward a new type of machine learning that they called "deep learning." In the previous AI version, an automated system had built the programs that selected the videos. However, just like spam AI, humans supervised the system and intervened so that it could evolve and make adjustments. Now, deep learning was so sophisticated that it could take over this supervision. So, in most cases, "there will be no humans making algorithmic adjustments, measuring these adjustments and then implementing new corrections," wrote the director of an agency that generated talent for YouTube in a text that deciphered the article on deep learning. "In other words: when Youtube says that it has no way of knowing why the algorithm does what it does, the statement is probably literal. It would be as if Coca-Cola supplied an AI-designed drink to 1 000 million soft drink machines without a single human being checking the contents of the bottles, and if the AI that supplied the drinks was programmed solely to increase sales, without worrying about health or safety" (FISHER, 2022, p. 162–163).

[21]"One of the most important characteristics of the impact of new technology on election campaigns is the significant increase in the number of actors in the campaign, regardless of party. Communication is no longer centralized, with a single individual source (be it a politician, a party or a media outlet) communicating with a large audience of individuals; rather it is decentralized, with many individual sources communicating with the audience of individuals. Today, anyone can show their support for a particular candidate on the Internet, upload a video with critical content or send emails promoting a candidate, without any official connection to the campaign. However, these activities can have a huge impact on the final result, causing controversial qualitative changes. New actors, from civil society organizations to isolated individuals, can play a key role [...], buying ads to strengthen or weaken particular positions. They can operate without ties to the official campaign and even work beyond national borders. In this context, the responsibility of the platforms in systems of financing is becoming increasingly important [...] to ensure

> Artificial intelligence simplifies for anyone not only the creation of disinformation,[22] but also its dissemination. In this context, profitability is related not only to a reduction in effort and time, but also to a reduction in the "economic cost" of disinformation, which ultimately makes AI an especially dangerous tool for democracy when it is used in bad faith (*male fidei*).
>
> (Denemark, 2024, p. 123)

Applications such as ChatGPT make it easier and cheaper to manipulate text, audio, and video files, while the sophistication of synthetic products

---

transparency and accountability in the placement of ads, spending and attribution, so that citizens may be better informed about the context in which electoral decisions are made. With these new actors, anonymous profiles profiles permitted by the social platforms have emerged. The weight that interpersonal communication is acquiring through social networks has led to the massive creation of bots, anonymous, automated and sometimes fake accounts that act as individuals online and increase the mass distribution of specific information, with the aim of creating artificial currents of public opinion which accept or reject certain people or ideas. By giving the impression that they have widespread support, these resources create a bandwagon effect, and other people accept the ideas shared by this apparent majority. This generates a herd effect, whereby individuals neglect personal responsibility and submit to the will of the collective. In addition, the voter's decision-making process is hindered by the creation and mass dissemination of false information via fake profiles, which are often automated. Anonymity even makes it possible for candidates and parties to run unofficial campaigns, taking advantage of the freedom to operate outside electoral regulations, since they can present themselves as ordinary citizens or use false identities to achieve greater impact on the election campaign. These options have created new and serious challenges for existing political and financing systems" (Rubio, 2018; Venice Commission, 2020, pp. 7–8). For more information on the changes brought about by third-party activities in the election campaign, see Rubio, 2023.

[22] "Creating AI-generated images has never been easier. With popular and easily accessible tools [...], users can obtain images in line with their indications in a matter of seconds. However, AI platforms have established some parameters to limit the use of their products. DALL-E does not allow users to create images of real people, and Microsoft's equivalent prohibits 'fraudulent imitation'. Midjourney only mentions 'offensive or inflammatory images of famous people or public figures' as examples of content that would infringe its community guidelines. Violent and pornographic images are also banned on these platforms. Other tools allow a wider range of options, and some people are using them. The Spanish collective *United Unknown* describes itself as a group of 'visual warriors, creators of videos and images'. They use deepfakes to create satirical images, usually of politicians" (Adami, 2024).

makes it increasingly difficult to identify them (Murphy, 2024). In simplified terms, image-generating AI provides lay users with digital manipulation tools that were previously reserved solely for studios that specialized in film production and special effects, at zero cost and without any minimum skill or effort requirements (Peirano, 2019; Schick, 2020, p. 8).

After all, although AI possesses a "structural ambivalence" (Beiguelman, 2021, p. 114) that is markedly favorable to the pursuit of objectives compatible with the common good (Muñoz Vela, 2022, p. 65), the fact remains that its systemic risks demand that it be closely monitored by the institutions constitutionally responsible for the normalcy, correctness, and integrity of the processes used to select our elected leaders. This attention should, of course, be on the information disorders that stem from AI solutions.

Disinformation generated by AI has been classified by the World Economic Forum (2024, p. 8) as the "main emerging risk" over the next few years, which is explained, in general, by the fact that:

> "(a) Deepfakes (realistic but fake images, videos and audio generated by AI) will induce people to believe falsehoods or make them suspicious of all images and recordings, given the possibility that they could be fake.
>
> (b) Propagandists will use GenAI to create hyper-persuasive arguments that back up false statements (for example, "the elections were rigged").
>
> (c) AI will make it possible to automate disinformation campaigns. Propagandists will use efficient AI robots instead of populating their troll farms with human workers.
>
> (d) AI will enable very specific and personalized disinformation campaigns ('microtargeting')".
>
> (Williams 2023)

It is worth noting that Dan Williams, although he assumes the existence of "genuine dangers" in relation to the "effects of AI on the information ecosystem", considers that, "at least in terms of Western democracies, the alarmism surrounding this issue is based on popular but erroneous beliefs about human psychology, democracy and disinformation." For him, there are four "overlooked" factors which, "taken together, [...] indicate that some concerns about the effects of AI-based disinformation on democracies are greatly exaggerated." The factors in question are: "(a) online disinformation is not at the

root of modern political problems; (b) political persuasion is extremely difficult; (c) the media environment is highly competitive and driven by demand; and (d) the establishment will have access to more powerful forms of AI than anti-establishment sources." Among other arguments, Williams (2023) argues that the users most likely to believe disinformation already have an information ecosystem adapted to their needs, and that "there are few reasons to believe that AI-based disinformation will increase the size of this audience."

Alongside disinformation, we can identify other threats arising from the use of this technology, such as digitally-driven polarization and cyberattacks on critical infrastructure. In the context of election campaigns, AI solutions enable a wide range of undesirable behaviors that can have an intensely negative impact on the framework of basic rights and freedoms necessary for the organization of honest, free and fair elections.

In this vein, the harmful use of synthetic technology has the potential to erode the basic framework of political guarantees that includes, among others, structural imperatives linked to the protection of free and informed elections against the influence of false information, the defense of the right to compete in fair conditions, and the guarantee of electoral normalcy,[23] particularly with

[23] "Electoral normalcy, in principle, refers to the maintenance of regular elections, with the preservation of the elements indispensable to their fundamental purpose of 'processing conflicts peacefully' (Przeworski, 2020, p. 232). That said, democratic peace around elections depends on the natural absorption of popular judgment, both by the losing parties and by the minority fringe of the citizenry. From this perspective, any public behavior, posterior or anticipatory, that tends to stimulate, in a broad or sectorial way, global rejection of the electoral method as the instrument for legitimizing new representatives would directly undermine the postulate in question. The imposition of conditions for the acceptance of the results, the premature threat that the public's judgment will be rejected and the loquacious questioning of the results of the polls (with weak, vague or false arguments) are typical examples of authoritarian practices that violently assault electoral normalcy. [...] In this context, the concept of electoral normalcy, [...] implies the pure, firm and irrevocable recognition of the following minimum elements: a) the recognition of elections as the only valid method for converting popular will into representative mandates; b) prior, unconditional and uncompromising acceptance of electoral rules, including the mechanisms for collecting and counting votes; c) respect for the authority of the established arbiters and the election results, with possible complaints being presented exclusively through established legal channels, accompanied by proof of the allegations; d) the understanding that the contenders are legitimate opponents, capable of assuming office in the event that they were to be elected; e) the recognition of the citizenry's political dignity and right to choose; and f) the maintenance of elections in a peaceful and non-conflictive setting, excluding all forms (physical, verbal and psychological) of blackmail, threats or violence" (Alvim *et al.*, 2023, p. 345–38).

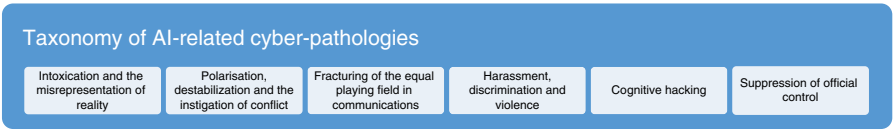

FIGURE 2.7 Taxonomy of AI-related cyber-pathologies. *Source:* authors' own elaboration

regard to respect for and acceptance of the result, which is a basic principle on which democracies establish themselves and take root (Levitsky; Ziblatt, 2023, p. 32).

In short, the malicious or antisocial use of AI in elections can give rise to systemic threats linked, for example, to the activities listed below, which are considered prima facie illicit because they conflict with the rules or principles that structure the electoral framework (Figure 2.7).

For the purposes of analysis, we propose grouping these behaviors according to their effects on democracy, establishing six different types of related actions that tend toward illegality and which are modalities of information disorders: (1) intoxication and the misrepresentation of reality; (2) polarization, destabilization, and the instigation of conflict; (3) fracturing of the equal playing field in communications; (4) harassment, discrimination, and violence; (5) cognitive hacking; and (6) suppression of official control.

2.2.1 Disinformation and the Manipulation of Reality

Infoxication and the *misrepresentation of reality* imply, firstly, the use of AI to automate the mass production and massive distribution of fake news, cheapfakes, deepfakes, and disinformation as well as the falsification of public opinion in virtual communities of any kind (astroturfing). At the same time, AI can obstruct access to factual reality through *information overload*, through mass dissemination tools (spreaders, spambots), and robot accounts used to spread misleading narratives (smokescreens) as well as intense propaganda campaigns (firehosing). Finally, *AI hallucinations,* although accidental, can also do significant damage to the quality of the information context.

2.2.1.1 Disinformation

Regardless of the topic, disinformation can take different forms, falling into 12 categories (Figure 2.8): (a) *fake news:* false news masquerading as journalism, disseminated with the purpose of spreading fictitious information

FIGURE 2.8 Taxonomy of disinformation. *Source:* authors' own elaboration (based on Alvim; Zilho; Carvalho, 2023) / Lumen Juris

or distorting the essential aspects of a real event; (b) *accidental disinformation:* false or decontextualized information that is shared due to ignorance, in an uneducated act of good faith; (c) *malicious disinformation*: false information disseminated in the knowledge that it is harmful and fraudulent; (d) *malicious decontextualization:* decontextualized information disseminated in the knowledge that it is harmful and fraudulent; (e) *numerical disinformation*: the fabrication, manipulation, or malicious interpretation of quantitative data which aims to fuel erroneous positions on controversial issues; (f) *informational degeneration*: information which is partially truthful but at times adulterated to induce error or inflict reputational damage; (g) *deepfakes:* the falsification of ultra-realistic audiovisual content to attribute reprehensible statements or behavior to specific political actors; (h) *discursive scene-setting*: controversial or damaging statements, fabricated through editing together snippets and erroneously attributed to institutions, authorities, relevant political actors, or third parties associated with them; (i) *conspiracy theories*: extravagant postulations entirely lacking in reliability (factual, legal, scientific, journalistic) and with damaging persuasive potential; (j) *rumors:* anonymous, controversial, and unconfirmed accusatory reports with damaging persuasive potential; (k) *strategic denialism*: the malicious planting of "doubts" based on false statements, premises, or data, with destabilizing or destructive objectives; (l) *visual disinformation*: graphic

content which conveys false or decontextualized messages with damaging potential (Alvim *et al.*, 2023, p. 144–145).

2.2.1.2 Infoxication

The term *infoxication* refers to the information overload to which Internet users are exposed on a daily basis. Information overload is one of the most effective forms of disinformation nowadays. To the extent that the excess of supply and noise makes it difficult to devote attention to, analyze effectively, and reflect on the information available, information intoxication represents a contemporary phenomenon which – in addition to inducing doubt, insecurity, and paralysis, and generating feelings of anxiety, incapacity and frustration (Loor-Cabal *et al.*, 2022, p. 358) – produces a proliferation of contradictory opinions that naturally "makes it difficult or even impossible to identify the objective truth" (Alvim *et al.*, 2023, p. 215).[24] As Carlos Bermejo illustrates:

> Information is like a river from which water has to be extracted. If the river is gigantic and flows at high speed, what we can assimilate from it will be very little, and so it can be said that the more information there is and the faster it is, the less we will know about the world.
>
> (Bermejo Barrera, 2021, p. 55)

In light of these issues, Marina Garcés (2017, pp. 75–76) states that, alongside the economics of attention, it is necessary to develop studies on the psychology and politics of attention. The former has to do with the pathologies produced by the same saturation of attention: anxiety, disorientation,

[24] Information intoxication, in this sense, is linked to the notion of infodemic, which "Designates the great flow of information that circulates through the network on a specific subject and that multiplies in an accelerated way in a short space of time, as occurred with the COVID-19 pandemic. The excess of unverified publications from unknown sources naturally damages the quality of public debate and makes it difficult to identify correct information, since it increases the costs of accessing appropriate information" (Cardiel Soto *et al.*, 2023, p. 148).

depression. The second is related to the political consequences and challenges that saturated attention generates. Basically, powerlessness and dependence. We cannot form an opinion about everything around us. The double limit of attention, the reception of data and information and its elaboration into opinions and knowledge, result in paralysis in the face of a disorienting scenario. An overwhelmed subjectivity is the one that today submits most easily to an uncritical adherence to the opinions, ideologies or judgments of others. Since we cannot form an opinion about everything around us, we follow those provided by others, already formed, and we trust them, without having the ability to subject them to criticism. Is this not the mechanism that Kant called heteronomy? The difference is that in other times heteronomy was based on ignorance of knowledge, on not having access to knowledge, whereas today it functions with its overwhelming accessibility and, therefore, inoperability. Each era and each society have their own forms of ignorance. From these their correlative forms of credulity are derived. Ours is an ignorance steeped in knowledge that cannot be digested or elaborated.

This excess of information leads to information inflation, leveling, and a consequent loss of authority on the part of traditional sources of information. This generates a type of selective "truth on demand" consumption of information: Facts that do not coincide with our point of view can be ignored. In such an ecosystem, the information cycle is shortened, as are the processes of verification, elaboration, and reflection. This reinforces the loss of the temporal horizon of politics, which is converted into an "omnipresent present" (Burguera; Cobacho, 2013) where every statement, every act, seems to start from zero without being determined by the past and without any obligation to the future. As Michael J. Sandel (2023, p.24) states:

> Facebook is free. And it inflicts damage on democracy. Its enormous power, free of regulation, allows external interference in our elections and the free dissemination, on an unprecedented scale, of hate speech, conspiracy theories, fake news and disinformation. These pernicious civic consequences are now recognized. Less obvious is the corrosive effect on our ability to concentrate. Hijacking our attention, collecting our personal information and selling it to companies that distribute advertisements tailored to our tastes is not only a threat to our privacy. It also corrodes the patient and attentive attitude towards the world that is necessary for democratic deliberation.

From a practical point of view, "info-saturation" (Dias, 2014, p. 8) can be stimulated by automated accounts and spam applications.[25] AI can also contribute to content development (text, image, audio, video) by generating media and news articles to feed the flow of messages in chat applications, social networks, video platforms, blogs, and digital vehicles with the assistance of "bot journalists" (Costa, 2021, p. 136), as revealed by an episode that took place in the United States in June 2022.[26]

These practices negatively impact the electoral environment, in particular because they jeopardize access to adequate and relevant information, which becomes blurred in a hectic scenario dominated by digital cacophony. From this perspective, Caldevilla explains that overexposure to a plethora of messages: (a) limits the visibility of important elements, and associated with this, (b) imposes a logic of selection and dismissal which is also harmful, because it creates a fragmented communication space, inhabited by biased and unchallenged minds molded by a self-referential logic (Caldevilla Domínguez, 2013, p. 35), which leads to decisions that are fragile and ill-informed.

On the other hand, the excess of information can jeopardize the attention received by false and harmful content, since low-quality messages also face much competition. In line with this view, some experts believe that the effects of deepfakes on voter behavior may not be as significant as imagined (Adami, 2024).

This inflationary universe is home, in some circumstances, to attractive and distracting narratives, which are generally used to distract the public from important issues, that foster a "change in the information cycle so that attention to the problem in question is relegated to the background" (Magallón Rosa, 2023, p. 123) for governments, actors, or political groups, thus reducing their negative exposure.

---

[25] In this vein, it should be noted that "a large part of the flow of false information is not organic or spontaneous, but maliciously plotted by a veritable fake news industry, operated by malicious agents, bots and fake profiles that take advantage of the digital platforms' models of omission and monetization. As a result, there would be no need to even consider restrictions on freedom of expression in these cases, which is another reason to justify the TSE's regulatory activity, including that which seeks to hold digital platforms responsible for propelling disinformation and making a lot of money from it" (Frazão, 2024).

[26] A network of fake journalists and political advisors was discovered, who used GenAI tools to create fake stories content, and even the (equally fake) photographs that illustrated their profiles. The fake faces, in this case, allowed these malicious agents to spread fraudulent content from fake accounts that simulated, in all respects, professional activity (Suying, 2022, p. 28).

> In the disinformation agenda, *smokescreens*[27] reduce the audience's field of vision, fulfilling the function of covering up unpopular or difficult to justify issues that may have an undesirable impact on the course of public dialog.
>
> (Cardiel Soto *et al.*, 2023, p. 61)

> Through their use, agents of communication manage to draw the community's attention elsewhere, investing in giving visibility to extravagant news items with impactful content which, due to the curiosity or uproar they generate, end up burying uncomfortable news.
>
> (Alvim *et al.*, 2023, pp. 214–215)

In the age of AI, *smokescreens* have received an additional boost with the appearance of content generation tools that are extremely useful for feeding the "scandal machine" (Strobl, 2022, p. 98). Sexual scandals, for example, are at the root of this unique agenda (Schick, 2020, p. 40; Suying, 2022, p. 25). These are especially effective not only for reputation destruction tactics based on narratives of inappropriate behavior – incorporated organically as an element to influence the vote in the context of the culture wars – but also to act as clarion calls for online activism, diluting the possibility of accessing (good) information as a precondition for the empowered exercise of citizenship.

Another way of influencing the formation of public opinion is through the use of *diffusion* applications, which are capable of automating the distribution of communication products across multiple platforms. With these tools, political content can achieve a much greater social reach in a few seconds than the visibility that would be standard if it were disseminated organically.

In many cases, amplification can go beyond automated distribution, being implemented not to amplify messages coming from a single source (usually a *figurehead*[28]), but rather to multiply the strength of narrative agendas promoted

[27] "Smokescreen [...] is an expression that is used when one wants to cover up one event with another; it is a figure, that is, a metaphor that is commonly used in the media and in politics, although its origin is on the battlefield. The term was first used in the military sphere to define something that takes the form of a burning straw, using fire as a tactic to obstruct the enemy's view" (Lifeder, quoted by Moreira Coelho; Castro Montenegro, 2023, p. 14).

[28] In specialist terminology, figureheads are "strongly ideologized actors who attract millions of followers and are responsible for proposing topics, slogans and frameworks, which they often expound first-hand. Adopting a posture of leadership and seeking out successive forms of agitation, they dictate orientations, trends and strategies, effectively guiding a

in different areas of cyberspace. This starts with a significant set of fake followers – such as ghost accounts, cyborgs, and robots – that together become an "explosive combination" capable of making voices in favor of fake news, conspiracy theories, discriminatory agendas, and hate speech (Mello, 2020, p. 153; Williams, 2021, p. 207) appear to resonate. In this hypothesis, synthetic intelligence helps to consolidate astroturfing operations and contributes to firehosing, which contaminates the virtual debate.

2.2.1.3 Fake Support (Astroturfing)

The notion of astroturfing serves to define the creation of "fake" popular movements (Mello, 2020, p. 28) capable of mobilizing networks by emulating apparently strong and massive public reactions but which are basically empty and unreal. Astroturfing therefore indicates influence campaigns that usurp or simulate virtual identities to add value and weight to particular causes or narratives (García-Orosa, 2021, p. 5), which then present themselves as majority causes backed by the solid numerical adhesion demonstrated by social network metrics. Within this logic, fake Internet users are created and mobilized to manipulate and deceive real people (Kovic *et al.*, 2018, p. 18), in many cases for the purpose of destabilizing and in contrary to the democratic agenda.[29] This includes actions in sensitive social processes such as wars and geopolitical or military conflicts.[30]

---

legion of users who replicate *en masse* the ideas sown on social networks and similar applications. In graphic language, the figurehead is the epicenter of micro-earthquakes of narrative crises, an influencer who openly and incisively uses their entourage to constantly oxygenate the living organism of disinformation" (Alvim *et al.*, 2023, p. 192).

[29] "That bots are a threat to democracy is clearly demonstrated by the data showing that far-right parties use them to promote their themes twice as much as the average of other parties. To analyze bot activity during the 2019 European elections, the British daily *The Independent* developed an algorithm that found that 12% of all Twitter tweets promoting and supporting far-right parties came from automated fake accounts: bots. In just two days, these accounts promoted Matteo Salvini's candidacy through more than 400 tweets. Moreover, the origin of most bots linked to far-right politicians can be traced back to Russia. In the same elections, according to the analysis, more than 6 700 bots reached some 241 million citizens, which represents approximately half of the entire population of the European Union" (Denemark, 2024, p. 127).

[30] Recent investigations have confirmed the presence of more than 40 000 inauthentic accounts on X which disseminated disinformation and propaganda posts seen by hundreds of millions of users on that platform just two days after the Hamas attack on Israel. Another example includes the presence of 600 000 automated Facebook accounts waiting to be deployed as well as the recent detection of 50 000 fake accounts conducting disinformation campaigns in Germany (Muñoz, 2024).

In a hyper-connected world (Arriagada, 2023), the majority of people develop opinions based on references found on the Internet, beyond their circles of trust. It is even common for news platforms and social networks to contain social buttons and comment sections through which users can express, support, and debate opinions. This hyper-connectivity conditions the dissemination of a piece of disinformation that is distributed massively thanks to the fact that people are hyper-connected, especially through interpersonal communication channels such as WhatsApp or Telegram. Moreover, hyper-connectivity facilitates and fuels the desire to maintain the coherence of a worldview with the group (Kahneman *et al.*, 2021), which can end up shaping political decisions in a way that is even more powerful than personal bias itself.

Bearing in mind that, from the perspective of mass psychology, an individual's beliefs are often influenced by the beliefs of others, a phenomenon which in the electoral context is known as the bandwagon effect, it can be deduced that astroturfing, by falsifying popular behavior, effectively performs the work of political persuasion (Zhang *et al.*, 2013, p. 2), given the interconnection between the principle of conformity and the power of social influences (Sunstein, 2020, p. 21).

In its traditional form, this phenomenon originated with the activation of automated or semi-automated fake accounts to manipulate engagement statistics (views, likes, and shares), and through more complex interventions that publish massive amounts of comments or posts strategically designed to influence public opinion[31] authored by bots,[32] cyborgs, or digital avatars. With the evolution of AI-based editing and infographic techniques,

[31] Within this logic: "[...] in the universe of campaigns, digital astroturfing can be seen as an artificial and deceitful strategic resource used by political actors who, from the top down, reproduce the mass behavior of autonomous individuals, creating currents of opinion that seem to emerge from the bottom up (Kovic *et al.*, 2018, p. 71). Closely related to some social adjustment phenomena, such as conformity bias (Sunstein, 2009, p. 15) and the herd effect (Cazorla, 2015, p. 36), astroturfing can present visible variations in a broader context when analyzed under dimensions that involve the actors involved, the chosen objects and the goals pursued" (Alvim *et al.*, 2023, p. 218).

[32] "A bot (apheresis of robot) is a computer program that automatically performs repetitive tasks on the Internet, tasks that would be impossible or very tedious for a person to carry out. Although there are different types of bots, in this context they would be fake profiles on social networks that are used to disseminate content and are even capable of interacting with other users in an automated way. These are called social bots" (Sánchez Muñoz, 2020, p. 35).

astroturfing can adopt new forms, such as the falsification of images or videos to inflate the number of people present at popular protests with the replication or digital insertion of *fake protesters*.

In 2023 Meta discovered thousands of accounts operated from China with stolen names and profile pictures that claimed to be Americans commenting on American politics and foreign affairs with the aim of "building an audience" and fomenting "polarization" (Bond, 2023).

Bots are "algorithms that, through AI and machine learning, learn to simulate human behavior for various purposes,[33] such as subscribing to the messages of a social profile and making its content viral ('follower bots'); using hashtags to obstruct debate ('roadblock bots'); and publishing positive comments about a political agenda or candidacy that imitate human language ('propaganda bots')" (Calvo *et al.*, 2019, p. 126), thus amplifying harmful and destabilizing narratives that foment confrontation (Caldarelli *et al.*, 2021, p. 157). In the manipulation industry, the importance of robots is confirmed because:

> Disinformation is most reliable when it is disseminated massively or when there is interaction with it, whether in the form of approval or disapproval. If disinformation is not shared or has no reactions (comments, likes, etc.), it means that it has no public reach and is therefore ineffective. To artificially create the impression that disinformation is being interacted with or shared from the outset (of its existence in the public arena), fake accounts come into play. In other words, bots are mainly used to amplify disinformation.
>
> (Denemark, 2024, p. 127)

[33] "Bots can be described as 'computer programs run by algorithms and designed to perform specific tasks online, such as analyzing and collecting data,' which means that they are often used by online service providers, such as the Google bot. However, these programs can also be 'created to publish content automatically, increase the number of followers, support political campaigns and spread disinformation.' The power of these bots lies in the fact that they are seen as real people, so that when a bot comments on or shares a piece of content, it appears that it has been commented on or shared by a living human being with autonomous thought. Therefore, bots help to spread disinformation both quantitatively (in the sense that there are many of them) and qualitatively (in the sense that, at first glance, they are unrecognizable and project themselves as real people)" (Denemark, 2024, p. 127).

Regarding social unrest, bots are used to exert more and better influence over public opinion.[34] In addition to creating content, "the bots' AI allows them to operate in a coordinated way, among themselves or with certain human users [...], to amplify or spread news, shift the focus of attention and increase the popularity of certain profiles." Through these activities, "they contribute to the strengthening of the digital presence of certain actors, as well as having an impact on the coverage that the press – especially the partisan press – gives to the major issues on the public agenda" (Shoai; López Molina, 2023, p. 253).

At election time, robots imitate "human behavior on social media platforms" and can be used "to amplify messages and spread false information," thus manipulating public debate and reinforcing the "sectarian bubbles of network feudalism" (Maher, 2022, p. 15). They can also be deployed "to create the illusion of widespread support for a specific candidate or issue" (Lins, 2023, p. 292–293), which contributes to inequality between political competitors (Calvo *et al.*, 2019, p. 126).[35]

On the other hand, cyborgs or sock puppets fulfill similar communicative functions to bots, with the simple difference that in this case the automation process is not absolute. From this perspective, sock puppets are fake digital identities driven partly by programming and partly by human intervention, through which members of a virtual community carry out attacks or support agendas pretending to be what they are not (Bu *et al.*, 2013, p. 366). Also known as hybrid bots, cyborgs therefore refer to semi-automated accounts.

[34] "In all the cases we have studied, the bots are directed by some social or political actor with the purpose of influencing human behavior. For example, one study identifies a large number of automated accounts around news portals that spread polarized messages and ideological attacks on certain actors in order to monetize the audience they build. Some authors have also associated bots with echo chambers, stating that they promote microsegmentation and the isolation of users on the network" (Shoai; López Molina, ibidem).

[35] In this outline: "The automation of political content during campaigns to achieve greater impact explains this depersonalization which is achieved, for example, with the creation on social networks [...] of fake profiles and bots whose participation in generating interactions aims, among other objectives, to share information to reach more people, create and amplify debates so that they acquire a greater level of relevance [...], increase a candidate's followers to make them appear more popular with voters, reduce criticism and also disable or intimidate political rivals" (Mateos Crespo, 2023, p. 245).

> [T]hat participate in cognitive warfare with a double advantage: insofar as they can avoid predictable behavior by acting in different ways and at different moments, they are *more convincing* and *impossible to track*. Whether for these reasons or because, due to the use of human capital, they tend to be more expensive, it is possible to argue, at least in principle, that cyborgs have an even greater negative juridical value than robots, given their greater ability to deepen the gaps between the different *actors*.
>
> (Alvim *et al.*, 2023, p. 188)

Mateos Crespo (2023, pp. 245–246) notes that during the most recent election, political parties openly acknowledged their awareness of disinformation practices while simultaneously denying their own involvement, instead attributing such activities to either opponents or unaffiliated supporters. However, he argues that this denial lacks credibility, citing evidence of coordinated actions between fake accounts and official campaign messaging, particularly demonstrated by the synchronized timing of cyber-activist operations and automated bot networks that aligned precisely with party strategies.

Numerous companies dedicated to disinformation operate in Israel, such as "Team Jorge," which used an automatic content creation system based on key words and distributed the content via fake accounts to promote negative narratives such as one accusing a former Mexican director of criminal investigations of being involved in torture and murder. Another example is the company Mind Force, which worked for the Angolan government and used numerous accounts to upload content in favor of the ruling party.[36] In Kyrgyzstan, political parties hired groups to spread propaganda using an AI-generated bot that created 150 fake profiles a day.[37] In Turkey in 2023, "bot armies" on X "bombarded" the platform with content in favor of President Erdogan and his party (Funk *et al.*, 2023). In China, AI companies must adhere to "core socialist values."[38] As a result, chatbots

[36] Available at: [https://freedomhouse.org/report/freedom-net/2023/repressive-power-artificial-intelligence]. Viewed: 3.5.2024.

[37] Available at: [https://freedomhouse.org/report/freedom-net/2023/repressive-power-artificial-intelligence]. Viewed: 3.5.2024.

[38] Available at: [https://www.theverge.com/2023/4/12/23680027/china-generative-ai-regulations-promote-socialism-chatgpt-alibaba-baidu?mc_cid=69b30ca9c0&mc_eid=66ec824c86]. Viewed: 4.5.2024.

created by Chinese companies do not give answers on sensitive issues such as Tiananmen Square and replicate the Communist Party of China's position on Taiwan. Other AIs, such as Alibaba and Baidu, are required to implement controls to ensure the "truth, objectivity and diversity" of the data they receive. Finally, in 2023, the Chinese Cyberspace Administration created five chatbots which are required to adhere to "essential socialist values" and delete content contrary to the interests of the Chinese Communist Party (Funk *et al.*, 2023).

A likely recent example of massive use of bots in digital influence activities points to the campaign of Pablo Marçal (a controversial coach and social-media influencer) for mayor in São Paulo.[39] The participation of bots, by the way, tends to be decisive not only in terms of the number of automated accounts themselves but also their posts. The biggest problem lies in the fact that these accounts, by nature, engage in infinitely more active behavior than ordinary users. After all, bots do not sleep, do not rest, and are created precisely to function as super-spreaders and super-posters.

2.2.1.4 The Hose Effect (Firehosing)

Both bots and sock puppets can be used to carry out intensive campaigns aimed at saturating and contaminating the communication environment with political information through the massive and simultaneous distribution of low-quality messages and content. Firehosing, in this context, is the expression used to graphically describe the hose of online falsehoods which materializes in the context of campaigns as a "common practice." (See, for example, the recent elections in India, Indonesia, Slovakia, Mexico, Brazil, and the United States.[40]) A voluminous mass of users and channels come together, in a synchronized and repetitive attempt to fill networks with an excess of low-quality messages that jeopardize access to factual reality (Alvim *et al.*, 2023,

[39] Available at: [https://noticias.uol.com.br/colunas/radar-das-eleicoes/2024/08/27/automacao-entenda-como-marcal-faz-para-todo-mundo-falar-dele-nas-redes.htm]. Viewed: 11.1.2024.

[40] In the 2016 US elections, 115 false pro-Trump and 46 pro-Clinton narratives were detected in just a few weeks, which were shared more than 38 million times on social networks. This content originated with or was amplified by fake news sites, which together had more than 159 million visits between 8 October and 8 November of the year in question (O'Connor; Weatherall, 2019, p. 4).

p. 215),[41] to inject dishonest opinions that contaminate the public imagination through "cascades of information." As Sunstein (2020, p. 53) points out:

> In an information cascade, people at some point stop trusting their information or their personal opinions. Rather, they decide based on the signals that others transmit. When this happens, the subsequent statements or actions of a few or many others no longer add any new information. People simply follow their predecessors. [...] Because so many people have done or said something – a politician is great, a product is dangerous or someone is a criminal – people think to themselves: 'How could they all be wrong?' The truth is that yes, they could be, all the more so if they are reacting to what others have said or done, and in this way they are amplifying the volume of a signal that influenced them.

2.2.1.5 The Misrepresentation of Reality

In this case, actors disseminate a symphony of rumors, conspiracy theories, and false or decontextualized news stories (Aminulloh *et al.*, 2022, p. 249; Salinas Olarte, 2023, pp. 334–338) on various channels and platforms, thus "taking advantage of an environment [...] conducive to the continuous and shameless propagation of partial truths and absolute fictions." The aim of this strategy is not so much to consolidate a "clear and dominant narrative," but to "sow distrust" among the audience, which usually happens when "lies come one after the other rapidly, repeatedly, and constantly" (Andrejevic, 2020, p. 23) to become "primary political factors" (Merenda, 2021, p. 26) on the electoral scene.

It is often cautioned that these tactics and tools have consolidated a structural environment of "informational corruption," among other reasons because they allow the massive, constant, and rapid circulation of contaminated statements. Within this flow, rumors and conspiracy theories circulate freely along with malicious claims, and messages and media that have been

[41] "Flooding the public with information; producing distractions to dilute attention and concentration by delegitimizing the sections of the press that provide accurate information; deliberately sowing confusion, fear and doubt; creating rumors or claiming that certain information is a hoax; and inciting persecutory campaigns aimed at obstructing the functioning of reliable channels of information such as the traditional press [....] are all methods used by those who want to implement their agendas and pervert public debate, according to Tüfekçi" (Prado, 2023, p. 170).

falsified to cloud reality are partially hidden behind a dense veil of inconsistent statements that weaken consciousness and complicate our understanding of the world. This compound of inaccurate statements makes up the complex of *information disorders*, which is a genre made up of three related sub-groups: (a) *misinformation* (false information transmitted accidentally, without the intention of causing harm); (b) *malinformation* (genuine information shared with the intent to cause harm, stemming from the exposure of private matters that should not enter the public sphere); and (c) *disinformation* (false information knowingly shared to do harm) (Wardle; Derakshan, 2017, p. 5).

Disinformation does not only include falsehoods that go viral due to the ignorant and unconscious actions of Internet users, but also – and in an especially impactful way – due to negligence or bad practice on the part of journalists. For example, the publication of stories without adequate checks (the confirmation of sources and facts).

AI-based journalism is a consolidated and rapidly expanding reality. In some cases, such as the Chinese portal Toutiao, algorithmic editors "trawl the Internet in search of content, using natural language processing and computer vision to collect articles and videos from a wide network of associated sites and contributors on to order basis. Next [...] the prior behavior of its users – clicks, reading, views, comments, etc. – is used to organize a highly personalized newsfeed, adapted to each person's interests. The application's algorithms even write headlines to optimize clicks" (Lee, 2021, pp. 133–134). In the Western world, the use of technology that speeds up the creation of content to cut the time needed to edit, collect, and analyze data and produce textual or audiovisual pieces of information has revolutionized the work of the print media and radio and television broadcasters (Fabbro; Pittaro, 2023, p. 113), providing support for an aspect of "algorithmic journalism" (Dörr, 2015) already present in the routines of very traditional organizations, such as the New York Times (Torrecillas Lacave; Fernández Martínez, 2023, p. 82), the Press Association (Coeckelbergh, 2022, p. 89), CNN, the British Broadcasting Corporation (BBC), The Guardian, and The Wall Street Journal (Pérez-Seijo; Vaz-Àlvarez, 2023, p. 63), among others.

In an era marked by the progressive incorporation of automated tools into the industry of news production, cases of disinformation can derive from phenomena known as *AI hallucinations*, which are no more than *responses* provided by generative tools that, although based on logical initial information, lead to incorrect (but convincing) conclusions, usually due to insufficient data or flaws in the training models (Athaluri

et al., 2023, p. 4).[42] Hallucinations are an "unwanted phenomenon in which linguistic models generate a meaningless text or one that does not faithfully reflect the original data provided" (Ji, cited by Oliveira; Siqueira, 2023, p. 117). This problem contributes to the environment of misleading sources within the general spectrum of information disorders. A similar problem arises with *deepfakes*. As occurs with the amateur public, deepfakes can also deceive professional journalists and be absorbed as reliable news sources by the media and communication services (Maher, 2022, p. 1).

As for the potential impact of the information derived from models based on GenAI, it should be noted that in February 2024 an X publication went viral in India, in which it was reported that Google's Gemini tool, when asked, replied that Prime Minister Narendra Modi had been "accused of applying policies that some experts have called fascist." The tool added that these accusations "are based on various factors, such as the BJP's Hindu nationalist ideology, its repression of dissent and its use against religious minorities." On the other hand, according to the same publication, when similar questions were asked of other important figures, such as Donald Trump and Xi Jinping, there were no clear answers (Mukherjee, 2024).

According to Guadián (2024), an AI Forensics study revealed that Microsoft's Bing chatbot demonstrated significant inaccuracies in electoral information during the German and Swiss elections, with 30% of basic electoral queries receiving incorrect responses. The analysis showed that the chatbot provided inaccurate information about candidates, polls, scandals, and voting data, while also misattributing sources, with errors becoming more prevalent in non-English queries.

AI also affects cybersecurity by making communications more vulnerable, heightening the threat of cybercriminals who, without time constraints and

[42] "A hallucination in AI is produced when artificial intelligence provides an assured answer, but one that cannot be explained by the training data, which leads it to make independent statements. According to the content generated by hallucinations is often meaningless or incorrect and can be classified into intrinsic and extrinsic hallucinations. In intrinsic hallucinations, the outcome (response) generated contradicts the original content. As an example, when asked about the approval of the Ebola vaccine: The generated summary 'The first Ebola vaccine was approved in 2021' contradicts the original content 'The first Ebola vaccine was approved by the FDA in 2019.' Extrinsic hallucinations, on the other hand, refer to the generation of an output that cannot be verified using the original source content" (Catalano; Lorenzi, 2023, p. 44).

with new capabilities, can compromise private communications for later use during the campaign, usually to damage rival candidates.[43] Such "information harmful to individuals, groups, organizations or states" – which forms the basis of *malinformation* (Hussain; Soomro, 2023, p. 28) – can be acquired through

[43] "The difference branches of Russian intelligence systematically penetrated the computer networks of the two major US political parties, their candidates and related organizations. A year before the elections, the FBI had alerted the Democratic Committee of the need to review its networks in view of the possibility that they had been infiltrated by a hostile actor. The party used to the services of the cybersecurity company CrowdStrike (Alperovitch, 2016), which was able to verify how the campaign's digital information had been compromised by at least two different groups of hackers, who had accessed thousands of files and emails from Democratic staff. According to the investigation, the digital footprint led to groups linked to the Russian state, which had used identical methods in previous operations of fraudulent information appropriation and leaking. The company established that the two groups were linked to two different branches of Russian intelligence: the group known as Fancy Bear was linked to the GRU (military intelligence) and the group known as Cozy Bear was linked to the FSB (civilian intelligence). Using very rudimentary identity-supplanting procedures ('social engineering'), the hackers managed to access the digital communications of a considerable number of Senator Hillary Clinton's campaign staff, including her closest circle. Both groups had pursued the same objective without apparent coordination and in ignorance of each other's activities, which highlights the importance that Russia attached to this type of operation, to the point of doubling the burden on different intelligence services. The scandal of Russian infiltration reached the country's mainstream media. Just a day after the publication of this news, a hacker who called themself Guccifer 2.0 published an allegation on the Internet in which they claimed to be responsible for the appropriation and filtering of Democratic Party documents, which they described as a 'personal project' with no connection to any state. To back up this claim, they published a new series of documents from the party, although they pointed out that most of them had been handed over to the Wikileaks portal. Guccifer showed great insistence when it came to discrediting the Russian role in these leaks, and even offered to interact digitally with journalists who wanted to obtain more details. These contacts, far from yielding the results desired by Guccifer, showed the inconsistencies and weaknesses of their claims, and were discredited by the US authorities as a campaign orchestrated by the Russian secret services themselves to generate confusion about responsibility for the hack on the Democratic Committee. The controversy over the origin of the infiltration was not an obstacle to the content of the leaked documents continuing to fuel electoral controversy. The Republican candidate and the media associated with his campaign did not hesitate to continue using these revelations to launch attacks on Hillary's honesty. Donald Trump himself would declare in an election rally: 'I love Wikileaks', without showing much concern about how these documents had arrived to the portal of the controversial Julian Assange" (Torres Soriano, 2017, pp. 7–10).

hacking, since AI reduces the entry barriers for cyberattacks by increasing the likelihood of attempts to hack cloud storage and email accounts as well as the usurpation of groups, communities, or profiles to (a) falsify identities, (b) obtain damaging information for future use, or (c) strengthen extremist agendas or foreign interference (Small, cited by Eddy, 2024; Vlachos, 2022).

In their connection with *leaks*, hacking, and the dissemination of private information (Balkan; Ülgen, 2023, p. 8), *disinformation* practices often also nourish (a) degrading sexist campaigns, closely tied to digital misogyny and political gender violence (Ochoa, 2023) as well as (b) insurrectional and *coup d'etat* strategies aimed at discrediting institutions and opponents by exposing their private lives. Furthermore, the specter of *malinformation,* in another context, represents the threat of degrading exposure as a tool to force political resignations, giving rise to low-cost blackmail that can emanate not only from information about the candidates themselves but also from people close to them (usually their own family).

In India, a significant incident demonstrated the malicious potential of AI manipulation when doctored images were deployed to discredit protesters who had accused a prominent political figure of sexual assault. The target of these protests was the president of the India Freedom Fighters Federation, who also held membership in the ruling BJP party. The AI-manipulated images were specifically designed to undermine the credibility of the protesters and their serious allegations (Adami, 2024).

In parallel, *disinformation* is intensifying with new synthetic AI technology, both in terms of the effectiveness of falsification techniques and their respective levels of sophistication and distribution capacity. As we have seen, AI can generate fake content and also automate the appearance of coordinated networks of inauthentic behavior, structuring "disinformation campaigns articulated through fake news media, and convincing profiles on social networks and other media, with the aim of sowing discord and undermining public confidence in the electoral process" (Eddy, 2024). However, it should be noted that these campaigns can have broader objectives. From Ingrid Bicu's (2024) perspective, the practices in question, depending on the circumstances, seek to: (a) maintain or obtain political power; (b) undermine social confidence in electoral results; (c) replace ineffective governments; (d) neutralize individual objectives; (e) obtain financial gains; (f) strengthen ideological convictions; (g) generate social disruption; and (h) influence the geopolitical landscape. The orchestration of illegal disinformation campaigns operates through a sophisticated manipulation of social-media ecosystems where fake news and fictitious profiles work in concert to achieve unprecedented levels of persuasive influence. These coordinated efforts create

profound cognitive dissonance among voters, systematically exacerbate political polarization, and erode public trust in electoral processes. The resulting damage extends beyond immediate electoral outcomes to fundamentally undermine the integrity of democratic institutions and the social fabric that sustains them. In this regard, it should be noted that:

> Before the explosion of GenAI, cyber-advertising companies all over the world had to write fraudulent messages and employ human troll factories to attack people on a large scale. With the help of technology, the process of generating fraudulent account holders is automated and becomes a weapon that requires only minimal human intervention. For example, microtargeting, the practice of directing messages to people based on digital tracking data such as their 'likes' on Facebook, was a cause for concern in the last US elections, although its main obstacle is the need to generate hundreds of variants of the same content to see which works with a specific group of people. What used to require a lot of manpower and was expensive, is now cheap and easy to achieve, with no barriers to entry. AI has democratized the creation of disinformation.
>
> (Van der Linden, 2024)

2.2.1.6 Superficial Falsifications (Cheapfakes, Shallowfakes)

At first, technology offers malicious agents the possibility of spreading false information through *cheapfakes (*also known as *shallowfakes)* which are, in short, cheap or superficial falsifications, in contrast to the concept of "deep falsification" that underlies deepfakes.[44] Although superficial falsehoods may be based on less sophisticated computing technologies, their results can be distributed along with AI solutions and algorithms.

Cheapfakes derive from unsophisticated adjustments made with simple and unprofessional tools, manifesting themselves, for example, through tampering with voices, images, or gestures using montages, tricks, and the contamination of subtitles. They therefore consist of "sound or audiovisual manipulation with small adjustments that expressly alter the meaning of the original material" (Maher, 2022, p. 7).

[44] According to Sean Maher, cheapfakes differ from deepfakes in the way they are produced and, above all, in their inherent capacity to create confusion, since the latter, due to their quality and level of representation, simulate the role of a reliable information source more effectively (Maher, 2022, p. 1).

An example of a viral cheapfake is a video produced by the nationalist BJP to sabotage the image of the leader of the main opposition party, Amit Malviya, at the start of his march across the country in 2023. On that occasion, various false reports were spread, from accusations that the flag of Pakistan (India's historical enemy) was going to be displayed during his march, to a video improperly linked to his campaign in which Gandhi was defiled with a sexually explicit song. It was later discovered to be a montage, with the original music taken from a Bollywood film (Mukherjee, 2024).

In this respect, Denemark (2024, p. 127) points out that "although deepfakes have a high level of quality in terms of resolution (clear images, steady voices), shallowfakes are still flooding the online space." They warrant our attention due to the fact that "they are even more difficult to prohibit," since "they may not be completely fabricated," as in the case of simply slowing down audio to simulate candidates with pronunciation problems and make it appear as if they were drunk, drugged, or senile, or as if they had low intelligence, a low reasoning ability, or deficient mental development. Specifically:

> Shallowfakes consist of manipulated images, texts or videos that do not use deep learning algorithms. As a result, they do not require the intense skill set and large volume of data that deepfakes do. Shallow fakes are more common in the information environment and provide a more basic method of influencing opinion. [...]
>
> The videos and photos edited in this context are created with little or no technology. They often decontextualize events and can range from crudely edited videos to elegantly coded images [...]. The manipulation of superficial falsifications using non-AI tools includes the deliberate slowing down or speeding up of audio or video in order to alter the real meaning and engage viewers. Splicing different parts of a video clip can drastically alter the original intention, as can editing out its beginning or end.
>
> (Suying, 2022, p. 32)[45]

[45] "Superficial falsifications, even if falsified and manipulated with simple editing tools, have an undoubted impact and should not be underestimated, especially when false information is used as a weapon to spread ethnic, religious or racial prejudice. These falsifications amplify existing tensions and fractures, which makes them ideal for misuse and for use in malicious and covert operations to spread disinformation, fostering misinterpretations of the original events by exploiting fears and prejudices. In short, although they are

This is the case, for example, with a video manipulated with the aim of jeopardizing the candidacy of Nancy Pelosi, Speaker of the US House of Representatives, which was spread to create the false impression that she was intoxicated or ill (DENEMARK, ibidem) as well as with an episode in which the speed of speech was reduced to suggest that President Lula had given a speech while drunk at an event in Bahia (UOL CONFERE, 2023).

However, the dilemma of digital disinformation on the web has worsened with the appearance of GenAI applications which facilitate the creation of synthetic communication falsehoods. From a technical point of view, generative tools work in the form of "generative pre-trained transformers," hence the acronym GPT (generative pre-trained transformer). Basically, these linguistic models offer texts similar to those produced by humans (*text to text*), as well as audios (*text to audio*), images (*text to image*), and videos (*text to video*) in response to the *indications* introduced by users. Within this logic, a GPT tool is capable of satisfying content demands more or less along these lines:

> [G]iven a topic sentence, it can produce possible paragraphs; given a question, it can give possible answers; given a topic and certain background information, it can write a possible essay; given a dialog, it can offer the transcript of a possible conversation. You can do all this on any topic about which there is information on the Internet. It is trained for the task by consuming this information.
>
> (KISSINGER *et al.*, 2021, p. 20–21)

GenAI tools reduce the economic and creative costs involved in fabricating false narratives, since they construct convincing stories, coherent theories, and complex arguments – even backed up by audio and images – on demand, based on brief instructions given in prompts *(*commands). With the reduction of barriers, GenAI is destined to become "the greatest tool for spreading disinformation that has ever existed" (CROVITZ, quoted by HSU; THOMPSON, 2023).[46]

---

not at the same technological level, shallowfakes can be just as effective and damaging as deepfakes" (SUYING, 2022, p. 33).

[46] A report by the International Panel on the Information Environment carried out a comprehensive analysis of the use of GenIA in national elections, documenting 215 incidents in 50 different countries throughout 2024. According to the study, the main uses found involve: (a) content creation (audio, video, image, text); (b) dissemination on social networks; and (c) to a lesser extent, hypertargeting of political messages. The majority of

2.2.1.7 Deepfakes

The concerns raised by the spread of deepfakes are, of course, no less alarming. The term "deepfake"[47] does not refer to synthetic content in general, but to video, image, or audio files manipulated by an AI tool that allows faces to be exchanged or voices to be modified, making false and unfavorable content seem real to those who receive it (GARRIGA *et al.*, 2024, p. 176; MAHER, 2022, p. 1).

From a technical standpoint, synthetic content generated by GenAI represents pure artificial creation that emerges independently of pre-existing material reality. In contrast, deepfakes operate through layered manipulation, where AI systems modify existing audiovisual content by systematically altering or augmenting specific features of sound and image (MAHER, 2022, p. 1). However, a broader perspective defines deepfakes as "any synthetic media – whether audio, video, or image – either manipulated or wholly generated through artificial intelligence," primarily deployed for disinformation purposes (SCHICK, 2020, pp. 8–9).

Murphy argues that the emergence of GenAI has fundamentally transformed deepfake production through its multimodal capabilities in combining text, images, audio, and video. He emphasizes how this technology has democratized sophisticated media creation, making it accessible to virtually anyone. As evidence, he points to companies like HeyGen and Synthesia, which provide affordable AI avatar video services in a market valued at

---

incidents (90%) are related to content creation, with a large number of episodes involving deepfakes and messages manipulated to mislead voters. More than two thirds of the cases were classified as having negative impacts. In 46% of the episodes, it was not possible to clearly identify the author/responsible party, which highlights the difficulty of tracing the origin of these actions and the risks to electoral integrity. Against this backdrop, the report proposes a set of recommendations aimed at mitigating the negative impacts of GenIA and subjecting it to democratic regulation. The measures include: (1) strengthening accountability and transparency; (2) guaranteeing due process for users; (3) clear rules for illicit or harmful content; (4) reviewing national electoral norms; (5) promoting media and digital education; (6) supporting research and monitoring; and (7) adopting international ethical standards (International Panel on the Information Environment IPIE, 2025).

[47] "The term 'deepfake' was first coined by a user called 'Deep Fakes' on Reddit, using a combination of the terms 'deep learning' and 'fake'. The user uploaded videos in which the faces of popular celebrities, such as Gal Gadot, were replaced by the bodies of pornographic actors and published these online in the 'r/deep_fakes' subreddit. [...] Trained with images of celebrities, the algorithm learned to replace the faces in each frame, generating convincing videos. Although this user was finally banned by Reddit, face-swapping techniques were freely shared online and have been introduced on the main social media platforms" (SUYING, 2022, p. 25).

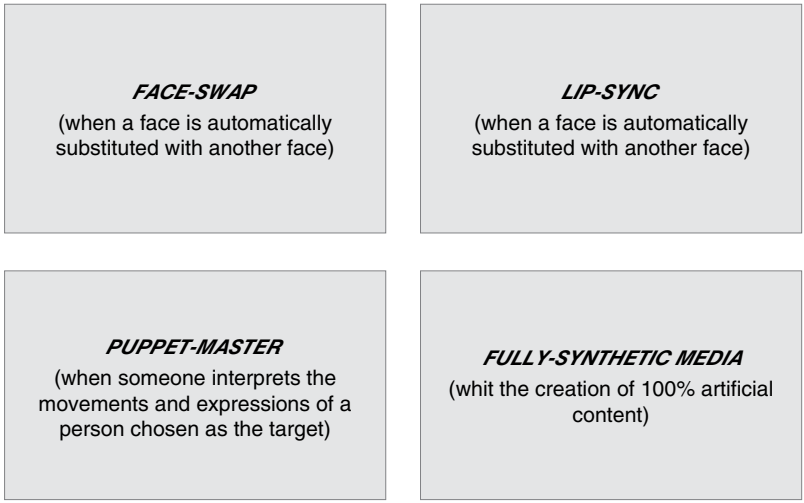

FIGURE 2.9 Main categories of video deepfakes. *Source:* authors' own elaboration based on SUYING, 2022, p. 24, with the addition of the last category

approximately $500 million by 2022. Furthermore, he notes that this same technology enables the generation of news content, the automation of bot networks, and the creation of chatbots designed to evoke specific political responses from users. These tools are used for different reasons: in some cases, in campaigns to influence voters' opinions or discredit candidates; in others, to undermine confidence in democracy as part of broader political objectives; there are also campaigns aimed at generating engagement and benefits (MURPHY, 2024).

Deepfake manifestations typically emerge in four distinct categories, each with its own characteristics and potential for manipulation that are outlined in Figure 2.9.

In other words, deepfakes consist of digitally adulterated media that "portray individuals saying and doing things they never did or said" (DENEMARK, 2024, p. 126). They are particularly dangerous because the manipulation, due to its meticulousness, easily disarms the skepticism of those who receive the information (SARLET; SIQUEIRA, 2022, p. 172). Thus:

> Deepfakes represent the greatest threat of this generation in terms of disinformation. From now on, images, videos and audio can be generated or manipulated by AI, using increasingly cheap, manageable and effective technology. Everyone has the power to show people in

> places they have never been to, doing things they have never done or saying things they have never said. In the wrong hands, this technology poses a serious threat to an already corroded information ecosystem, and to the way in which we understand and move through the world.
>
> (Schick, 2020, p. 20)

It is worth noting that, as we have seen previously, audio falsifications have already shaken up the scenario in different elections. Along with those already cited, we can mention how in the municipal elections in October 2024 in Brazil, presumed audio falsifications based on AI jeopardized precandidacies in the states of Sergipe, Amazonas, and Rio Grande do Sul (Góes, 2024).

The effectiveness of deepfakes has increased with the development of generative adversarial networks (GANs), which basically function as neuronal modules that compete with each other in a scheme in which one network (identifier) trains itself to detect real images and videos, and the other network (generator) continues to learn over time to generate better results that are more capable of circumventing or deceiving the detection algorithm (Suying, 2022, pp. 24–25).

For more than a century, we have been accustomed to understanding photographic, audio, or video media as the strongest guarantee of reality, as is graphically reflected in the popular expression "a picture is worth a thousand words," basically because these materials reliably recorded events that actually took place in time and space. Today, however, photorealistic media are generated by deep learning networks without any connection to the concrete world (Filimowicz, 2022). Thus begins an era in which any assertion of truth can be corroborated by absolutely false and almost irresistible (audio or visual)[48] proof. As for audio deepfakes, experts point to an even greater risk of information intoxication, since sound deepfakes "have a

[48] The strength of deepfakes has to do with the fact that, in general, people tend to trust their intuition and their associative processes more than their critical/cognitive processes. This theory was confirmed by a study centered on the perception of visual media which found that people who were presented with images were more convinced of the veracity of fake news than those who received it without visual support. Naturally, this phenomenon tends to magnify when the images arrive in video format or when the news is corroborated by voice recordings. In these cases, it is likely that people believe what they see or hear almost unconsciously (Denemark, 2024, p. 126).

higher quality" than their video equivalents and are also less traceable, since, compared to fake videos, they have less contextual information that can be used for verification purposes. In places where messaging platforms are used by a huge portion of the population and where there is a cultural tendency to send audio messages through them, deepfakes have become a "powerful tool in the information war during polarized electoral contests" (BIONI *et al.*, 2024). In this context, "deepfakes have a more direct effect on the audience's psychology than other types of media. The form of human cognition predisposes us to be influenced by visual evidence, and this is exactly what happens when we watch videos or audios whose quality prevents the artificiality from being easily and accurately deciphered by our ears or eyes" (SUYING, 2022, p. 27).

Deepfakes, therefore, "can make disinformation campaigns even more disturbing," along with their potential use to wrongly attribute reproachable words or actions to relevant *actors*. As they become increasingly convincing, these fabrications – derived from facial recognition AI – "can create particularly dangerous forms of fake news" and be deployed on the days or even hours before the polls open to influence votes or "sow distrust and confusion" (BENDER, 2022, p. 27).[49]

There have been some recent examples of this: in 2024, a fake BBC page promoting scams emerged from an AI-generated video of Rishi Sunak;[50] in Ukraine, an AI-generated fake video of President Zelensky asking his army to give up their weapons and surrender to the Russian army was distributed by the hacking of a Ukrainian news portal;[51] in Venezuela, government-controlled media broadcast videos generated by AI contained news

[49] "For example, in 2021 Russia was accused of using deepfakes to deceive senior EU officials and obtain information about a Russian opposition movement. In addition to impersonating public officials, deepfakes can also be used to create false public statements and influence or even disrupt electoral processes. For example, states that rely on independent commissions to redraw the boundaries of electoral districts often allow citizens to testify about how they would like the maps to be drawn. Fraudulent testimonies have been used in the past and will be increasingly convincing with the advance of this technology" (BENDER, ibid).

[50] Available at: [https://www.theguardian.com/technology/2024/jan/12/deepfake-video-adverts-sunak-facebook-alarm-ai-risk-election]. Viewed: 3.5.2024.

[51] Available at: [https://www.elconfidencial.com/tecnologia/novaceno/2022-03-17/hackers-rusos-difunden-un-video-falso-de-zelensky-ordenando-la-rendicion_3393225 surrendering]. Viewed: 3.5.2024.

presenters from non-existent international channels;[52] in the United States, a voice that simulated that of Joe Biden encouraged non-participation in the New Hampshire Democratic primaries;[53] in Mexico City, a running candidate leaked audios to various friendly journalists that were detrimental to his rival, who denied that these were genuine and asserted that they had been produced with AI;[54] in India, an anti-BJP message delivered by a Bollywood star was simulated that criticized Prime Minister Modi;[55] in Pakistan, an AI-generated video of Trump promising his support for former prime minister Imran Khan was also distributed;[56] also in Pakistan, during the 2024 election, an invented video of the vice-president of the PTI party, Sher Azfal Khan Marwat, calling for a boycott of the election was distributed;[57] in South Africa the world famous rapper Eminem was simulated giving his support to the opposition EFF party;[58] in Indonesia, a fake audio pretending to be a discussion between two leaders of the same political party was also distributed;[59] in Slovakia, a few hours before the first round of the 2023

[52] Available at: [https://www.technologyreview.es/s/16125/ia-generativa-censura-y-microinfluencers-marcaran-las-elecciones-de-2024]. Viewed: 3.5.2024.

[53] During the 2024 US primaries, automated phone calls urging people to abstain from voting in New Hampshire were reported that used cloned voices of celebrities and President Biden himself (Elliot; Kelly, 2024). Available at: [https://www.washingtonpost.com/technology/2024/02/06/nh-robocalls-ai-biden]. Viewed: 4.5.2024.

[54] Available at: [https://elpais.com/mexico/2023-11-02/un-polemico-audio-con-la-voz-de-marti-batres-echa-mas-lena-al-fuego-en-la-carrera-por-la-candidatura-de-ciudad-de-mexico.html]. Viewed: 4.5.2024.

[55] Available at: [https://economictimes.indiatimes.com/news/elections/lok-sabha/india/deepfakes-of-bollywood-stars-spark-worries-of-ai-meddling-in-india-election/articleshow/109487075.cms?from=mdr]. Viewed: 4.5.2024.

[56] Available at: [https://newschecker.in/fact-check/ai-generated-video-shared-to-show-donald-trump-supporting-pakistans-imran-khan]. Viewed: 4.5.2024.

[57] The false media made reference to another fake video, broadcasted the previous day, in which the former chief of staff Imran Khan appeared calling for a similar boycott. The false nature of the video was verified by the Soch Fact Check agency (Soch, 2024).

[58] Available at: [https://africacheck.org/fact-checks/meta-programme-fact-checks/will-real-slim-shady-please-stand-eminem-video-endorsing]. Viewed: 4.5.2024.

[59] Available at: [https://www.cnnindonesia.com/nasional/20240124064555-617-1053546/anies-buka-suara-soal-hoaks-rekaman-dimarahi-surya-paloh]. Viewed: 4.5.2024.

presidential elections, a deepfake was leaked in which the leader of the Progressive Party said he had bought votes from an ethnic minority;[60] in Brazil, during the 2022 presidential campaign, a deepfake of a well-known media outlet reported that Jair Bolsonaro was leading in opinion polls (Funk *et al.*, 2023).

Moreover, in the 2023 US primaries, Ron DeSantis used an AI-generated image of Trump embracing Dr. Anthony Fauci, who was responsible for managing the COVID-19 pandemic and was held in very low regard by certain groups of Republican voters;[61] Donald Trump was also the protagonist of images in which he posed next to African-American voters generated by AI in an effort to attract the black vote;[62] also in the United States, in February 2024, an audio deepfake of the Texas governor praising Putin appeared.[63]

In Mexico, deepfakes created by voice cloning have played an important role in the disinformation market, as demonstrated, for example, by the cases in which a candidate (Marcelo Ebrard, from the Morena party) "confessed" that he would vote for his opponent, or in which a candidate (Martí Batres, Head of the Federal District Government) criticized an important ally with the aim of discrediting him. GenAI technology has become very cheap and is within the reach of any campaign, making it increasingly difficult to detect. The dissemination system is through pseudo-news outlets that collect and reproduce false statements, even in some cases with the added benefit of paid advertising. In Batres's case, boosting the content would have involved an investment of at least 32000 pesos (Canedo, 2023), equivalent to about 2000 US dollars.

The issue becomes more important insofar as increasingly sophisticated deepfakes have the collateral effect of "undermining the credibility of legitimate information, turning false information into reality and reality into something that is possibly false" (Denemark, 2024, p. 127). Politicians around the world are already using GenAI as a scapegoat (excuse) to dismiss unfavorable evidence against them, claiming that this information was

[60] Available at: [https://www.bloomberg.com/news/articles/2023-09-29/trolls-in-slovakian-election-tap-ai-deepfakes-to-spread-disinfo]. Viewed: 4.5.2024.

[61] Available at: [https://www.npr.org/2023/06/08/1181097435/desantis-campaign-shares-apparent-ai-generated-fake-images-of-trump-and-fauci]. Viewed: 4.5.2024.

[62] Available at: [https://www.bbc.com/portuguese/articles/cqqwrdr32veo]. Viewed: 4.5.2024.

[63] Available at: [https://twitter.com/nexta_tv/status/1753857326849589512?s=48&t=7U06CMqa3d40yKYwVC2fAg]. Viewed: 4.5.2024.

generated by GenAI. This tactic undermines the truth and destabilizes the concept of reality. In this sense, social control over political events is made more difficult since the widespread knowledge of GenAI's existence creates, for leaders involved in real scandals, a quick – and broad-spectrum – way of nullifying the impact of their audio-visually documented actions (Bump, 2024; Mukherjee, 2024). Danielle Citron and Robert Chesney, when discussing the "*liar's dividend,*" explain that convincing the public that fictions are out of control is a very effective formula for evading responsibility for reprehensible acts (Citron; Chesney, 2019, p. 1785).[64] In this context, this technology devalues the "epistemic value" of audiovisual evidence to the benefit of political impunity (Labuz; Nehring, 2024).

Hence, regardless of their capacity to convince, deepfakes devalue the information landscape since their mere existence is sufficient to create a general environment of doubt and confusion (Maher, 2022, p. 7).[65] At this rate, even if they move rapidly to discredit the falsehoods that affect them, the actors involved will have lost, at minimum, their initial state of self-assurance, losing trust and social support. As Leong Suying (2022, pp. 31, 34) points out, mere doubt or suspicion is enough to generate distrust and with it, damage in the political sphere. Therefore, deepfakes, even if they fail to convince (or are denied), do usually have some kind of effect.

In February 2024, an Instagram account, which serves as an unofficial hub for supporters of presidential candidate Ganjar Pranowo, shared a 48-second deepfake video, generated by AI, in which the deceased former president of Indonesia, Suharto, criticized the current government of Joko Widodo (RTW, 2024). In another video, also an AI-generated deepfake, this time three minutes long, Suharto spoke of his legacy as president. Bearing in mind that he died in 2008, it is unlikely that the

[64] Also with the intention of suppressing certain information, deepfakes have been used to discredit or intimidate investigative journalists responsible for surveys or reports capable of negatively affecting the image of political leaders and prominent figures (Samoilenko; Suvorova, 2023, p. 507).

[65] In February 2024, an Instagram account which serves as an unofficial hub for supporters of presidential candidate Ganjar Pranowo shared a 48-second deepfake video, generated by AI, in which the deceased former president of Indonesia, Suharto, criticized the current government of Joko Widodo (RTW, 2024). In another video, also an AI-generated deepfake, this time three minutes long, Suharto spoke of his legacy as president. Bearing in mind that he died in 2008, it is unlikely that the video would be taken as authentic by a considerable part of the population. However, its dissemination had a viral effect: it was viewed by more than 4.7 million people on the social network X, and served to strengthen the candidates linked to the party that is synonymous with Suharto, Golkar (CNN, 2024-b).

video would be taken as authentic by a considerable part of the population. However, its dissemination had a viral effect: it was viewed by more than 4.7 million people on the social network X and served to strengthen the candidates linked to the party that is synonymous with Suharto, Golkar (CNN, 2024-b).

The challenge posed by deepfakes is further compounded by the current limitations in detection capabilities. As Suying (2022, p. 38) notes, no technology exists that can reliably identify these manipulations with complete certainty, and the available detection tools remain largely confined to specialized experts, thus creating a significant asymmetry between creation and verification capabilities. In their current phase, detection tools can, in a best-case scenario, indicate probability or "provide a signal" that is more or less satisfactory but never definitive. AI technology evolves rapidly, increasingly eliminating the clues that make debunking possible, which also affects the full exercise of political rights in the dimension of access to jurisdiction.[66] In addition, debunking methodologies, although successful, tend to be valid only for a certain period of time (Mukherjee, 2024). It is worth highlighting the case of the Slovak election, where a deepfake was released during the day of reflection, making any conclusive denial with an irrefutable technological basis impossible until days after the election was held (Meaker, 2023).

Besides that, deep falsehoods of erotic content (deepnudes) have been used to embarrass, humiliate, harm, and eventually force the withdrawal of candidates, especially female. In some countries, criminal practice can constitute a hypothesis of political gender violence, implying a prison sentence, as in the case of Brazil (art. 359-P of the Penal Code). Despite the legal provision, however, in recently organized elections some candidates for mayor of São Paulo, Rio de Janeiro, Taubaté and Bauru, had their faces inserted into adult content produced with AI (Soares, 2024). Alejandria Ocasio-Cortez, an American congresswoman, became also a famous victim of political attacks based on deepfake porn, starting in 2019.[67]

---

[66] After all, as Goltzman; Soares (2024) point out, the complaints filed before the Brazilian Superior Electoral Court carry the procedural burden of proving fraud. Given that the parties bringing the case have to convince the court, the difficulty of proof tends to limit the prospects of success in these actions.

[67] According to cybersecurity firm Home Security Heroes, the online presence of deepfake videos in 2023 increased by 550% compared to 2019. Within the total of this content, 98% represented pornographic videos. But it is necessary to look at this data from a gender perspective as 99% of the so-called deepfake porn or "fake porn videos" involved women. For many of them this meant being held accountable for the veracity or falsity of their actions (Dubini, 2024).

2.2.2 Fragmentation, Polarization, Destabilization, and Instigation of Conflict

In a context of disappointment, upheaval, and latent indignation, "negative partisanship" (Klein, E., 2020, p. 24),[68] "incendiary language" (Del Rey Morató, 2007, p. 198), and aggressive communication dominate the "toolbox of discursive production" (Leslie, 2021, p. 243) to establish a conflictive relational outlook which is incompatible with moderation, mutual acceptance, and constructive exchange.[69] The populist technique, adopted by more and more political actors, "feeds on negative emotions;" mockery is used to "dissolve hierarchies" and authorities become "objects of ridicule" (Da Empoli, 2020, p. 21). In this manner, the confusion between truth and lies intermixes with the effects of fiction, humor, and terror that fuel citizen perception, giving rise to the sentimentalization of politics and enshrining conflict as a method. "Sentimental democracy" (Arias Maldonado, 2016) has prompted some authors to lament what they consider to be "the end of the Enlightenment" (Lasalle, 2017).

The crisis of confidence facilitates social animosity, which leads people to succumb "to the charms of anti-politics" (Innerarity, 2018, p. 22), incorporating conspiracy theories (Salinas Olarte, 2023, p. 338) and simplified, provocative, and unconstructive information[70] to the point of

[68] The expression defines a behavioral phenomenon that is, to a large extent, the result of polarization in which the definition of preferences is not driven by positive feelings (sympathy, trust, synergistic thinking) in relation to the available options but by negative emotions (repulsion, resentment, aversion) nourished in opposition to distinct and opposing segments (Klein, 2020).

[69] "A notable feature of Infocalypse is that it is becoming increasingly difficult to form a reasonable consensus on how to represent or perceive the world. Too often, it can seem that we are forced to 'choose a side'. In the Infocalypse, even agreeing on a framework of common 'facts' within which a rational debate can take place can be extremely difficult. As the number of people who become politicized in our contaminated information ecosystem increases, well-intentioned efforts are directed toward winning arguments on increasingly intractable issues (race, gender, abortion, Trump, Covid-19...), culminating in a doom-loop of partisanship. Neither side can persuade or convince the Other in the Infocalypse: each attempt only risks entrenching further division. Ultimately, this growing division in society will not be solved unless and until attention and energy can be redirecting into addressing the structural problems of our broken information ecosystem" (Schick, 2020, p. 12).

[70] "Indignation taints everything with clichés: our biggest problem is the political class, there are too many of them; enough parties, end them all; it does not matter who does it, nobody makes the right decisions or they make them too late, they spend the day rambling; let's not play on emotions, the left and the right no longer exist; they are incapable of agreeing; it is possible, but they do not want to; they do not represent us; the more transparency, the better; it is all down to a lack of ethics" (Innerarity, 2018, p. 22).

assuming "aggressive identities" (Lasalle, 2021, p. 185). These, through digital manifestations of hatred and intolerance, devalue the public sphere, delegitimize democratic institutions, and alter "civic protocols (with their rules about what can be said)" (Kiffer; Giorgi, 2019, p. 16), which lowers the quality of the electoral process. Thus, emotional communication gains strength due to the fact that "false narratives often depend on how people feel about the issue, rather than on what they think about it" (Bowman; Cohen, 2020, p. 225).

The experts' analysis of this communicative culture reveals three critical impacts on democratic discourse: (a) "compromises civilized debate and the possibilities of understanding, turning adversaries and their sympathizers into enemies positioned in a relationship of threat;" (b) diverts the discursive agenda, "which stops orbiting around public needs to, in many cases, devote energy to 'culture wars' over non-existent problems;" and (c) "it brings illegitimate advantages that diminish the rate of procedural justice and depress the conditions for the full exercise of active political rights (conscious and well-informed voting) and passive political rights (equality of conditions, based on equivalent respect for the rules of the game)" (Alvim *et al.*, 2023, pp. 117–118).

From this perspective, technological tools, in addition to boosting "pseudological" visions (Rodríguez Ferrándiz, 2018, p. 81) that consolidate the post-truth, serve to strengthen other categories of harmful discourse which are used as a lever to mobilize those segments of the population that feel cheated, betrayed, or excluded. In this order of ideas, affects are used insistently, through the methodical dissemination of "emotional weapons" (Loveless, 2021, p. 64) that submerge consciousness in a "broth of hatred," stimulating the "desire to eliminate the other" (Kiffer, 2019, p. 38).

All over the world, the most recent electoral cycles were characterized by the generalization of "linguistic abuse, with an increasingly notable presence of aggressive discourse full of hatred and intolerance, as well as the widespread dissemination of fake news and attacks on institutions," facilitated – or even stimulated[71] – by the nature of virtual media (Alvim *et al.*, 2023, p. 26),

[71] "Before the arrival of the Internet, 'being active' meant finding a party, joining it, campaigning, knocking on doors, distributing leaflets and attending meetings. Now, from the comfort and safety of their own homes, far-right activists can participate in politics by watching videos on YouTube, accessing far-right websites, chatting on forums, annoying people via voice chat services like Discord and trying to convert users on the main social media platforms like Twitter and Facebook. The fact that all this can be done anonymously greatly reduces the social cost of activism" (Mulhall, 2022, p. 25).

which fosters viscerality, multidirectional fury, and a "herd effect" (Vallespín, 2021, p. 91; Fischer, 2023, p. 131).[72] A recent survey of electoral authorities in 73 countries revealed that the majority (52%) have been the object of digital campaigns attacking their reputation and disinformation (Bicu, 2024). In this sense:

> [W]e observe the undesirable normalization of communicative practices that tend to suppress the enlightened spirit and peaceful antagonism that traditionally characterize the environment of election campaigns, now degenerating into progressively radicalized processes that reproduce, in many ways, the characteristics of an *Information War* or, more appropriately, a permanent *Cognitive War,* in which a clash of rhetorical falsifications with high levels of aggressivity prevents access to factual reality, in the form of "psychological operations" (Bradshaw; Howard, 2017) that produce nothing more than defective opinions and hostile behavior.
>
> (Alvim *et al.*, 2023, p. 68)

Party–political disputes in this tangled mess are exposed to a set of risks derived from the progressive expansion of social acceptance of the new

[72] "In our fairly recent history, we have decided that these impulses are more harmful than beneficial. We have replaced the tyranny of cousins with the rule of law (almost), prohibited collective violence and eliminated the incentive for group behavior. But there is no way to completely neutralize instincts, only to contain them. Social networks, by directly accessing the emotions of the most visceral groups, evade this wall of containment and, in the right circumstances, tear it all down, causing these primal behaviors to overflow into society. When you see a post that expresses moral indignation, 250 000 years of evolution are set in motion. You feel obliged to participate. It makes you forget your own moral sense and accept that of the group. And it makes you feel that causing damage to the target of your indignation is something necessary, and even a pleasure. [...] Behind a screen, far from our victims, there is no sense of guilt at seeing the pain in the rest of the people we have hurt. [...] In the real world, if you interrogate someone for wearing a hat in an expensive restaurant, you yourself will be marginalized, a punishment for disobeying the rules against excessive displays of anger and harassing your fellow diners in a restaurant. On the Internet, if the others pay attention to your outburst of anger, it is probably to be part of it" (Fisher, 2022, pp. 131–132).

communication pathologies – driven by the multiplication[73] of influencers,[74] "symbol figures" (MULHALL, 2022, p. 24), and histrionic spokespeople that peddle provocative[75] and damaging discourses – which address, in an isolated or aggregated way, and in a constant and connected manner: (a) *disinformation content,* (b) *radical content,* and (c) *extremist content*. All this flows

[73] "These [extremist] movements can be compared to a many-headed hydra. If a prominent activist or leader falls into disgrace, it is not a fatal blow. Others will emerge, and the discredited will be discarded. The most important factor is that these movements are truly transnational. Although their main concern is usually local or national problems, activists invariably contextualize them in a continental or even global way. Often, activists from all over the world come together for brief periods to collaborate on certain issues, and these networks act as a link that transmits information to the four cardinal points of the planet. An Islamophobe enraged because a fast-food restaurant in their city serves halal chicken can write a post on a social network and the story will spread. If a particularly influential activist with a large following on social media picks it up, this local story will be adopted and distributed by like-minded Islamophobes all over the world and will act as further 'proof' of the threat of 'Islamification'" (MULHALL, 2022, p. 25).

[74] According to the Global Network on Electoral Justice's Glossary: Digital media and elections, an influencer is "Someone who affects or changes the way other people behave. Recently, companies have paid online influencers to show and describe their products and services on social media, encouraging other people to buy them. Although mainly applied in the commercial field, this concept or personality has recently entered the political and electoral sphere." The figure is the subject of specific regulation in countries such as France (The French Parliament passes the "influencers" law - DW - 01/06/2023), Brazil (`https://www.tse.jus.br/legislacao/compilada/res/2019/resolucao-no-23-610-de-18-de-dezembro-de-2019`), and Spain (`https://www.boe.es/diario_boe/txt.php?id=BOE-A-2022-11311`).

[75] Edward Erikson considers the transformative weight of social networks in social communication and in the process of creating links between the political class and the public, to present the concept of "political fandom" (ERIKSON, 2008), which explains the issue that underlies the usual provocation against opponents and their sympathizers. From this point of view, it has been considered that people tend to process political events more through self-identification than through rational consideration of their own interests. In this context, it is easier for politicians/candidates to cultivate grassroots support through an effective management of digital popularity, via the exploitation of theatrics such as grandiloquent tweets or video clips with impactful statements on social networks, than through real social recompense for a mandate in the form of concrete achievements. In short, "searing" is worth more than accomplishments and, as such, it is possible that today building a new hospital has less value than an insult on social networks or a handful of attractive memes.

effectively through platforms, relying on the help of AI algorithms to germinate and thrive in information bubbles.[76]

Edward Erikson (2008) considers the transformative weight of social networks in social communication and in the process of creating links between the political class and the public to present the concept of "political fandom," which explains the issue that underlies the usual provocation against opponents and their sympathizers. From this point of view, it has been considered that people tend to process political events more through self-identification than through rational consideration of their own interests. In this context, it is easier for politicians/candidates to cultivate grassroots support through an effective management of digital popularity, via the exploitation of theatrics such as grandiloquent tweets or video clips with impactful statements on social networks, than through real social recompense for a mandate in the form of concrete achievements. In short, "searing" is worth more than accomplishments and as such, it is possible that today building a new hospital has less value than an insult on social networks or a handful of attractive memes.

The dynamic relationship between algorithms and user behavior creates a complex feedback loop where collective actions shape and are shaped by automated recommendations. These systems effectively encode human prejudices and tendencies into their operational logic and amplify existing biases through continued online interactions. Evidence of this phenomenon can be observed through the algotransparency.org platform's analysis of YouTube's recommendation algorithm, which reveals how initial searches for political

[76] "As a result of various characteristics of digital platforms, such as their monetization models, a favorable environment has been created for the emergence of a veritable fake news industry, one which is even associated with attempts to manipulate voters, especially in favor of authoritarian candidates. The problem of information bubbles goes far beyond partiality and the transmission of a single view of events, which would be worrying enough. We are witnessing the creation of veritable parallel worlds, fueled by facts that are blatantly false and far removed from reality. In this context, the use of artificial intelligence for political and partisan purposes, by allowing the creation or manipulation of images, sounds and videos – so-called deep fakes – presents a new and much more serious dimension of risk, with the potential to prevent the identification of what is real and what is not. Not in vain, the disturbances of January 8 were created and fueled by diverse content produced by artificial intelligence, without digital platforms having acted to contain the spread of falsehoods. This is acknowledged by the Meta Supervisory Board itself, which at the time pointed out the platform's serious omissions" (Frazão, 2024).

content systematically evolve toward extremes. Whether beginning with conservative figures or liberal politicians, the system progressively guides users toward increasingly radical content – from climate denial to anti-immigration rhetoric, or from socialist perspectives to conspiracy theories – to create ideological echo chambers that deepen with each interaction (WILLIAMS, 2021, p. 256).

In effect, extremist networks opt for insistence as a strategy for naturalizing ideas considered intolerable, which – due to the effect of time – become tolerated or, at least, exempt from objections (Figure 2.10). Specifically, they work with the theory of the "Overton Window," as Michele Prado explains: "The theory states that 'talking about what was previously unthinkable will move the Overton Window and ideas that were previously radical will end up entering the mainstream.' In this way, the limits of acceptable public discourse gradually widen. Thus, Holocaust deniers and neo-Nazi groups have increased the frequency of their discourse on chans and other digital platforms to normalize their extremist concepts" (PRADO, 2023, p. 98).

In the medium and long term, disinformation campaigns deepen fractures and intensify nationalist, ethnic, racial, religious, and ideological tensions (WARDLE; DERAKSHAN, 2017, p. 4). False narratives promote agendas that foment sociocultural divisions and lead to the exacerbation of

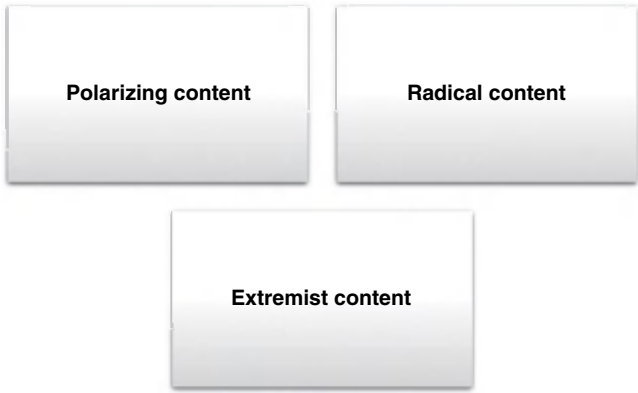

FIGURE 2.10 Taxonomy of discursive pathologies (harmful discourses).
Source: authors' own elaboration

political and emotional polarization.[77] From this perspective, they represent a "threat to democracy," since the transformation of adversaries into enemies "infringes the supreme value of pluralism and the right to difference" as well as the equal dignity of individuals and groups who should be free to compete for a space in the "direct political debate." In this sense, "it is often said that political polarization is the precursor to totalitarian regimes," as history has demonstrated (Martín Guardado, 2023, p. 213–214).

2.2.2.1 Polarizing Content

Polarizing content is content which fosters social division and accentuates it. In this way, fragmentation is simultaneously a cause and a consequence. As a result of "truth on demand," in which everyone is initially fed information as they wish, only to end up being fed automatically by algorithms, what we used to call the public sphere – understood as a kind of imaginary place where social and political dialog took place – becomes fragmented into bubbles (Pariser, 2011) in such a way that each person develops this conversation in their own little niches, unified by community and identity ties. In these

[77] "[D]isinformation on social networks is our daily bread. But it also influences a very specific area on which political scientists have been focusing in recent years: polarization. Algorithms only show the user a thousandth part of the iceberg that is the totality of a social network. There are no half-measures, all content is automatically selected according to the algorithm's orders and structure. As a result, the posts that a common user sees are profiling their behavior and way of interacting with others. The groups 'us' and 'them' are formed. Users tend to gather together in groups that are similar to one another, and respond to the same interactions in a similar way. This phenomenon reflects the human need to feel part of a group; solitary or alternative opinions tend to be avoided. It is not usually the average person, but rather the media and political parties who decide what is debated. [...] Debates are usually launched with two types of stance: for or against. [...] This way of dualizing public opinion has consequences. [...] the process is exacerbated by the dazzling speed with which the media works to report the news: the need for eye-catching headlines so that users 'click' on the articles or news items published that moment. All of this results in a dynamic in which information is not properly verified: rectifying a previous error does not have the same repercussions as the news bomb of the moment. This is the environment in which the user interacts, fake news or news that tells only half-truths polarizes the user and the environment in which they move. At the same time, a large group of polarized users will follow and feed back into the environment, generating news that is in line with their demands. There is no way of verifying all the information to which we are exposed and we often have to rely on what we are told and assured" (Vivas Escribano, 2023, pp. 357–358).

closed information environments, extreme positions and polarization increase as a direct consequence of the fragmentation and isolation that make it more difficult to practice the habit of getting to know the other and putting oneself in their place.

> The notion of "bubbles" is often associated with the growing polarization and radicalization of political debate. [...] due to the use of algorithms and hashtags, social networks are responsible for the creation of "bubbles" (echo chambers) in which users are only exposed to information and opinions that interest them. This can happen spontaneously – users only follow people, hashtags (usually references to certain topics) and media from a particular social network that they like or are interested in – and/or automatically: through algorithms, social networks identify users' preferences and select or prioritize the content they want to view.
>
> (Barcellos; Terra, 2022, p. 278)

Opinion becomes collective, of the group, and one's own thinking adapts to that of the tribe, which becomes the only standard measure. From that moment on, "our loyalties and prejudices are governed by instinct and rationalized as calculation" (Klein, 2020). Thus, reason is replaced by the rationalization of group prejudice, which is often confounded with reality. The story replaces the fact and the slogans become "truths," more the fruit of faith than of analysis. One's own truth excludes that of others and it is always "the other" who lies or manipulates. Reality becomes only what confirms our point of view and demonstrates how wrong the point of view of others is.

Judgment marks you out as being or not being "one of us," depending only on who has issued said judgment. Criticism of others is aggravated, and judgment of our own group is suspended, with a double edge that celebrates in those close to us what we cannot tolerate in our adversaries. At the same time as asking for understanding in the judgment of the group, incurring resounding silences, a literal interpretation of others is demanded, even when this goes against common sense and always with ostentatious gesticulations. The most disproportionate and ridiculous thing about the other's thinking is selected, when its content is not directly altered, in order to reinforce one's own viewpoint. Even when they end up recognizing what has happened, when there is no other alternative, they always end up exonerating the guilty party from responsibility: as Orwell stated, "there is no crime,

absolutely none, that cannot be condoned when 'our' side commits it." There is a congenital inability to recognize in one's own what is decried in others, and it is determined that any action, no matter how illegal or immoral, is acceptable in order to win or to defeat the other. Even when one side changes its ideas, it is better to change ideas than to change sides. This is war and in wartime anything is permissible.

Polarization transforms our relationship with others. It changes the way we think not only about ideas that do not coincide with our own, but also about those who hold opposing viewpoints. Those who think differently are silenced and expelled, and in the name of tolerance, "Zero Tolerance" of all those who do not coincide with our way of thinking is embraced. The moderate is confused with the equidistant, and both are labeled as traitors. As if sharing objectives prevented us from pursuing them in different ways, as if the moral obligation to defend a just cause exempted us from doing so in an intelligent way, those who try to understand the other are punished, while we reject any kind of agreement, which is presented as irrefutable proof of a total absence of conviction, thus forcing us to choose a side. Those who embrace the "true faith," with a religious conception of politics, punish the lukewarm, look down on the dissident, on those they believe incapable of understanding reality, and assume that destroying the enemy is the only mechanism for achieving the survival of their vision (the only acceptable one), of democracy, and the nation.

In this way, the need to capture attention in homogeneous groups generates the need to always raise the tone a little more, whether in the background or in the form, which in turn causes a veritable competition for attention in which these closed groups end up polarizing in an almost natural way, driven by radicalization dynamics which are in permanent feedback (Rubio, 2018). Agreement, in any field, becomes impossible. When positions become unacceptable for the adversary, only the sides remain. Every decision, no matter how complex, is resolved in a binary choice. All or nothing, black or white... with the secret hope that by only offering two alternatives, the citizen will have no choice. "Us or chaos," even when, as in the well-known graphic joke, the rigged alternative hides the fact that "we" too are chaos. The legitimacy of the other side to engage in politics, or even to run the government, is questioned, and any action to push it aside ends up being justified, even if it goes beyond the legal framework. This fuels a dangerous form of anti-politics and refuge for activists, mobilized minorities who are gaining power at the expense of all those who reject the sheer absence of rules of the game. In this way, polarization finds its point of no return when, paradoxically, it achieves its goal.

In the electoral arena, where these dynamics have shown their effectiveness, *polarizing content* can be grouped into three main forms: (1) *political disinformation* (or disinformation among peers) consists of the construction and propagation of false and decontextualized statements that defame the image of opponents and indirect opponents; (2) *anti-system* disinformation embodies strategies to depreciate public bodies (electoral institutions, in particular), as part of coupist populism's playbook; and (3) *counter-civic disinformation* that aims to influence the exercise of the vote, spreading warnings and disinformation in order to sabotage electoral processes, by discouraging or hindering the exercise of this right (*voter suppression*).[78]

Proving GenAI's potential for carrying out covert interference campaigns, OpenAI recently revealed the discovery of the malicious use of its platform in five destabilization operations carried out systematically by foreign actors. Using the ChatGPT tool to generate and translate news articles and short comments that fed fake accounts, the operators of the scheme sought to influence public debates around certain agendas on social media (Carrascón, 2024). The five operations detected involved: (a) a Russian operation targeting Ukraine, the Baltic countries, and the United States, focusing on the creation of politically motivated comments for distribution on the Telegram app (Operation "*Bad Grammar*"); (b) a Russian initiative to produce comments in several languages, using platforms such as X and 9GAG for dissemination (Operation "*Double"*); (c) a Chinese network that used AI to create texts in several languages and manage online platforms (Operation *"Spam Camouflage*"); (d) an Iranian scheme that produced and translated lengthy articles for publication on affiliated websites (Operation "*International Union of Virtual Media*" – IUVM); and (e) an Israeli commercial enterprise that created articles and comments for social-media platforms such as Instagram, X, and Facebook (Operation "*Zero Zeno"*) (Matoruga, 2024).

[78] A recent example of counter-civic disinformation was the dissemination of a deepfake video in which Muhammad Basharat, a candidate in the Pakistani elections, appeared asking his voters to stay at home and boycott the vote. Strategies like this, also detected in fake speeches by President Biden, demonstrate the opening of a new variety of damaging strategies which involve the misuse of candidates' images and voices to spread disinformation among their own voters. Bearing in mind that people tend to resist messages that come from figures they don't like or trust, these deepfakes would be in a position to amplify their persuasive potential (Adami, 2024).

2.2.2.2 Radical Content

Radical movements[79] are, by definition, bearers of a "closed mindset" that excludes and invalidates everything that contradicts the "purity" (ZAYAN, 1984, p. 777) of their political, ideological, and moral ideas. In the closest sense, the term "'radical' is distinguished from 'moderate', 'conventional' and even 'legal', terms that can also be applied to the aims and means adopted by a political movement." From this point of view, radicalism "indicates a more extreme, absolute and uncompromising commitment" to a particular world view (WRONG, 2000, p. 645), which fosters inflexible public behavior, governed by a general logic of rejection of external groups (*outgroups*). This behavior, which is always negative, can involve different degrees of risk and harmfulness, according to the idea of the "pyramid of hate," devised by Dr. Gordon Allport, Professor of Psychology at Harvard University (Figure 2.11).

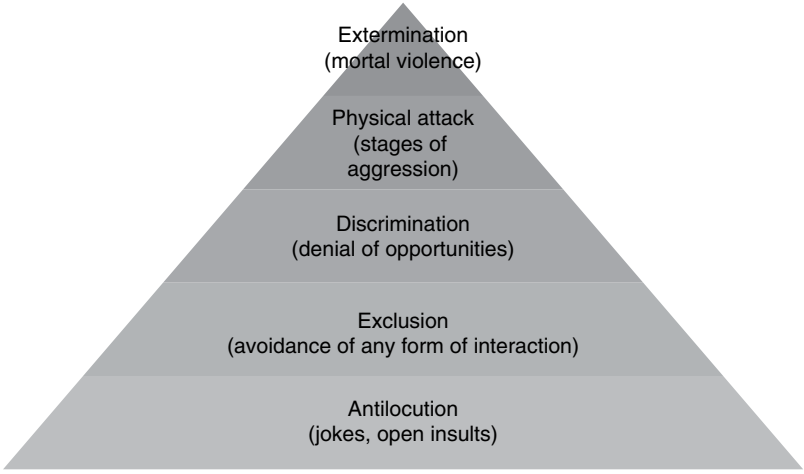

FIGURE 2.11 Allport's pyramid of hate. *Source:* authors' own elaboration based on WILLIAMS, 2021, pp. 30–32 / Estante virtual

[79] "The word 'radical' comes from the Latin *radix*, meaning 'root'. This word was used by some Greeks philosophers as equivalent to 'principle', 'foundation', 'cause' or 'primary reason' of things. By this they meant the real or conceptual origins or foundations of something. The *radical*, therefore, was what was rooted, what was deep, what had a solid basis. Today, in politics, this word refers to a person who has deep-rooted convictions or who holds firm elementary principles. A person like this usually maintains an uncompromising and inflexible attitude towards their principles or their politics and goes to the ultimate consequences without making concessions. Generally, they are in opposition to *system,* hold a critical attitude towards existing institutions and advocate their abolition or reform" (BORJA, 1997, p. 845).

In this sense, *radical content* expresses intolerant statements that, in some way (express or implicit,[80] direct or indirect,[81] particularized or general[82]), deny, attack, or nullify the moral dignity of people, groups, or divergent opinions, regardless of the underlying reasons. Therefore, these statements reflect the vaguer notion of "hateful ideology" used by the Brazilian Superior Electoral Court (TSE),[83] insofar as, although they do not really question the central institutions of democracy (especially the technique of electoral legitimation), they oppose certain liberal principles (freedom of the press, separation and independence between powers, etc.), questioning the legitimacy of non-elected powers (specifically, the judiciary) and, more specifically, the dignity and rights of minorities and egalitarian humanism in a more general sense (Mudde, 2019, p. 21; Mulhall, 2021, p. 28; Levitsky; Ziblatt, 2023, p. 133).

[80] Radicalism entails political, social, and legal risks, although it often uses communicative mechanisms to present its friendlier face. In this context, it usually receives a "rhetorical softening" aimed at "reducing stigmatization." In terms of discrimination, for example, the most common tactic is "ethnopluralism," defined as "a defensive strategy against immigration and cultural incompatibility, based on the construction of a national affiliation, in which ethnic, cultural or religious criteria are emphasized and condensed into ideas of collective homogeneity, linked to authoritarian political models". "This serves as a disguise for racist, xenophobic and anti-liberal discourses, as well as reinforcing reactionary concepts of the radical and extreme right that express a return to tradition, homogeneous communities and cultural values (patriarchal and hierarchal systems) that have been destroyed, according to their followers, by liberalism. [...]. Positioning oneself as a defender of ethnopluralism creates the false illusion that there is no racial and/or nationalist content and motivation, that its defenders are just individuals protecting the values, traditions, sovereignty and culture of their peoples in the face of the supposed onslaught of the global liberal elites whose desire is, with multiculturalism, to annihilate supposedly non-Western values and homogenize societies (which is a paradox given that these groups support the defence of homogeneous communities and condemn the introduction of ethnocultural individuals who are different to the group)" (Prado, 2023, pp. 194–196).

[81] "The methods used are diverse" and involve "sharing the bait" (dog whistling). "In politics and in the online manipulation of public opinion, [dog whistling] acts as a coded message for a certain group or individual to suffer massive attacks" (Prado, 2023, p. 157).

[82] An example of this is "stochastic terrorism" (random), "rhetoric which, although hostile, does not explicitly tell someone to carry out an act of violence against a group, place or institution, but a person who feels threatened may be motivated to do so. Therefore, this rhetoric incites random violence" (Prado, 2023, p. 233).

[83] Article 9-E, point IV, of the Res. n° 23.610/2019 of the TSE (Brazil).

As a systemic problem, radical ideas are sometimes the basis of conspiracy theories, within which disgust and prejudice are consolidated into a degenerate narrative that persists over time. This is the case, for example, of hate narratives that project an image of the Jews as greedy aspirants to world control, of Native Americans and Africans as primitive, or of Muslims as barbaric invaders, terrorists, or sexual predators (George, 2020, p. 147). The same happens with the exploitation of the "paranoid style" based on extravagant theses predicting that cultural Marxism, globalism, and progressive ideology are leading society as we know it to collapse, or even extinction (Prado, 2023, pp. 197–219).

In graphic terms and as an example, we would point out that this radical content can be found in the habitual rhetoric of the extreme right, in line with what researcher Michele Prado (2023, p. 80), among others, asserts:

> In general, the rhetoric of the *extreme right* makes explicit the tendency to dehumanize the groups it considers inferior. Adjectives such as "rats," "criminals" (in relation to immigrants), "scum," "lice" are often used in far-right discourse. To this is added rhetoric and belief in existential threats and apocalyptic visions such as the "end of Western civilization" [or the end of the evangelical lineage or Christian civilization].

In this vein, generalizations are also very common, as can be seen, for example, in the phrase "all terrorists are immigrants," uttered by Hungarian far-right leader Viktor Orbán (Bauman, 2017, pp. 34–35), and in the phrase, "Why are we having all these people from shithole countries come here?" uttered by Donald Trump in reference to African nations, El Salvador, and Haiti (Viennot, 2021, p. 19).

Extremist currents, on their part, are moving in a clearly anti-democratic direction, taking up a hostile anti-system struggle that is openly opposed to the order based on popular power (Mudde, 2019, p. 21).[84] In this vein, the insurgency against constitutional democracy is

[84] "Although 'extreme right' is a broad term, it needs to be subdivided into two categories: the democratic radical right and the more extremist extreme right. Political scientist Cas Mudde explains that the extreme right 'accepts the essence of democracy, i.e. popular sovereignty and majority rule,' while the radical right opposes fundamental elements of liberal democracy, in particular the rights of minorities, the rule of law and the separation of powers. Although this is a useful distinction, it should be pointed out that much of the radical right's acceptance of democracy is tactical or performative. Several currents of the radical right, though not all, can currently be described as 'populis', which Mudde defines

based on "language as a weapon" with which to imprison public opinion in a "permanent state of siege," in which each and every attack (verbal or physical, even terrorist) can be justified as an "act of defense and prevention" (Strobl, 2022, p. 17).

2.2.2.3 Extremist Content

Extremist content, in this area, oversteps constitutional boundaries to advocate, on the one hand, coupist, conspiratorial, and destabilizing theses against the acceptance of election results, in favor of deposing civil power and establishing regimes of force, and on the other, the incitement of chaos, whether through the calling of strategic stoppages (truck strikes, strikes by productive sectors, or strikes by essential services), calls for acts of vandalism, or mass protests in defense of authoritarian or anti-democratic agendas, as a strategy to increase social unrest and provoke recurrent mobilization. The same category includes the open defense of dictatorial agendas, such as the summary illegalization of parties, the exclusion of trade unions, the closure of Congress, purges at the top of the judiciary or other guarantee bodies (including electoral bodies), intervention in or closure of press organs, and the defense of torture. On a third front, public stances in favor of open violence against minorities or opposition currents, groups, or specific institutions are also radical. This violence can be physical or symbolic, punctual or systemic, with a destructive purpose – as in the case of terrorist movements[85] – or not, but it always engenders a climate of acrimony, threat,

as 'a ideology which considers that society is ultimately separated into two homogeneous and antagonistic groups, the pure populace and the corrupt elite, and asserts that politics ought to be an expression of la volonté Générale [general will] of the people'" (Mulhall, 2021, p. 29).

[85] "Terrorist activity is characterized by the desire to impose certain points of view through the use of violence and the socialization of pain, thereby generating a sense of insecurity and fear. Today, the phenomenon of terrorism is a threat fueled by the ease of travel, digital interconnection and social models that generate malaise, exclusion or hatred." The association between terrorism and radical politics goes back a long way, from nineteenth-century European anarchism to current jihadist movements and has also formed part of separatist movements in Spain, Ireland, and Algeria, among many others (Pérez-Francesch, 2022, pp. 718–719). The links between extremism and terrorism are, in fact, extensive, and disinformation and harmful content, with the assistance of new technologies – including AI – act as a powerful force in stimulating and recruiting young people to join violent movements (Gandhi, 2024).

or persecution. These modalities, as Cherian George (2020, pp. 145–146) points out, have great potential for sparking debate:

> Our attention tends to focus on extreme discourses; these already use language that violates social norms, such as openly racist epithets, or contain explicit calls to action, such as incitement to violence. But these expressions are only a small part of a hate campaign and do not work on their own. Hate campaigns are made up of multiple messages, with multiple layers and slightly interconnected, spread by different actors over years or decades. The master narratives, which often focus on the noble characteristics of the group and often hark back to a mythical golden age that demonstrates the true potential of the community, provide the backdrop. Past traumas can also be highlighted to increase the community's sense of victimization and heighten the need to prevent "danger." A complementary grand narrative presents a particular external group as intrinsically unworthy due to certain irremediable cultural, religious or ideological characteristics. The effect of these grand narratives is to cultivate an exclusive primary identity, a sense of belonging to a community that is exceptionally valuable and needs to be defended.

Obviously, radical and, above all, extremist messages are part of the classic repertoire of *foreign interference* campaigns[86] in which AI operates in an "astonishing" way (Rebollo Delgado, 2023, p. 74), making fake accounts circulate and distributing segmented content with a strong impact on people's behavior (Bender, 2022, pp. 26–27; Heller, 2021, p. 27) – whether to weaken elected governments, undermine democracy (Kakutani, 2018, p. 132), or to

[86] "The global social order is now clearly subdivided into two worlds: the western one, in which democracy and the enforcement of rights prevail to a greater or lesser degree; and the autocratic one, which is submerged in a distorted informational obscurantism. Globalization has interconnected these two worlds, and this has both positive and negative aspects. The former includes the knowledge that can be gained of countries that until recently we were barely aware even existed [...]. The most negative element for the West is vulnerability to information or disinformation. More specifically, we could talk about disinformation, or the generation of information confusion. Fake news is a major problem today, both at state and individual level, and its aim is to undermine the ideological freedom of citizens. It constitutes a true threat to the state, social organization, and also to the individual, attacking the heart of the social and democratic rule of law, and that which is most essential to the free configuration of the person" (Rebollo Delgado, 2023, pp. 73–74).

ensure the victory of candidates more aligned with certain geopolitical currents, spreading "false and polarizing messages" (MINOW, 2022, p. 357) – at an exponential rate.

The literature even points out that some foreign governments maintain professionalized structures to produce disinformation within specialized centers, whose work has often been denounced by experts. Some examples are the Internet Research Agency, which operates on behalf of the Russian government from a four-story building in St Petersburg (POMERANTSEV, 2020), and the Disinfo Lab, created and run by intelligence officials from the Indian government (MAKHERJEE, 2024).

"National borders have long been permeable to new ideas and trends, including those fomented with a deliberately malign purpose" (KISSINGER *et al.*, 2021, p. 104), which is why technology has been used – mainly by authoritarian governments – to disseminate controversial and dishonest content in various areas with the aim of destabilizing the political framework, social agreements or elections, and popular consultations in strategic countries (BECHIS, 2021, p. 127; MAHER, 2022, pp. 13–16; POMERANTSEV, 2020; SUYING, 2022, p. 23), as seen in the recent US elections, the Brexit referendum in the United Kingdom, the peace process in Colombia, and elections in France, Austria, Germany, Italy (MOORE, 2018, p. 22; SCHICK, 2021, p. 21), Brazil (RUBIO; MONTEIRO, 2023), and Slovakia (DENEMARK 2024, p. 123), among others.

Russia's role in the 2016 US elections is well-known, when the Russian Glavset created bots and web pages dedicated to influencing the votes of different communities (PEIRANO, 2019), such as African Americans.[87] The Chinese are also notorious for this type of campaign. In addition to the spamouflage campaign against the US mentioned above,[88] influence campaigns against Southeast Asia generated with AI to benefit China's geopolitical interests have been detected as well as a continuum of disinformation campaigns involving deepfakes, memes, and news generated by AI in Taiwan. There are many more examples of foreign interferences.[89]

---

[87] Available at: [https://www.motherjones.com/media/2022/09/disinformation-russia-trolls-bots-black-culture-blackness-ukraine-twitter]. Viewed: 4.5.2024.

[88] Available at: [https://www.isdglobal.org/digital_dispatches/pro-ccp-spamouflage-net-work-focuses-on-us-election]. Viewed: 4.5.2024.

[89] Available at: [https://www.microsoft.com/en-us/security/business/security-insider/reports/east-asia-threat-actors-employ-unique-methods]. Viewed: 4.5.2024.

During the 2017 French presidential elections, multiple disinformation attempts were detected against candidates such as Emmanuel Macron, with the spread of fake news and smear campaigns on social networks. For their part, in the run-up to the 2019 general elections, reports emerged about the use of disinformation tactics to question the legitimacy of the electoral process and revive ethnic tensions, especially in relation to the migration crisis. In the 2017 German federal elections, coordinated efforts were identified to spread fake news and discredit established political parties, especially on issues such as immigration and the economy. In Italy, during the 2018 electoral campaign, there were attempts at disinformation aimed at undermining the credibility of political leaders such as Matteo Salvini and Luigi Di Maio, by manipulating public opinion on issues such as immigration and the economy. In Brazil, disinformation played an important role in the most recent presidential elections, generating the ideal breeding ground for the failed assault on democratic institutions on 8 January 2023.

Within the varied practices of destabilization and conflict, hoax campaigns on the *web* are common, taking the form of the escalating dissemination of "false messages with emotional and alarmist content that demand action," such as stockpiling food or emptying their bank account (Silva; Cendón, 2022, p. 29), playing an important role in increasing tensions (Utami, 2018, p. 85). With regard to electoral processes, *hoax* messages are usually linked to false accusations of fraud, usually accompanied by a *call-to-action* linked to a multitude of fervent responses (rejection of the results, petitions for annulment or recounts, pleas for military intervention, and similar delusions).

Under the same logic, AI can also be used to support transversal tactics, linked to the execution of *psychological operations* (*psyops*) which encompass "practices that catalyze disinformation narratives, especially useful in highly polarized scenarios, with the constant introduction of explosive topics that are likely to deepen divisions" (Alvim *et al*., 2023, p. 224). Based to a large extent on "selected information and indicators designed for specific audiences," psychological operations entail important social risks since they focus on strategic *targets* susceptible to manipulation and act on them selectively, taking into account moral compasses, emotions, beliefs, fears, and desires (Gelev; Popovska, 2020, p. 58).

Radical and extremist groups can use AI to trigger waves of *flaming*, inciting "flammable debates" that lead to the exchange of insults on a large scale, radiating over a large area of the online surface (aranha, 2014, p. 125). As a practice that detonates heated verbal debates, based on

differences of opinion (BALOCCO, 2016, p. 503), historical disagreements or feelings of latent intolerance, *flame wars,* associated with culture wars,[90] jeopardize the regular occurrence of electoral processes. These generate instability, intensify conflict, and make it considerably more difficult to maintain peace and mutual respect, all preconditions for the normal and regular development of the processes in which elected public officials are chosen.

Along with the tactics already examined, AI can help spread radical and extremist content through: (a) *large language models (LLMs)*, such as ChatGPT, which make it "easier and less costly to create persuasive and engaging messages;" (b) synthetic media and deepfakes, which allow the improper manipulation of images to create false and alarmist narratives with great emotional impact; (c) the *microtargeting* of political messages through a personalized focus, which makes them more effective; (d) *bots* that reproduce human behavior in a convincing way; and (e) systems of recommendation and access to data, which give rise to the formation of echo chambers in which denigrating narratives spread with greater intensity (GANDHI, 2024, pp. 15–22).

[90] "The concept of culture wars indicates the activation of conflicting, incompatible and incommensurable visions of a specific issue (education, property, weapons, etc.). Culture wars are, in this sense, a socio-discursive and therefore political phenomenon that constructs and, at the same time, expresses ways of understanding the world and ways of being in the world. [...] culture wars translate social life and political dynamics into a conflict of identities, into a permanent game of the expression and differential recognition of 'us' and 'them' (LINDAMAN, 2002). Culture wars are therefore phenomena endowed with a certain theatricality associated with dynamics of political polarization, both ideological and affective (TORCAL; MAGALHÃES, 2022). A relevant aspect of culture wars is that they transform elements of social life into recognizable group signs that activate a whole implicit worldview. For example, actions of not wearing a mask in a space where it is recommended become signs that set a recognizable social discourse in motion and mark the actor as part of a group that identifies itself according to this social discourse. These smaller cultural wars, which we could call 'cultural battles', are part of the contemporary populist discursive strategy and serve to transfer the cultural wars of the political elite to the general public" (AGUADO TERRÓN; VILLAPLANA JIMÉNEZ, 2023, pp. 203–204).

2.2.3 The Rupture of Equity and Neutrality in Communication

In an environment of permanent technological innovation, access to certain technology can determine the equity of electoral processes.[91] Although it would not be reasonable to oblige all political actors to use the same technology, it is necessary to ensure that technology offers its services equally to all contenders. From the moment that selection algorithms dominate the supply of content on social networks, candidates and parties have no way of guaranteeing, by themselves, that their messages actually reach the real recipients, unless – of course – they are willing to pay for it (Prado, 2022, p. 188). Today, the "invisible hand of the algorithms" (Maradeo, 2021, p. 11) organizes *feeds*, *stories,* and *timelines,* as well as defining the hierarchy of search engine results. Thus, along with the precipitous fall in the consumption of information produced by traditional media,[92] in the world of the networks, the flow of information is heavily conditioned by the *design of* the

[91] "Social networks have forever changed the way information is accessed, disseminated and shared. Algorithms mediate and intervene in the selection and promotion of content based on the ultra-segmentation of users' profiles and preferences. Their intervention, which is neither transparent nor subject to external regulation, makes them responsible [...] and moves social networks away from being the mere third-party content platforms that their executives have so often claimed them to be. [...] The algorithm of social networks ends up forming niches, echo chambers of similar attitudes and visions [...]. One's own opinion seeks the 'like', the mass approval [...], and the filters of our bubbles block any response or disagreement. If this is so, algorithmic mediation [...] decisively influences the formation of opinions and perceptions, determines public debate and fosters polarization around shared narratives" (Álvarez, 2023, pp. 174–175).

[92] "[Today] Who gets their news from printed newspapers? How many from radio or television? How many from websites? ¿And social networks? Two-thirds of those interviewed in a survey carried out in 26 countries said that they use social media for this purpose; an even greater number of people receive more information from news selected by online algorithms than by human editors; and only 1 out of every 10 interviewees searches for news on the Internet. People rely heavily on apps and social networks. More than 100 newspapers stopped publishing daily and started publishing weekly, and between 2004 and 2014 at least 664 newspapers went bankrupt. The newspaper business model, based mainly on advertising, is no longer working, because advertising funds have shifted to cheaper and better-targeted online platforms. Currently, 89% of the money spent on online advertising goes to Google or Facebook, and between 60% and 70% of all advertising revenue goes to Internet companies" (Minow, 2022, p. 359).

platforms.[93] This undoubtedly raises very serious questions about the preservation of equality between competing forces.

AI, in combination with other digital technology, shapes the general framework of ideas and opinions that will reach each user, thereby designing the window through which Internet users view the world. Given that they organize everything that will be seen, read, or listened to, network algorithms contribute to the rise of a fragmented world and can "push voters" toward a concrete political current, both through personalized ads (Niyazov, 2019) and subtle influences that affect the formation of beliefs, the development of consciousness, and the consolidation of ideas and decisions.

> Search engines pose another challenge: ten years ago, when search engines were based on data mining (and not on automatic learning), if a person searched for "gourmet restaurants" and then "clothes," their search for the latter would be independent of their search for the former. In both cases, a search engine aggregated all possible information and offered options [...]. But today's search engines are guided by models based on human behavior. [...] If a person searches for "gourmet restaurants" and then "clothes," they may be shown

[93] Max Fisher states that digital platforms have "swallowed up" the big names in machine learning to "build machines that learn exactly what combinations of text, images and sounds will help to keep hooked." For example: "YouTube's system is looking for something that has a wider reach than monthly subscriptions. Its all-seeing eye tracks every detail of what you see, how long you see it, what you click on next. It supervises all this for 2 000 million users, accumulating what is surely the largest set of data on audience preferences ever gathered, which it constantly examines in search of patterns. One of the algorithm's most powerful tools is the affinity between topics. If you watch a cat video until the end, [...] YouTube will show you more of these videos when you return to the site. It will promote, above all, the cat videos that it has found most effective in capturing your attention. [...] Like practically all internet users, I've been through this. I ride a bike [...] and, when I lived abroad, I used to look for videos of trails in the neighborhood to get an idea of the terrain. The system started recommending videos to me that it had never occurred to me to look for: professional rides, tests of new models. It worked; I watched more videos. Over time, the recommendations became more extreme. Impressive crashes, collisions between 10 bikes, maneuvers for those who do not fear death. Although these did not do me any harm, they were not entertaining either; they were simply something that got me involved and were out of the ordinary [...]. The effect is that you attract the user to increasingly stimulating variations of what interests them. If the interest is cats or bicycles, the impact is minimal. If it's about politics, health or other more serious social issues, the consequences can be disturbing" (Fisher, 2022, pp. 148–149).

> designer clothes instead of more affordable alternatives. Designer clothes may be what the user is looking for. But there is a difference between choosing from a wide range of options and taking action – in this case, making a purchase; in others, adopting a political or philosophical position [...] – without ever knowing what the initial range of possibilities or implications was, entrusting a machine with the pre-emptive configuration of options.
>
> (KISSINGER *et al.*, 2021, p. 32)[94]

Platforms have assumed the role of "*modern-day guardians*," acting as "agents who, due to their centrality, exercise all forms of control over the information they create on the *network*, such as selection, *framing, timing*, repetition, retention, among others." In addition, they exercise acute power over individuals who are connected, since they can: "(i) 'trap' users within the network; (ii) protect rules, information, users and communities from unwanted entries; and (iii) even keep ongoing activities within the network from being disturbed." In the light of these considerations, Ana Frazão considers the idea of open platforms and neutral content to be "fanciful," given that, in a dimension that goes beyond algorithmic programming, they also exert an influence in the field of moderation level, because:

> Moderation allows platforms to regulate the discourse of their users, imposing policies and terms of service and managing the flow of illegal content automatically, through software or manually, either before it is published (*ex ante moderation*) or afterwards (*ex post moderation).*
>
> Practical experience has shown that the role of many of these agents is not limited to moderation and the consequences of eliminating and blocking content, but also encompasses the management of

[94] "Throughout the day, a person benefits from and contributes to enormous amounts of data. The scope of this data and the options for consuming it are too immense and varied for human minds to process on their own. The individual comes to rely, often instinctively or unconsciously, on software processes to organize and select necessary or useful information, choosing the news they want to watch, the films they want to see and the music they want to listen to based on a combination of previous individual choices and widely popular selections. Experiencing this automated selection can be so simple and satisfying that it is only noticeable when absent: for example, try reading the news on someone else's Facebook feed or watching films using someone else's Netflix account" (KISSINGER *et al.*, 2021, p. 97).

> how content is categorized and how some gains prominence to the detriment of others. [...]
>
> As a result, limiting the question of responsibility to the "creation" of content has become insufficient to put power and responsibility on an equal footing, given the clear proof that "the moderation and recommendation of algorithm systems determine what content will be seen and listened to by users".[95]
>
> (Frazão, 2022, pp. 561–562, with reference to Gilliespie)

At present, "the journalist is [no longer] the privileged mediator, the guardian of what is or is not news, of what is or is not newsworthy." The press no longer acts as a gatekeeper, controlling the information that flows about candidates and their parties, selecting what seems worthy of being reported (Prado, 2022, pp. 86, 87, and 90). On the other hand, ideas and statements flow from multiple poles, with the personalization of searches and content recommendation algorithms as the central cogs in the process of selecting what does (or does not) reach the eyes and ears of consumers on digital platforms.

Content curation definitely has an impact not only on information consumption but also on our daily lives. Search engines work on the basis of sophisticated algorithms that "select and classify the results according to a series of factors, such as relevance, authority and popularity, which translates into the presentation of information considered most relevant and reliable" according to unknown criteria. This logic has a great effect on the balance of electoral contests, particularly since the results that appear at the top of search pages can affect users' perception of the candidates or the issues that motivated their search for information (Lins, 2023, p. 295). The same happens with the flow of information that takes place on social networks:

[95] "From this perspective, a new front in the discussion is opening up, which no longer focuses solely on the control of content, but on the architecture of the platforms themselves, that is, their design and compatibility with minimum standards of responsibility, care and transparency, in order to mitigate the negative externalities that this model imposes on society. It would be worrying enough if there were no duty of care in relation to the content circulating on the platforms, but it is even more worrying when the monetization system itself can create incentives for and great profits as a result of the dissemination of illegal or inappropriate content" (Frazão, 2022, p. 563).

> Although social-media platforms (and their AI) tend to present themselves as content-agnostic, not only their community rules, but also their filtering and presentation of information can influence the way in which it is created, aggregated and perceived. When AI recommends content and connections, classifies information and concepts and predicts users' preferences and objectives, it can inadvertently reinforce certain individual, group or social choices. In fact, it can encourage the distribution of certain types of information and the formation of certain types of connections, while discouraging others. This dynamic can affect social and political outcomes, regardless of the intentions of the platform's operators. Every day, individual users and groups influence each other rapidly and on a large scale through countless interactions, especially when these are determined by complex recommendations based on artificial intelligence; as a result, operators may not have a clear understanding of what is happening in real time. And the complexities are magnified if the operator injects (consciously or unconsciously) their own values or objectives.
>
> (Kissinger *et al.*, 2021, pp. 102–103)

Current elections, in turn, suffer the exposure of their integrity to so-called *algorithmic interference*, insofar as classification mechanisms, however simple they may seem, can influence people's beliefs and habits to the point of altering the results, as revealed by an experiment discussed by Max Fischer (2023, pp. 151–152):

> In a 2015 experiment, Americans were asked to choose between two fictitious candidates based on what they found about them on the Internet. Each participant found the same 30 search results in a controlled version of Google, but each found these results in a different order. The participants gave greater psychological weight to the results with the highest ranking, even when reading all thirty. The effect, concluded the researchers, could affect up to 20% of the voting intentions of undecided participants. The author of the study, Robert Epstein, [...] stated in an article: "The next president of the United States may win election not only through television advertising or speeches, but also thanks to Google's hidden decisions."

The repercussions stem from the fact that, in today's society, "the media environment is highly competitive." Individuals "have a limited attention span" and therefore have to be "selective about the information they seek," which in practice means "being selective about which sources to pay attention to and which to ignore." In this setting, content producers (media corporations, media outlets, content producers, opinion formers, conspiracy theorists) compete for attention in a game that is known to be tough (Williams, 2023), unfair, and exclusionary. The vast majority of posts made on social networks do not achieve visibility or engagement, so the messages that end up being catapulted by influencing strategies gain a more than privileged position in the battle for persuasion.

In electoral terms, the "discriminatory hierarchy of visibility" (Shoai; López Molina, 2023, p. 255) generates processes of exclusion and censorship by default (Kissinger *et al.*, 2021, p. 97; Metaxas, 2020, p. 247), and therefore produces asymmetries in the balance of opportunities, since it puts direct competitors at an advantage or disadvantage, depending on how the invisible spirits[96] of codification behave.

The negative externalities, moreover, go beyond the *pars conditio* to also jeopardize the substantive basis of the vote, since algorithmic intervention creates information bubbles that block access to a plurality of opinions. Plurality, in this setting, is reduced, since most people consume "news selected by algorithms that are based on previously evidenced interests and which only echo what they already know, without offering opportunities to learn, understand or believe what others are hearing as news" (Minow, 2022, p. 357).

[96] "Facebook altered the news feed of 2 million politically committed people, who began to receive a greater proportion of bad news, no longer having access to videos of cats and the like; when their friends shared a news story, it appeared in their feed, without taking into account the mathematical interference behind the phenomenon. Unlike the decisions made by television channels such as Fox News or MSNBC to highlight one news item over another, algorithmic adjustments are invisible to users and therefore immune from comments or criticism. Apparently, during the 2016 US elections, fake news, especially items in favor of Trump or against Clinton, attracted participation and obtained better results than legitimate news, generating income for Macedonian teenagers who created websites and spread messages on the Internet that repeated polarizing content" (Minow, 2022, pp. 372–373). For more information see Peirano, 2019, chapter 7.

In this context, the filtering methods used by social-media platforms influence the way in which people become informed about, access, and shape their perception of reality:

> [N]ews depends not only on the visibility of private companies, but on the decisions made by dispersed, powerful, private and competing companies that are not in the news business, carrying out deliberations that are invisible to those affected and blatantly manipulated by enemies of the nation and treacherous opportunists. Information deserts affect many local communities; these are uninformed about the actions of the government and the regional population, denying many people what they need to govern themselves and demand responsibility from others. The architecture of virtual communication allows anonymous messages initiated by robots to influence the algorithms that determine what news individuals receive. Digital companies benefit from the news links shared by users without investing in the apparatus needed to investigate, verify and report the news, which ends up undermining people's ability to obtain and trust in [legitimate] news.
>
> (Minow, 2022, p. 374)[97]

As a result of the aforementioned, algorithmic governance is attracting the attention of control bodies and public policymakers and is at the center of concerns related to protecting the integrity of elections in the digital age (Kofi Annan Foundation, 2020, p. 40).

[97] "The management of Facebook and Google insist that they are technology companies, not news companies. They rely on algorithms and not on human editorial decisions to select what people see; their aim is to keep consumers' attention, not to cover the news. Sheryl Sandberg, Facebook's chief operating officer, explained that: 'We are very different from a media company (...) In essence, we are a technology company. We hire engineers. We don't hire journalists. No one is a journalist. We don't cover the news. Despite these statements, more and more people are receiving news via social networks: through links and forwarding messages to other people. Emily Bell rightly observed that 'social networks have not only swallowed up journalism, they have swallowed up everything. They have swallowed up political campaigns, banking systems, personal histories, the leisure industry, retail, even government and security'. Digital platforms, apparently neutral, are easily manipulated by vendors and extremists who use search optimization and digital clicks to achieve self-serving goals, while guaranteeing profits for Google and Facebook" (Minow, 2022, pp. 363–364).

On the other hand, breaches in equity can occur in the context of content moderation on social networks.[98] In a parallel dimension to compliance with court orders when illegal content is detected, platforms can also, in principle, sanction their users in the event of flagrant non-compliance with conditions of use or community policies. Bearing in mind, therefore, that by unilateral decision networks can eliminate or reduce the visibility of messages (*shadow banning*); suspend accounts, groups, and channels; or ban profiles, there is no doubt about the importance of maintaining an impartial and neutral balance in the treatment and moderation of content on social networks, without the influence of political or partisan inclinations.[99] If, on the one hand, proactive moderation activities are important to ensure an attentive and active fight against false and damaging narratives, then on the other

[98] "One of the most controversial issues is the internal moderation procedures for inappropriate content that can lead to the removal of such content, which have often been criticized for their opacity and the defenselessness they generate in those affected, which can also lead to inequality between individuals or groups when it comes to participating in public debate. As an alternative to the 'notice and take down' system, some authors have proposed a 'notice and notification' system, whereby the user who is allegedly the offender is notified to take relevant measures or to accept responsibility. This system, they argue, is less detrimental to freedom of expression and user autonomy and does not oblige technology companies to decide on the lawfulness of content. On the other hand, this latter system delays the reaction, which in many cases amounts to rendering ineffective any withdrawal measure that could be taken at a later date" (Sánchez Muñoz, 2020, p. 118).

[99] Aline Osorio argues that moderation under legitimate conditions presupposes, among other things: "(i) the definition of clear and practical application rules; (ii) their orientation towards the fulfillment of non-discriminatory and legitimate objectives in light of the design and purposes of each platform; and (iii) procedural requirements that imply the provision of due process in the case of the adoption of consequences for users and the continuous monitoring of the quality of moderation decisions" (Osorio, 2022, p. 105). In the same vein, "the Transatlantic High Level Working Group on Content Moderation Online and Freedom of Expression [...] has defined a series of basic principles that must be respected in content moderation mechanisms in order to avoid the effect of self-censorship: (1) clearly defining the specific damage to be prevented on the basis of evidence and not conjecture; (2) integrating transparency throughout the procedure, so that the public and all interested parties can accurately assess the impact of moderation; (3) guaranteeing due process so that the authors of the withdrawn material can have an effective recourse to appeal the decision; and (4) taking into account the risks posed by excessive dependence on automated mechanisms and include an adequate number of human reviewers to correct errors" (Sánchez Muñoz, 2020, p. 120).

hand, the issue poses a fundamental problem, since it ultimately means empowering private companies to regulate the traffic of messages on social networks, through the application of "rules that do not have a democratic origin" and via "technical mechanisms which are often automated and not very transparent" (SÁNCHEZ MUÑOZ, 2020, p. 119).

Aline Osorio (2022, p. 105) argues that moderation under legitimate conditions presupposes, among other things: "(i) the definition of clear and practical application rules; (ii) their orientation towards the fulfillment of non-discriminatory and legitimate objectives in light of the design and purposes of each platform; and (iii) procedural requirements that imply the provision of due process in the case of the adoption of consequences for users and the continuous monitoring of the quality of moderation decisions." In the same vein, "the Transatlantic High Level Working Group on Content Moderation Online and Freedom of Expression [...] has defined a series of basic principles that must be respected in content moderation mechanisms in order to avoid the effect of self-censorship: (1) clearly defining the specific damage to be prevented on the basis of evidence and not conjecture; (2) integrating transparency throughout the procedure, so that the public and all interested parties can accurately assess the impact of moderation; (3) guaranteeing due process so that the authors of the withdrawn material can have an effective recourse to appeal the decision; and (4) taking into account the risks posed by excessive dependence on automated mechanisms and include an adequate number of human reviewers to correct errors" (SÁNCHEZ MUÑOZ, 2020, p. 120).

2.2.4 Harassment, Discrimination, and Political Violence

Social-media platforms are designed for users to come together and share content of personal interest to them. These environments appear free and friendly: people feel that they are choosing the content they want to interact with, that their options are unlimited, and that they can decide who to connect with (or not). However, by participating in what appears to be a harmless, safe, and private environment, people provide a large and valuable set of data about themselves and their networks. Users and their data become a product that platforms can, without exaggeration, literally sell (BAKIR; MCSTAY, 2018, p. 155; MOROZOV, 2018, p. 53). This combination between the flow of information and the data it generates makes networks the ideal breeding ground for influential disinformation messages aimed at significant

social identities that can be manipulated according to emotions (Oyserman; Dawson, 2021, p. 185–186).[100]

In the same way that they connect people who know each other and get on well, or who share similar beliefs and tastes, social networks can also be a battleground on which individuals and groups who feel and think differently confront one another, whether over minor and apparently trivial issues or over sensitive and highly relevant questions. In this context, just as digital communities are home to users who are understanding, tolerant, and sympathetic to difference, they are also home to hordes of intolerant, provocative, and uncivilized individuals who are fond of anti-democratic and anti-social practices, as expressed in different forms of discrimination, hatred, and political violence. As Ayesha Siddiqua (2021, p. 325) summarizes, "due to the surprising increase in the number of digital hate groups and cyber-hate activities [...], the space of social networks has become a breeding ground for extremists." In effect:

> Virtual harassment has been a blind spot for the big platforms for many years. The problem became publicized with Gamergate (2014), the first public settling of accounts with harassment of women in the online gaming community. This problem continues to plague social networks, where it advances with giant steps each time incidents of harassment or the silencing of minority voices are made public. However, the problem has grown beyond these applications and has caused new forms of harm. Online harassment has created serious political, technical and structural vulnerabilities that have been exploited by malicious actors and that sometimes go unnoticed – or are not a priority – for the responsible authorities.
>
> (Heller, 2021, p. 19)

[100] Emotion is "a sentiment that tends to distort our perception of reality because it is nourished by non-rational elements, that is, by what is not verifiable. From some time now, emotions have taken over society, and fake news takes advantage of this fact. This emotional or affective turn, which can be seen in different areas of citizenship, is fertile ground for 'communicative democratization', but also for the construction of 'post-truth'. This is why disinformation moves so well in environments where emotion prevails over facts (Vosoughi *et al.*, 2018), as is the case with social networks" (Carratalá; Peris-Blanes, 2023, p. 239).

Digital crime, in this context, is not the result of a single cause, but the combination of several factors, among which should be mentioned, firstly, the very architecture of the networks, which is responsible for the formation of emotional echo chambers that breed intolerance (Fisher, 2022, p. 21)[101] and which are, moreover, the practical result of a political culture based on the constant replication of aggressive behavior which, from a top-down perspective, has "mimetic effects" on the social base (Alvim *et al*., 2023, p. 454). It also derives from the excessive moralization of political debate, which fosters the rise of dogmatic thinking[102] and the multiplication of subjects who: "(a) give absolute priority to their own convictions; (b) deny their opponents any trace of moral integrity; and (c) place the ethics of conviction at the center of political activity, subjecting every debate to a permanent moral judgment" (Vallespín, 2021, p. 51).

In this vein, the massification of aggression as a mirror effect of the behavior of political leaders can be seen, for example, in the evident correlation between the populist strategy of attacking press organizations and professionals adopted by Donald Trump and Jair Bolsonaro, and the corresponding exponential increase in attacks on journalists on social networks in the United States and Brazil (Costa, 2021, p. 110; Magallón Rosa, 2023, p. 118).

---

[101] "A democratic society is a dynamic order that is susceptible to learning, which also presupposes an individual and collective will to admit mistakes, correct historical injustices and forgive one another. A democracy is not simply a dictatorship of the majority, but provides a procedure not only for deciding and electing, but also for debating and deliberating together. It is an order in which anything that is not fair or inclusive enough can and must be readjusted. This also requires a culture of error, a culture of public debate characterized not by mutual contempt, but by reciprocal curiosity. For political actors, recognizing mistakes in their own ways of thinking and acting is as essential as it is for agents of the media and members of civil society. Forgiving oneself sometimes is also part of the moral texture of a living democracy. Unfortunately, structural conditions, as well as the social uses of communication on social networks, are a growing obstacle that prevents this culture of debate in which it would also be possible to recognize one's own mistakes and ask for forgiveness" (Emcke, 2020, p. 168).

[102] In discussions on the web, dogmatic thinking is often revealed in the crystallization of a "binary logic." According to Sakamoto, examples of this superficial and conflictive form of reasoning can be found in this series of false syllogisms: "Did you criticize spending on the World Cup? You don't like football." "Did you take to the streets to protest? You want to overthrow the government." [...] "Are you left-wing? You must take a vow of poverty." "Right-wing? Flog the workers." "Do you defend the homeless? If you don't take them home, you are a hypocrite." "Do you despise male chauvinism? You are a fag" (Sakamoto, 2016, pp. 31–32).

Commenting on the factors that contribute to the radicalization of the contemporary democratic environment, Judge Luís Roberto Barroso (2020, p. 88) considers the presence of political, social, and cultural causes:

> The political causes lie in the crisis of representation of contemporary democracies, in which the electoral process fails to give voice and relevance to citizens. 'They don't represent us' is the slogan of the moment. [...] The economic and social causes lie in the large contingent of workers and professionals who have lost their jobs or have seen their prospects for social advancement reduced, becoming less relevant in the world of globalization, the new knowledge economy and automation, which have weakened more traditional industries and activities. [...] Finally, there are the cultural-identitarian causes, derived from the fact that there is a contingent of people who do not profess the cosmopolitan, egalitarian and multicultural creed that drives the progressive agenda of human rights, racial equality, feminist policies, gay marriage, the defense of indigenous peoples, the protection of the environment and the decriminalization of drugs, among other modernities. These people, who feel disadvantaged or excluded in the world of what is "politically correct," cling to traditional values that give them security and the dream of recovering a lost hegemony.

In the midst of this process, polarization is consolidated and becomes ever more extreme, to the point where political radicalism begins to "spill over into everyday life" (Nunes; Traumann, 2023, p. 14), forming a conspiratorial culture that, in extreme cases, degenerates into traumatic episodes that involve, for example, coup attempts (Alvim *et al.*, 2023, p. 351) and conflicts that can even lead to deaths (Birch, 2021), all this with the backdrop of a visible process of "discrediting elections" (Fachin, 2021).[103] In this manner, the proliferation of fake news not only misleads: it poisons the very fountainhead of democracy (Moore, 2018, p. 13).

[103] "This is of course polarization: a process of delegitimizing the system, but also a process that aims to exclude and delegitimize certain people as members of the political community, under the false pretence of democratic legitimacy. As it goes beyond political fragmentation and in fact can occur without its existence, we speak of the so-called 'participatory postverdad' in which democracy only pretends to be a 'democracy of appearances' in which the emotional prevails over any rational debate" (Martín Guardado, 2023, p. 217).

Recent history has seen emblematic cases of post-election violence. In 2007, the confrontations in Kenya caused 1 113 deaths and more than 3 500 injured and forced more than 350 000 people to leave their homes. In clashes left more than 3 000 people dead and more than a million people had to leave their homes. In 2014, at least 400 lives were lost in Bangladesh, and more than 200 people died in Pakistan in 2018. Not even the so-called most advanced countries are immune. In the European election in the United Kingdom, there were attacks, and the Catalan independence referendum in Spain was also marked by acts of violence (Birch, 2021, pp. 1–2), and we should recall that the storming of the Capitol in the United States after Donald Trump's defeat also resulted in six deaths.

In another example, Trump, in the context of his attempt to return to the White House in early April 2024, said: "The Democrats say 'please do not call them [illegal immigrants] animals, they're human.' I say: 'No, they are not humans, they are animals'" (CNN, 2024). For Michael Sandel: "The populist tradition has for a long time contained two strands: a politics that moved the people against the elites, inequality and irresponsible economic power, and a politics that empassed nativism, racism and antisemitism. Bernie Sanders was inspired by the former. Trump was inspired by both. His hostility toward immigrants, represented by his promise to build a wall on the border with Mexico, echoed the nativist tendency of the populist tradition. And his racist rhetoric recalled the populism of George Wallace, the segregationist Alabama governor who organized a strong independent bid for the presidency in 1968" (Sandel, 2023, p. 383).

The "open hostility to verifiable reality" (Snyder, 2017, p. 64), the "contempt for facts, the substitution of emotion for reason and the corrosion of language are diminishing the value of truth," inflating a "toxic polarization," and injecting intolerance as a "virus" in the political system (Kakutani, 2018, p. 10–27). In this corrosive environment, political interactions are compromised by the growing presence of harassment, intimidation, and discrimination, carried out not only for biopsychological (Fisher, 2022, p. 49; Williams, 2021, p. 14), cultural, and educational reasons but because this negative climate is easily and effectively exploited for political ends (Bruzzone, 2021, p. 78; Marques, 2021, p. 139; Murgia, 2019, p. 96).

> Extreme polarization is a phenomenon umbilically linked to the return of global populism, the concept that [...] places the center of debate in the moral sphere, to the detriment of political platforms and proposals. It can be found in the debate about Hindu nationalism

> in India, the fight against drug trafficking in the Philippines, anti-immigration policies in Western Europe, anti-system candidates in Chile and Argentina and Donald Trump's "Make America Great Again" nativism.
> How have political opinions come to dictate who we live with? Why has tolerance for divergent opinions lasted so little? Why does it seem that elections never end even once the results have been announced? [...]
> These are issues of our time. In *The Bitter End*, [...] John Sides, Chris Tausanovitch and Lynn Vavreck carry out a brilliant autopsy of the 2020 US elections – in which Democrat Joe Biden defeated Republican Donald Trump – proposing the replacement of the concept of partisan polarization with that of fossilization. According to them, political opinions have become petrified and now constitute part of each voter's identity. [...]
> This hardening of the sides leads voters to become fanatical supporters who turn their preferences into what defines their position in society. By assuming that the colors of Lulaism or Bolsonaro are those of someone who supports Flamengo or Fluminense, the voter ceases to worry about a rational comparison of the two politicians and turns their choice into part of their identity. In the same way that a fan's identity does not disappear with the end of the party on the pitch, win or lose, the voter's identity [...] does not expire with the end of the elections.
>
> (Nunes; Traumann, 2023, pp. 14–20)

For conceptual purposes, *harassment* implies "an action that degrades, humiliates or embarrasses a person; it stems from discriminatory motives and has the effect of nullifying or harming an individual in the enjoyment of their rights." Intimidation, in a related vein, "implies the causing of fear or dread, and can manifest itself in the form of physical threats, threatening expressions, emotional manipulation, verbal insults or deliberate embarrassment," with the aim of "preventing an individual from continuing with their work or paralyzing their activities through an attack" (Szabó, 2020, p. 35). On the other hand, *discrimination* is a supremacist (and unacceptable) form of social treatment, in which collectives or members of historically or situationally persecuted groups are subjected to words or attitudes that nullify their dignity and their right to equal treatment (Mitchell, 1970, p. 153). Finally, *political violence* can be understood as a "tool of exclusion" (Birch, 2021, p. 7) that uses symbolic coercion (in the form of threats or

online attacks) to weaken, hinder, or prevent other people's political participation, leading to: (a) silencing, through shame; (b) loss of opportunities due to harassment; or (c) definitive withdrawal as a result of excessive pressure.

The interconnection between radical and extremist individuals and cells raises major concerns (Mulhall, 2021, p. 25). At election times, spirits heat up and violence on the networks causes not only disruptions to normal practices and procedures but also damage has a greater impact on candidates from minority and socially vulnerable groups in a context where "disinformation and hate speech go hand in hand" (Carratalá; Peris-Blanes, 2023, p. 240).[104] The "dark side of networks," in this context, thrives as a strong factor in disaffection, which often leads to committed and promising figures abandoning politics (Wagner, 2020, p. 2).

Deepfakes, for example, are often used in contexts of race (Kim, 2022, p. 52) and gender (Maher, 2022, p. 7; wagner, 2020), operating in the latter case as instruments of blackmail and harassment (Suying, 2022, p. 25), forging pseudo-scandals of a sexual nature that undermine women's chances of political success, and on many occasions, result in the voluntary or involuntary exclusion of female (Ochoa, 2023) or transgender (Siverio Luis, 2023) candidates.[105]

Regardless of the underlying reasons, the Internet is especially vulnerable to avalanches of coordinated attacks involving, among other things,

[104] It should be noted that, in the context of the links between disinformation and hate speech, the idea of hate "does not refer to a psychological state, but to a strategy of political competitors who, although they certainly seek to exploit the emotions of their followers, carry out their strategies in a calculated and cold manner" (George, 2020, p. 146).

[105] "[T]he momentum of feminism in recent years, brought about by global movements such as #NiUnaMenos and #MeToo, has led to a rethink of how the issue of gender equality is approached from a political perspective. But this impulse, and as a reaction, has generated polarization among progressive forces, as well as a reactionary drive against feminism. In fact, from the extreme right, [...] the attacks against the movement are continuous and well-articulated through social networks. In the face of this reaction, minority groups such as those represented by the LGBTQ acronym have not only been among the most active and visible, but have also taken on the role of disseminating feminist values in public opinion, emphasizing one of the movement's central struggles: structural and intersectional violence against women, including trans women. This visibility and participation in the debate has, in fact, outlined a context of confrontation and polarization in which the LGBTQ collective has not only been attacked by conservative and extreme right-wing sectors, but also from a more traditional understanding of feminism that questions the inclusion of trans women in the feminist struggle" (Montagut *et al.*, 2023, pp. 257–258).

doxxing[106] and synthetic attacks for the purpose of harassment (MINOW, 2022, p. 373). Discrimination plays an important role in these spaces, but it is true that in political confrontations, cybercriminals wreak havoc under a partisan or ideological guise. Here, communication as a weapon takes shape in the form of *shitstorms*,[107] following a more or less circular logic: "maintain the escalation [of offenses and accusations], produce new scandals and launch new stories to distract." As Steve Bannon said: "Flood the zone with shit!" (STROBL, 2022, p. 105), attacking opponents with low-quality content and derogatory comments, controlling the narrative and blocking debate (FORTI, 2021, p. 158).

In this context, "trolls"[108] hurl "shameful and provocative messages to disrupt online communities and harass individuals until they leave" (MINOW, 2022, p. 373) as well as "militarize political discourse to drown out others and suppress the voters of the traditional parties" (MOORE, 2018, p. 23). In this setting:

> Online *trolling* is a notorious disruption strategy that consists of publishing provocative content, often including deliberately misleading or meaningless comments, with the intention of provoking other users into engaging in a meaningless debate or fight [...]. *Trolls* amplify spurious and misleading content, since their activity artificially increases support for a public figure or an issue [...][109] Political parties and companies often pay trolls to manipulate virtual interactions for political or commercial purposes.
>
> (SAMOILENKO; SUVOROVA, 2023, p. 512)

---

[106] The practice of "revealing private information [...] in order to promote harassment or put it at risk" (CARDIEL SOTO *et al.*, 2023, p. 81) is known as doxxing, with attacks on candidates and their families being quite common.

[107] "[S]hitstorms invoke a modern form of moral lynching, derived from the combined and insistent publication of outrageous statements, verbalized in unison by a pack of aggressive users, clearly interested in filling the public imagination with negative thoughts about a given individual, collective or institutional objective. [...] like other disinformation and harmful practices, they are often driven by inauthentic behavioral tools" (ALVIM *et al.*, 2023, p. 217).

[108] On the Internet, trolls are users who deliberately send offensive or provocative messages designed to attack, ridicule or offend opponents, followers, and sympathizers, almost always promoting disinformation about whoever they are addressing (CARDIEL SOTO *et al.*, 2023, p. 269).

[109] "In fact, trolling (the deliberate use of offensive attacks to provoke a public reaction and thus set the agenda for debate having amplified the reach of the message) and memes that seek to stir up public opinion would be the hallmarks of the alt-right and alternative right-inspired movements around the world" (PRADO, 2023, p. 99).

Along with trolling, meme campaigns are also used to raise awareness and damage the image of opponents or specific groups, in this case using humor as a harmful tool, capable of ridiculing a target and making them feel embarrassed or uncomfortable (SAMOILENKO; SUVOROVA, 2023, p. 512). In the context of public debates, humor seems to be a great asset in political struggles, among other things because memes, in their apparent celebratory joviality, can be used as an instrument for constructing, spreading, and disseminating disinformation narratives (MELLO, 2020, p. 34). In this environment, healthy and artistic humor is replaced by "verbal violence," materializing under the dishonest guise of "provocation, sarcasm and jokes" (CHARAUDEAU, 2022, p. 68) and often putting the intrinsic morality of the supporters of certain ideologies in doubt (hence the consolidation of offensive identity markers, such as the term "Bolsominion," used to define the followers of Jair Bolsonaro in Brazil).

AI plays a central role in these strategies of aggression and violence, since a large part of the crimes that take place on these networks are based on the automated manipulation of images or videos, which enhances their credibility, and/or distributed with the help of computer propaganda (SAMOILENKO; SUVOROVA, 2023, p. 514), according to segmentation data or psychometric profiles. Technology, in these cases, serves clearly antidemocratic purposes, since exclusionary extremism erodes political pluralism as well as dignity and the principle of equality, and must be vehemently rejected by institutions.

2.2.5 Cognitive Hacking

In the current configuration of societies, the issue of personal data processing and storage is of great importance and can, as Bruno Andrade points out, "impact on the democratic regime."[110] On the one hand, expanding the possibilities of available information enables more effective public policies to be

[110] "AI is fueled by data and the growth of its application and use for the common good depends on it. At present, and until we have a new regulatory context, [...] the lawful processing of personal data requires the express consent of its owner for the specific purposes for which it is to be processed or transferred to third parties, without prejudice to the possibility of other more exceptional purposes" (MUÑOZ VELA, 2022, p. 68), such as the prohibition of unauthorized commercialization, the guarantees of unauthorized commercialization and anonymity, and the possibility of a unilateral withdrawal of consent.

designed, but on the other, it raises sensitive questions about privacy[111] and intimacy[112] as well as about informational self-determination, the free development of personality, and freedom of expression (Souza, 2022, pp. 30–42).

With the discovery of ways to use data-based knowledge to design direct and personalized persuasive approaches – used for the first time by Rodrigo Duterte in 2014 in the Philippines (Moore, 2018, p. 10) but widely publicized with the Cambridge Analytica scandal in 2016 – personal data has become a marketing asset and the basis of the main business model of large social-media platforms. It is available to anyone who can pay for it, whether fashion or technology brands or candidates and political parties, and has become the main element of the strategies offered by election consultants who specialize in marketing and communication.

The digital ecosystem generates an immense volume of personal data under constant scrutiny. While governments, corporations, and social networks engage in massive data collection for analysis based on their interests, much of this information becomes "dark data" – vast pools of disorganized or low-quality data unsuitable for big data applications. With information volume doubling annually, approximately 90% remains unanalyzed. Currently, only 10% of user-generated data undergoes processing for citizen data

[111] "The intrusive convenience of smart technology constantly disrupts personal privacy and reduces the possibility of anonymity. Communication service providers are actively seeking to monetize big data *by* developing new services or selling them to third parties. Most of the data is collected through 'cookies', which allow Internet companies to focus on microtargeting and offer products based on our previous behavior and preferences. As points out, we live in Bentham's panopticon: our digital lives are aquariums into which we enter at will. Today, facial recognition technology can use data mining algorithms and statistical re-identification techniques to identify a person's name, location and even the first five digits of their social security number based on facial features publicly available on social networking sites" (Samoilenko; Suvorova, 2023, p. 509).

[112] "Another concern that marks our time is the threat to privacy, that is, the sphere of people's lives that must be protected against invasion by other individuals, companies or the state. Technological platforms have great potential to violate this right to privacy, since they rely on: the user's personal identification, which includes information such as name, address, marital status, occupation, financial data, tax returns, etc., and information about each person's behavior, preferences, interests and concerns, obtained from online browsing. Various scandals [...] have led countries around the world to pass data protection laws. There have also been attempts to address the risks of economic concentration in technology companies and the fairest way to levy taxes on their activities" (Barroso, 2020, p. 83).

applications, though this percentage increases as AI-driven big data capabilities advance (Rebollo Delgado, 2023, p. 56). Data is the new raw material, the "new petroleum," necessary for the algorithms that decode reality to be incorporated into the business model[113] that has revolutionized the traditional practices of marketing, advertising, and political advertising.

The integration of AI into advertising, particularly in political contexts, enables unprecedented precision in content delivery and persuasion. Advanced systems now analyze browsing data to optimize digital material for maximum impact at precisely calculated moments. On platforms like Facebook, AI systems evaluate various human-created components – text, graphics, and hyperlinks – to determine the most effective combinations for specific individuals at specific times. The technology has evolved beyond mere optimization to autonomous content creation, potentially eliminating human involvement in advertising design. Through detailed personal data analysis and micro-segmentation, these systems achieve both cost efficiency and enhanced effectiveness in influencing target audiences (Bender, 2022, p. 26).

This establishes the ideal circumstances[114] for trained algorithms to calculate scenarios with unprecedented efficiency, making it possible for

[113] In this frame, "all the technological requisites are present for the algorithms to know the probabilities [of your blood test] more precisely than your flesh-and-blood clinician. In the same way that a mobile phone anticipates, as your finger presses the letter 'v' on WhatsApp, that you intend to write 'virus' or 'viral', a system that has access to your laboratory's database and thousands of other databases will be able to project the probable trends of the results. These probable trends will be based on the statistical patterns extracted from the data of your previous tests, combined with the data of millions or thousands of millions of other patients. An accumulation of numbers inside your mobile phone will be able to predict, with a good margin of mathematical certainty, when the terrible diagnosis will appear for you. Why do systems want to know about your future health? Exactly: because this prediction has market value, especially for insurance companies" (Bucci, 2023, p. 55).

[114] "Machine learning based on big data has attracted a lot of interest due to the availability of large amounts of data and the increase in (and decreasing cost of) computing power. Some researchers talk about a 'data tsunami'. We all produce data through our digital activities when we use social networks or buy products online. This data is of interest to commercial agents, but also to governments and scientists. It has never been easier for organizations to collect, store and process data. This is not only due to machine learning; the expanded digital environment and other digital technologies play an important role. It is now cheaper to store data and computers are more powerful. All this has been important for the development of AI in general, but also for data science" (Coeckelbergh, 2022, p. 85).

interested parties to obtain exceptionally valid readings for planning new ways to influence individual behavior. Within the "digitalized market" in particular, reducing uncertainty is worth a fortune for those who "run [or want to run] the political machinery," since "knowing what the other doesn't know" is a very powerful weapon in a competitive setting. Data extraction "generates the perfection of mechanisms for the profitable exploitation of uncertainty" and in this way, "information is a treasure to help the limited few circumvent free competition" (Bucci, 2023, pp. 107–114) in a logic that applies to both the economic sphere and political–electoral disputes.

The intensive exploitation of large amounts of personal data, facilitated by the confluence of AI with big data techniques,[115] has become a "first-class economic resource with a direct impact on the platform economy and the media ecosystem" (Aguado; Martínez, 2023, p. 278). It has been extrapolated, in a generalized way and with no turning back, to the universe of marketing and advertising, both in retail and in the service of political, ideological, and electoral battles. This establishes "a new fundamental type of multifaceted market centered on data processing, a market that brings together platform users who generate data, data buyers (advertisers and data brokers) and platform service providers who benefit from the assignment, sale and internal use of data" (Rieder; Sire; Cohen, cited by Prado, 2022, p. 81). "Nowadays, several companies have begun to specialize in collecting personal data, either to use it directly to improve their business or as an intermediary provider for other companies" (Souza, 2022, p. 23).

The transformation of social media into powerful propaganda tools began with Facebook's 2012 emergence as a dominant advertising platform – not through deliberate manipulation but through profit-driven innovation. The company leveraged its core assets of reach, attention, and personal data to create unprecedented targeting capabilities for advertisers. However, the foundational architecture of surveillance-based advertising originated with Google around 2000, which developed an intricate, automated advertising infrastructure minimizing human oversight while maximizing algorithmic efficiency. This system enabled precise message delivery to any individual

[115] "The Code of Good Practices in Data Protection for Big Data Projects of the Spanish Data Protection Agency uses it to refer both to the 'set of technologies, algorithms and systems used to collect data at a scale and variety not achieved until now' and to the 'extraction of valuable information through advanced analytical systems supported by parallel computing'. We are therefore referring to the collection and processing of a massive volume of data that cannot be processed using conventional methods, to be analyzed using artificial intelligence algorithms and techniques" (Sánchez Muñoz, 2020, p. 20).

worldwide, optimized for engagement at minimal cost. While revolutionary for commercial advertising, this frictionless infrastructure proved problematic for democracy as it failed to distinguish between selling products and propagating ideologies, thus allowing propagandists to target susceptible voters with precisely calibrated messages designed to provoke specific reactions (Moore, 2018, pp. 14–15).

In this context, big data can help answer crucial questions for effective campaign planning, such as: "Who is my voter? What do people think of my proposals? or even Did the voters like the post I published yesterday?" (Rais *et al.*, 2018, p. 74).

> Data mining[116] has become, in this competitive and addictive setting, a fundamental technique for feeding the libraries from which artificial intelligence systems will be nourished. Access to information on one's own population is just as important as denying it to potential adversaries, for various reasons. Firstly, because of its potential to provide intelligence on the behavior of individual actors, as well as on the general context in which they operate. Secondly, because of the possibilities this knowledge opens up for the manipulation of subsequent information directed towards this population. And thirdly, because of what it supposes in terms of denying access to the data of a portion of the population which, in many cases, would provide the diversity needed for the entire database to be coherent.
>
> (Gómez de Ágreda, 2023, p. 212)

2.2.5.1 Psychographic Segmentation

Of course, the division of society into smaller groups for the purposes of segmented communication is nothing new in mass communication.[117] However,

[116] "The identification of patterns in large amounts of information (big data) is sometimes called 'data mining', using an analogy with the extraction of valuable minerals from the earth. However, the term is erroneous, since the aim is to extract patterns from the data, to analyze the data, not [simply] to extract the data itself" (Coeckelbergh, 2022, p. 82).

[117] "Data mining can uncover patterns unknown to human users. Used for a long time for product and brand marketing research, it is now applied (usually through deep learning) to big data: huge collections of text (often in several languages) or images such as scientific reports, medical histories and statements on social networks and the Internet. The applications of big data mining include surveillance and counterintelligence, as well as the monitoring of public behavior by governments, political leaders and social scientists.

AI marks a before and after in the political sphere, as it opens up a more sophisticated and effective range of possibilities by offering "psychographic segmentation" (profiling) as a modern alternative to the previous macroscopic segmentation based on population statistics.

In fact, *demographic profiling* is a well-known technique in sociology and has been used repeatedly by political forces throughout history. It allows large populations to be segmented according to data such as age, education, employment, place of residence, religion, and ethnicity. From demographic profiles, a strategist can develop a campaign that is more targeted at groups such as young people, if they have reason to believe that this segment could be decisive. In addition, demographic analyses make it possible to see that the opponent's electorate is made up of older people with lower levels of education who live in poorer regions and to develop general strategies to reach this segment. However, in a much more developed sense, "*psychographic profiling makes it possible to* focus not only on groups, but on each individual according to whether they are, for example, extrovert or introvert, the articles they buy or the way they react on the Internet to different political programs" (Denemark, 2024, p. 130). Martin Moore (2018, pp. 9–10) explains the importance of these transformations:

> Almost half a century later, political campaigns are almost unrecognizable. Official campaigns are nourished by a central office that provides them with mountains of voter data, subjected to complex algorithmic models and used to send messages distorted with millimetric precision to the most coveted voters. You are no longer an anonymous resident of 43 Belvedere Avenue. Hundreds of "data points" know you because they capture everything you buy, how much you earn, what you read, what you see, who you know and what interests you. Combine all this with the data from opinion polls, and any candidate will know whether they should shower you with attention, ask you for a donation or even discourage you from voting. Unofficial campaigns – set up by private individuals and organizations, pressure groups and us, the uneducated public in

---

These studies can compare the evolution of opinion among different subgroups, men/women, young/elderly, north/south, etc. For example, the British research institute Demos [...] analyzed thousands of Twitter messages related to misogyny, ethnic groups and the police. It is possible to investigate sudden bursts of tweets following specific events ('Twitter incidents') to discover, for example, changes in public opinion regarding the reaction of the police to a specific incident" (Boden, 2020, pp. 91–92).

> general – have changed even more. Now we all have access to such an arsenal of digital tools that we can take up arms and fight for our own message on the same battlefield.

From the moment we enter the "captivity of the platforms" (Morozov, 2018, p. 56), we consent to them being able to use not only our registration information, but also records of "interactions and relationships" that include friends, connections, preferences, and everything related to our activity on the network. "This data, together with the metadata, or data about the data, in addition to being used by the provider to direct mechanisms for personalizing use, can be transferred to third parties, even and especially for commercial purposes." In this context:

> [A] circular system in which the increasingly constant use of social networks supplies providers with large amounts of data that allow them to personalize use according to user preferences, increasing the degree of use upon generating more data.
>
> (Biolcati, 2022, p. 123)

The first problem is that this data processing is mainly carried out "using techniques that remain hidden, beyond the control of interested parties and far outside any state supervision," presumably under the pretext of safeguarding commercial secrets protected by industrial production patents. Under the pretext of the legal protection of business activity, infinite amounts of data collection and data use are concealed on a daily basis, thereby escaping any kind of effective control, whether from an ethical or legal perspective. From a regulatory point of view, the establishment of mechanisms to guarantee greater transparency and, in particular, clear knowledge of the purpose of these processes, is therefore imperative (Rebollo Delgado, 2023, p. 109) so that informational self-determination does not become an empty aspiration.

The second concern stems from the fact that taking into account the secrets hidden behind the data makes it possible, according to modern AI techniques, to develop new forms of manipulation and control that are personalized, targeted, and almost unavoidable. What happens is that these tools allow "recipients to be segmented thanks to the processing of a large amount of data," which in turn creates "the possibility of sending personalized messages to different groups of people, messages that can be different or even contradictory" (González-Torre, 2020, pp. 67–68) as well as harmful, incendiary, and false.

The third problem is that information manipulation, along these lines, is based on narrative actions that are commonly associated with the exploitation of weaknesses, beliefs, and prejudices, thus establishing connections with limited rationality and a high level of stimulation of the emotional apparatus which often exploit the dynamics of social panic linked to ties forged due to "loss aversion bias" (Garrigues Walker; González de la Garza, 2020, p. 91). Radical and extremist candidates and parties use and abuse attractive games that "fantasize existential dramas," placing antagonistic currents or identities in a threatening relationship (Del Rey Morató, 2007, p. 171), often in association with campaigns of hate and prejudice (George, 2020, p. 146).[118]

The Cambridge Analytica case exemplifies sophisticated manipulation of social network communications through advanced psychological profiling. Their methodology combined psychographic analysis with behavioral psychology to predict and influence voter behavior by exploiting cognitive biases. Through detailed psychological profiling, the company identified susceptible individuals for targeted influence campaigns, using selective information distribution and disinformation techniques, as revealed through whistleblower testimonies and British media investigations (González-Torre, 2020, p. 67).

It is crucial to note that GenAI is increasingly employed to develop more credible and targeted *phishing campaigns* on a large scale, with the aim of crafting personalized disinformation messages to influence electoral processes (Small, as cited by Eddy, 2024). Phishing, in this context, refers to the practice of sending fraudulent messages that appear to come from a reputable source and are designed to manipulate individuals into disclosing sensitive information or engaging in actions that compromise personal or organizational security. While traditional phishing relied on generic

[118] "Hate campaigns are a communication strategy used by groups involved in territorial conquests, genocides, pogroms and other crimes against humanity, and have been used throughout history. In less extreme cases, they attack vulnerable minorities to activate fear and identity politics as a means of accumulating power. Hate campaigns invariably incorporate disinformation, that is, deliberately misleading content ranging from lies and falsehoods to the use of misleading framing, selectivity in the presentation of events and the misrepresentation of context to show people and events in a false light. Few forms of discourse hit the public harder than hate propaganda. When it comes to this type of expression, liberal faith in the marketplace of ideas tends to be misguided. The communities attacked tend to be politically, economically and culturally fragile and therefore incapable of defending their rights against adversaries who wish them harm" (George, 2020, p. 146).

deceptive messages, modern AI-powered approaches demonstrate an unprecedented ability to mimic legitimate sources and create compelling narratives that exploit fundamental aspects of human psychology – particularly our inherent trust in familiar sources and our instinctive response to perceived threats. Perhaps most concerning is the emergence of *spear phishing* as a refined weapon in this digital arsenal. These attacks leverage AI systems to methodically harvest and analyze network data, precisely replicating language patterns, communication styles, and voice characteristics. The resulting fraudulent communications demonstrate such sophistication that they can effectively breach even well-established security protocols and manipulate informed decision-making processes (Muñoz Vela, 2022, p. 64).

2.2.5.2 The Use of Fear

In this context, psychometric secrets enable the construction of linguistic blackmail that exploits the weaknesses of the psychological–sensory apparatus, and fear in particular (Prado, 2023, p. 160).

These forms of discursive blackmail, generated by AI, appeal, among other possibilities, to the discourse of the enemy, to the rhetoric of elimination, and to the language of fear; in other words, they appeal to the repeated stimulation of collective fears and paranoia which, among other historical experiences, opened the door to Nazism (Levitsky; Ziblatt, 2023, p. 34). Modern campaigns end up contaminated by an excessive load of negative emotions, expressed not only in practices that foment hatred and acrimony but also in minor manifestations of disgust, ridicule, and mockery. This is contrary to the dignity necessary in politics and, due to their cumulative effects, are equally capable of replacing rational decision-making with a weakened system dominated by group-think, pack behavior, and collective hysteria.

Under these conditions, the political sphere regresses to a system that promotes segregation and instability and which profits from a climate of mutual distrust, systematic opposition, and escalating conflicts. This is perhaps, to use an emblematic expression, the broth in which the "unhappy consciousness of the neo-reactionary" is prepared (Hui, 2020, p. 49).

> The use of references based on emotions represents a substantial change in argumentation. Persuasion is no longer based on arguments, facts and figures, but on the ability to generate a discourse that interests the audience. Furthermore, disinformation appeals to stories that polarize the audience, to nostalgic arguments based on historical traumas and other rational mechanisms that have multiplied. The

> emotional approach promotes polarization [...] without nuances. [For this reason] In the fight against disinformation, the production of noise and objective information is not enough [...]. Disinformation acts as a social parasite that radicalizes pre-existing problems. [...] Imprecision in terms of the possibility of 'hurling poison' contaminates the source of information[119]. The population becomes infected because it shares everything on social networks and becomes self-convinced of the informative value of a tweet [...]. It is not necessary to intervene in the entire information chain, it is sufficient to introduce a few messages into the public conversation.
>
> (Manfredi Sánchez *et al*., 2023, p. 197)

2.2.5.3 Bespoke Cheating

In the field of cognitive hacking, it should also be borne in mind that, in addition to attractive messages, AI algorithms can be adjusted to boost false messages adapted to the orientation (political, ideological, religious, etc.) of specific audiences (Garriga *et al.* 2024, p. 180) in a context in which disinformation increasingly has greater persuasive power.[120] An analysis of

[119] Obviously, however, smokescreens can use other agendas and formats, operating fundamentally from any medium or agenda capable of causing shock, curiosity, or emotional reactions, as is the case with memes and viral content of all kinds. Continuing with this subject, Campos Mello recalls that, in authoritarian populism, the creation of fake news (factoids) and attacks on journalists have often been used as smokescreens (Mello, 2020, pp. 111–112).

[120] "Correctly executed disinformation cannot, therefore, be based on a mere alteration of punctual data about a situation, but must be based on individualized knowledge of the target audience, their prejudices and assumptions, and on a prior construction of the scenario in which the events are presented. Gathering data on audiences, configuring prior narratives to support the disinformation of states of opinion are activities that require not only a long period of time before act of disinformation, but also an environment conducive to the correct reception of the messages. Therefore, they are activities that necessarily have to precede the outbreak of hostilities between the contenders" (Gómez de Ágreda, 2023, p. 220). "Studies carried out in the field of computational social science conclude that you can get to know an individual's personality based on how they express themselves, their actions, the character of what they publish, how much and how often they do it, or how they interact, among other infinite options. All this digital trail or footprint brings about the configuration of what is known as a *digital self* or *virtual personality* that is manufactured, not from what we really are and our singularities, but from what mathematical algorithms and big data consider us to be" (Rebollo Delgado, 2023, pp. 55–56).

vulnerabilities using cookies makes it possible to manufacture tailor-made deception oriented around topics carefully selected to trigger behaviorist responses that affect the unique and personal capacities of each individual. The aggravating factor is that at this point, algorithmic AI can be reinforced with the advanced and creative ingenuity of GenAI.

The manipulation of personal opinions increasingly relies on microtargeting – a sophisticated form of personalized advertising driven by algorithmic analysis of users' digital footprints, including browsing history, search patterns, and social-media activity. University of Cambridge research demonstrates the profound depth of personality insights obtainable through social-media engagement: a user's Facebook "likes" can reveal crucial demographic and ideological characteristics, with 300 likes potentially providing deeper personal insights than those held by intimate partners. While this technology enables legitimate electoral advertising, it also facilitates rapid dissemination of disinformation, often outpacing correction mechanisms and thereby threatening electoral integrity (Lins, 2023, pp. 290–291).

It is sufficient to recall that, in general terms (a) big data algorithms[121] make it possible to read and understand large amounts of data relating to personal tastes, preferences, and behavior, which in turn makes it possible to use profiling and micro-segmentation techniques to design and send targeted messages that exploit feelings such as fear, prejudice, and other conditions or vulnerabilities that influence voting; (b) the algorithms that personalize search engine results, select and recommend content, and organize feeds and timelines profoundly and artificially alter the flow of communicative actions, ultimately prioritizing political expressions that will obtain

[121] "We are referring to: a large amount of data about people (their beliefs, their preferences, their concerns, their habits, their fears, their hopes...); formulas that allow this data to be analyzed and classified, as well as, based on this, the personal profiles of the subjects, in order to segment the recipients; and finally, the possibility of using the results of these tactics to send messages, in a massive way, through the network, for a political purpose. Let's see how it works: the process would begin with the collection of a large amount of data about people (big data); then there would be a processing of this data to group and segment the population, which would be carried out by artificial intelligence using algorithms; then, taking advantage of the knowledge acquired about these groups of people, there would be a selection of content and topics (issues) that would be sent to these – already segmented – recipients through the new media, always with the aim, rather than informing or debating, of motivating their behavior; and all of this [....] with the intention of achieving a political goal. Motivate political behavior, create moods, breeding grounds, get out the vote in elections, benefit or damage political figures, etc." (González-Torre, 2020, p. 57).

greater reach and visibility over statements and ideas that will fall on deaf ears, and thus invariably determining the final destination of the messages and the fate of the respective messengers; and (c) bots, acting in concert, can flood networks with false or hate-filled narratives, launch harassment and humiliation campaigns, and divert public attention from relevant issues with false flags and smokescreens.

In summary, it can be concluded that GenAI is revolutionizing the industry of persuasion, implementing techniques to convince people that are free of traditional operational limitations, and have a degree of effectiveness never seen before.

2.2.5.4 The Protection of Privacy

Technology that affects privacy can have direct effects on voting, especially in terms of its exercise in a free, conscious, and (well) informed manner,[122] in contact with ideas that flow in a plural environment. Aside from its associated risks, AI reduces costs and progressively increases its accuracy and efficiency, all of which drives its use as a widespread trend (Hernández Peña, 2022, p. 59).

Electoral normalcy has already been compromised by an exacerbated belligerence that stems, in part, from the fragmentation caused, among other things, by these segmentation processes. In today's world, not only people, but also discourse itself is habitually segmented. The "loss of a global discourse [...] threatens the necessary universal vision of society" and with it, "the very idea of the general interest, which is undoubtedly one of the essential concepts for the articulation of any social and political system." If

[122] "The current system of data extraction [...] can jeopardize an important pillar of democracy, which is access to the facts on which voters' rational processes of choice and deliberation of depend. Therefore, the manipulation of information described in the previous sections has very perverse effects on democracy. However, in addition to information manipulation, other strategies have been used to compromise or nullify people's own free will, exploiting their subconscious so that they are not even aware of what is happening. So, while *fake news* and information manipulation can erode democracy because they make it difficult to make rational decisions, enhanced by bias and the limitations of rationality, digital manipulation can erode democracy by suppressing people's own autonomy and decision-making capacity. [...] At the end of the day, digital manipulation techniques allow the use of a power that could not even be considered persuasive, since its ultimate goal is to have total control of people's minds, not exactly through persuasion, but through complete subjugation" (Frazão, 2022, p. 567).

discourse centered on general interest is replaced, "another essential notion, that of social cohesion, will also be automatically threatened." Therefore,

> [W]ith the triumph of these practices, design to influence political ideas and behavior, shared public debate is replaced by a plurality of partial debates, with the idea that these are the ones that best determine how people vote, and this phenomenon of segmenting the public message, apart from questioning essential notions, is related to, and takes to an extreme, the heralded fragmentation of modern social conflict and connects to the crisis of the grand liberation narratives.[123]
>
> (González-Torre, 2020, p. 68)

The "cookie apocalypse" brings about the "feudalization of the Internet," undermining privacy and configuring a new landscape in the information circulation ecosystem within which pluralism and rational debates end up spinning out of orbit due to the effect of personalized approaches that erode the egalitarian dimension of campaigns, boosting disinformation and the dynamics of exclusion and prejudice in connection with a model of privilege promoted by the worrying advance of machine learning (Aguado; Martínez, 2023, p. 275).

In short, as has already been pointed out, data-driven politics can jeopardize the integrity of elections:

> In the new digital campaigns, data has become an essential asset and control over its use has therefore become a priority in order to guarantee free elections. Today, electoral campaigns have become "data-driven campaigns" and electoral competitors are systematically relying on data to have more and better information about voters, to communicate with them more effectively and to better evaluate their campaign activities. Information, communication and evaluation are, therefore, three pillars on which data-driven campaigns are based, although it is undoubtedly the first two that have the most problematic impact on equal opportunities [...].

[123]The author adds that the idea of human dignity can also be affected by processes of segmenting individuals and points out that, based on this premise, some bodies, such as the European Union's Group on Ethics in Science and New Technologies, are already debating limiting the classification of people based on autonomous systems and algorithms, especially when those affected have not given their consent. There is talk of the "right not to be measured, profiled, analyzed, harrassed, induced or nudged." Other rights examined along the same lines are the "right to receive information without interference" and the "right to be free from surveillance" (González-Torre, 2020, p. 75).

> Used effectively, data can help [...] to better understand the voters, which allows [candidates] to address them in an individualized way by talking to them about the issues that interest them most and, therefore, enables a much more efficient use of economic resources. When big data analysis tools are further refined with the assistance of artificial intelligence, electoral competitors will be in a position to know how each individual thinks and, above all, feels, that is, they will be able to know which emotional keys need to be pressed in order to condition their behavior in one direction or another. [...] Where there is power, there can be abuse, if the measures to prevent it are not put in place.
>
> (Sánchez Muñoz, 2020, pp. 90–91)

According to Óscar Sánchez, there are two possible courses of action to guarantee a minimum equilibrium. The first is to strengthen the right to data protection through legislative reforms that also include measures to maximize transparency in the processes of data collection, analysis and use. The second is the mitigation of positions of control, in particular "to prevent control over the data of millions of people from being concentrated in the hands of a single company, or a very small group of them," bearing in mind that this new "datopoly" is currently one of the world's greatest dilemmas (Sánchez Muñoz, ibidem).

2.2.6 The Dispersion of Control

In light of the foregoing, new technology poses serious threats not only to electoral processes, but also to the "proper functioning of democracy itself," since democratic regimes require not only that all citizens have the right to vote but also that "they are able to express their preferences in a deliberation process governed by free and informed debate on the main issues at stake." From this point of view, John Tasioulas (2019, p. 87) expresses his concern that democratic processes in the current climate could be compromised by technological activities that are undetectable or imperceptible, whether these involve the manipulation of illicitly obtained data, the construction of deepfakes, or the robotic distribution of disinformation content at different levels.

Due to their characteristics, the architecture of networks relies on economics, omnipresence, ease of access, and above all, anonymity to bring together "in their ranks libertarians and neo-Nazis, paleoconservatives, supremacists and openly misogynist groups" as well as racists, extremists, radicals, and authoritarians of all kinds "who seek to march on parliaments and broaden the acceptance of [....] anti-liberal agendas" (Prado, 2022, p. 99), anti-democratic agendas, or in the same direction, destabilizing

agendas with economic, ideological, or geopolitical objectives. In this terrain, various "armor-plating tactics" (Alvim *et al.*, 2023, p. 219) are employed to provide attackers with a safe haven (Caiani; Susánsky, 2021, p. 175) in which to carry out their subversive activities without fear of the law.

AI also plays a role in these issues since it makes remote cyberattacks possible as well as providing solutions that undermine accountability by adding various layers of anonymity and impunity. In addition, new technology facilitates hidden and imperceptible forms of manipulating content selection and suggestion algorithms, thus altering the information supply system. Search engine optimization (SEO) engineering, for example, cracks Google's secret codes to ensure the hegemony for certain websites, blogs, news, and publications in the search results ranking. In the same vein, generative applications can reinforce clickbait practices, giving false and harmful content tempting titles that will attract many visits and then be further boosted by the networks' own algorithms (Gillespie, 2020, p. 335) in an unwholesome feedback loop.

Opaque strategies – as José Manuel Muñoz explains – affect the pillars of governance and, in a related way, democratic values (Muñoz Vela, 2022, p. 55). Bearing in mind that electoral integrity depends, above all, on the "general imperative of respect for the legal order" (Alvim, 2016, p. 233), it is understood that the use of AI, by interfering with natural control mechanisms, endangers the constitutional principles that guarantee society's proper development. The transnational nature of network platforms exposes democratic protection to some "system failures" (Moore, 2018, p. 149), and the anarchic conception of the cybersphere compels us to seek formulas capable of protecting the rule of liberty, both within and outside the electoral sphere.

When this is the case, the courts must remember that "constitutional principles, considered as the backbone of state order, must remain in force at all times" and cannot be violated, not even in a "partial or insignificant" way. The basic principles of electoral discipline, in this framework, have a "permanent degree of primacy," and therefore, compliance with them conditions the lawfulness of political practices and the very validity of the results (Del Rosario Rodríguez, 2019, p. 15).

At this point we come up against the difficulty that official bodies, especially electoral bodies, have in providing an effective response to this type of threat. Their technical limitations – and the impossibility of enforcing their regulations in the virtual space, which resists submitting to national laws – produces a technological dependence on the large platforms in which, faced with the material impossibility of offering solutions, they seem to place their hope and trust. This involves a sort of privatization of part of the electoral justice process, which is often justified by the absence of alternatives but which has serious consequences for electoral rights and legal security.

BIBLIOGRAPHY

Adami, Marina. How AI-generated disinformation might impact this year's elecions and how journalists should report it. *Reuters Institute*, 15 de marzo de 2024. Available at: [How AI-generated disinformation might impact this year's elections and how journalists should report on it | Reuters Institute for the Study of Journalism (ox.ac.uk)]. Viewed: 28-03-2024.

Aguado, Juan Miguel; Martínez, Inmaculada J. Inteligencia artificial y privacidad: la transformación de la publicidad digital y su impacto en el ecosistema de medios. In: Vázquez-Barrio, Tamara; Salazar García, Idoia (eds.). *Inteligencia artificial, periodismo y democracia*. Valencia: Tirant lo Blanch, 2023, p. 265–282.

Aguado Terrón, Juan Miguel; Villaplana Jiménez, F Ramón. Guerras culturales, desinformación y moralización del discurso público. In: Carratalá, Adolfo; Iranzo-Cabrera, María; López-García, Guillermo (eds.). *De la desinformación a la conspiración: política y comunicación ante un escenario híbrido*. Valencia: Tirant lo Blanch, 2023, p. 201–216.

Alperovitch, Dmitri. Bears in the Midst: Intrusion into the Democratic National Committee. *Crowdstrike Blog*, June 15, 2016. Available at: [https://cyber-peace.org/wp-content/uploads/2018/11/Bears-in-the-Midst_Intrusion-into-the-Democratic-National-Committee-%C2%BB.pdf]. Viewed: 02/04/2024.

Álvarez, Antón. IA y mediación de los algoritmos: las iniciativas de autorregulación de las plataformas digitales ante conflictos democráticos. In: Vázquez-Barrio, Tamara; Salazar García, Idoia (eds.). *Inteligencia artificial, periodismo y democracia*. Valencia: Tirant lo Blanch, 2023, p. 171–185.

Alvim, Frederico Franco. *Curso de Direito Eleitoral*. 2. ed. Curitiba: Juruá, 2016.

Alvim, Frederico Franco; Zilio, Rodrigo López; Carvalho, Volgane Oliveira. *Guerras cognitivas na arena eleitoral: o controle judicial da desinformação*. Rio de Janeiro: Lumen Juris, 2023.

Aminulloh, Akhirul; Artaria, Myrtati; Surya, Yuyun Wahzu Izzati; Qorib, Fathul; Hakim, Lukman. Firehose of falsehood model in the 2019 Indonesian Presidential Election. *Jurnal Komunikasi*, 15(2) (2022), p. 249–263.

Amorim, Gabriel. Candidatas à prefeitura de São Paulo têm rostos inseridos em conteúdo adulto feito com IA. **Desinformante,** 25 September 2024. Available at: [https://desinformante.com.br/timeline/candidatas-prefeitura-sp-deepnudes]. Viewed: 10.29.2024.

Andrejevic, Mark. The political function of fake news: disorganized propaganda in the era of automated media. In: Zimdars, Melissa; McLeod, Kembrew (eds.). *Fake news: understanding media and misinformation in the digital age*. Cambridge: The MIT Press, 2020, p. 19–28.

Aranha, Glaucio. Flaming e cyberbullying: o lado negro das novas mídias. *Ciberlegenda*, 34 (2014), p. 122–133.

Arias Maldonado, Manuel. *Democracia Sentimental: política y emociones en el siglo XXI*. Madrid: Página indómita, 2016.

Arriagada, Eduardo. *Hiperconectados: Cómo comunicarse en el siglo xxi*. Santiago de Chile: Forja, 2023.

Athaluri, Sai Anirudh; Manthena, Sandeep Varma; Kesapraga, Krishna Manoj; Yarlagadda, Vineel; Dave, Tirth; Duddumpudi, Rama Tulasi Siri. Exploring the boundaries of reality: investigating the phenomenon of artificial intelligence hallucination in scientific writing through ChatGPT references. *Cureus*, 15(4) (2023), p. 1–5.

Bakir, Vian; Mcstay, Andrew. Fake news and the economy of emotions. Problems, causes, solutions. *Digital Journalism*, 6(6) (2018), p. 154–175.

Balkan, Ekin; Ülgen, Sinan. A primer on mimsinformation, malinformation and disinformation. *Cybergovernance and Digital Democracy Programme*, 2023. Available at: [Report-Disinformation-Malinformation-Misinformation.pdf (edam.org.tr)]. Viewed: 25-03-2024.

Balocco, Anna Elizabeth. O flaming (ou violência verbal em mídia digital) e suas funções na esfera pública. *Linguagem em discurso*, 16(3) (2016), p. 503–521.

Barcellos, Ana Paula de; Terra, Felipe Mendonça. Liberdade de expressão e os desafios da democracia digital. In: Branco, Paulo Gustavo Gonet; Fonseca, Reynaldo Soares da; Branco, Pedro Henrique de Moura Gonet; Velloso, João Carlos Banhos; Fonseca, Gabriel Campos Soares da (eds.). *Eleições e democracia na era digital*. Brasília: Almedina, 2022, p. 263–285.

Barroso, Luís Roberto. *Sem data venia: um olhar sobre o Brasil e o mundo*. Rio de Janeiro: História Real, 2020.

Bauman, Zygmunt. *Estranhos à nossa porta*. Rio de Janeiro: *Zahar*, 2017.

Bechis, Francesco. Playing the Russian disinformation game. Information operations from Soviet tactics to Putin's sharp power. In: Giusti, Serena; Piras, Elisa (eds.). *Democracy and fake news. Information manipulation and posttruth politics*. New York: Routledge, 2021, p. 118–131.

Beiguelman, Giselle. *Políticas da imagem: vigilância e resistência na dadosfera*. São Paulo: Ubu, 2021.

Bender, Sarah M L. Algorithmic elections. *Michigan Law Review*, 121(3) (2022), p. 489–522.

Bermejo Barrera, José Carlos. *La política como impostura y las tinieblas de la información*. Madrid: Ediciones Akal, 2021.

Bicu, Ingrid. Challenges for electoral officials in the information environment around elections. IDEA International, 31 de janeiro de 2024. Disponível em:

[https://www.idea.int/theme/information-communication-and-technology-electoral-processes/election-officials-challenges-information-environment-around-elections]. Viewed: 04.03.2024.

Biolcati, Fernando Henrique de Oliveira. Eleições e a importância do engajamento dos provedores de redes sociais no controle das fake news. In: Branco, Paulo Gustavo Gonet; Fonseca, Reynaldo Soares da; Branco, Pedro Henrique de Moura Gonet; Velloso, João Carlos Banhos; Fonseca, Gabriel Campos Soares da (eds.). *Eleições e democracia na era digital*. Brasília: Almedina, 2022, p. 121–144.

Bioni, Bruno; Almeida, Virgilio; Mendes, Laura Schertel. Inteligência artificial e ameaça a integridade das eleições. *Folha de São Paulo*, 17 de febrero de 2024. Available at: [https://www1.folha.uol.com.br/ilustrissima/2024/02/inteligencia-artificial-e-ameaca-a-integridade-de-eleicoes.shtml]. Consultado: 19-02-2024.

Birch, Sarah. *Electoral violence, corruption and political order*. Princeton: Princeton University Press, 2021.

Boden, Margaret A. *Inteligência artificial: uma brevíssima introdução*. São Paulo: Unesp, 2020.

Bond, Shannon. 2024 elections are ripe targets for foes of democracy. **NPR,** 29 December 2023. Available at: [https://www.npr.org/2023/12/29/1220087754/2024-elections-targets-foes-democracy-disinformation]. Viewed: 31.10.2024.

Borja, Rodrigo. *Enciclopedia de la Política*. Ciudad de México (1997).

Bowman, Nicholas David; Cohen, Elizabeth. Mental shortcuts, emotion, and social rewards. In: Zimdars, Melissa; McLeod, Kembrew (eds.). *Fake news: understanding media and misinformation in the digital age*. Cambridge: The MIT Press, 2020, p. 223–233.

Bradshaw, Samantha; Howard, Philip N. The global organization of social media disinformation campaigns. *Journal of International Affairs*, 71(1) (2017), p. 23–32.

Bruzzone, Andrés. *Ciberpopulismo: política e democracia no mundo digital*. São Paulo: Contexto, 2021.

Bu, Zhan; Xia, Zhengyou; Wang, Jiandong. A sock puppet detection algorithm on virtual spaces. *Knowledge-Based Systems*, 37 (2013), p. 366–377.

Bucci, Eugênio. *Incerteza, um ensaio. Como pensamos a ideia que nos desorienta (e orienta o mundo digital)*. Belo Horizonte: Autêntica, 2023.

Bump, Philip. Trump demonstrates why Artificial Intelligence Will be so usuful to him. *The Washington Post*, 13 de marzo de 2024. Available at: [Trump

demonstrates why Artificial Intelligence will be so useful to him – The Washington Post]. Consultado: 22-03-2024.

Burguera, Leyre; Cobacho, Angel. El derecho al olvido de los políticos en las campañas electorales. In: Corredoira y, L; Cotino, L (eds.). *Libertad de expresión e información en internet: amenazas y protección de los derechos personales*. Madrid: CEPC, 2013, p. 501–520.

Caiani, Manuela; Susánsky, Pál. Radical-right political activism on the web and the challenge for European democracy. A perspective from Eastern and Central Europe. In: Giusti, Serena; Piras, Elisa (eds.). *Democracy and fake news. Information manipulation and posttruth politics*. New York: Routledge, 2021, p. 173–187.

Caldarelli, Guido; De Nicola, Rocco; Petrocchi, Marinella; Saraco, Fabio. Information spreading and the role of automated accounts on twitter. In: Giusti, Serena; Piras, Elisa (eds.). *Democracy and fake news. Information manipulation and posttruth politics*. New York: Routledge, 2021, p. 157–172.

Caldevilla Domínguez, David. Efectos actuales de la "sobreinformación" y la "infoxicación" a través de la experiencia de las bitácoras y del proyecto ID avanza "Radiofriends". Revista de Comunicación de la SEECI, March 2013, v. 17, n. 30, pp. 34–56.

Calvo, Dafne; Campos-Domínguez, Eva; Díez Garrido, María. Hacia una campaña computacional: herramientas y estrategias online en las elecciones españolas. *Revista Española de Ciencia Política*, 51 (2019), p. 123–154. https://doi.org/10.21308/recp.51.05

Campos-Domínguez, Eva; Redondo García, Marta; Cala Siria, Reyes; Rodríguez Pallares, Miriam; Fiuri, Érika; Risueño, Iván. La organización y estrategia de la cibercampaña de los partidos españoles: 2015–2016. In: Dader, José Luis; Campos Domínguez, Eva (eds.). (coords.)*La búsqueda digital del voto. Cibercampañas electorales en España 2015–16*. Tirant lo Blanch, 2017, p. 229–299.

Canedo, Aldo. IA: "voice cloning", la neva arma en la guerra sucia electoral. *La Verdad*, 3 de noviembre de 2023. Available at: [IA: Voice Cloning, la nueva arma en la guerra sucia electoral (laverdadnoticias.com)]. Consultado: 25-03-2024.

Cardiel Soto, Roberto Heicher; Alvim, Frederico Franco; Rondon, Thiago Berlitz. *Glosario contra la Desinformación*. Ciudad de México: Instituto Nacional Electoral, 2023.

Carrascón, Ignacio. Operaciones encubiertas y usos maliciosos de la IA para influir en las elecciones. **Newtral**, 7 October 2024. Available at: [https://www.newtral.es/ia-influencia-elecciones/20241007]. Viewed: 10.08.2024.

Carratalá, Adolfo; Peris-Blanes, Àlvar. (Des) 'infroentetenimiento' en los magazines televisivos de actualidad: sesgos y bulos a propósito de la 'ley trans'. In: Carratalá, Adolfo; Iranzo-Cabrera, María; López-García, Guillermo (eds.). *De la desinformación a la conspiración: política y comunicación ante un escenario híbrido*. Valencia: Tirant lo Blanch, 2023, p. 237–254.

Catalano, José, Victor Rodrigues; Lorenzi, Bruno Rossi. Sem referências: o ChatGPT sob a perspectiva latouriana do duplo clique. *Faz Ciência*, 25(41) (2023), p. 38–58.

Cazorla, Ángel. Bandwagon (efecto). In: Crespo Martínez, Ismael; D'adamo, Orlando; García Beaudoux, Virginia; Mora Rodríguez, Alberto (eds.). *Diccionario enciclopédico de comunicación política*. Madrid: Centro de Estudios Político y Constitucionales, 2015, p. 36–37.

Charaudeau, Patrick. *A manipulação da verdade. Do triunfo da negação às sombras da pós-verdade*. São Paulo: Contexto, 2022.

Citron, Danielle K; Chesney, Robert. Deep fakes: a looming challenge for privacy, democracy, and national security. *California Law Review*, 107(12) (2019), p. 1753–1820.

CNN. Trump chama imigrantes de "animais" e intensifica foco em imigração ilegal. Available at: [https://www.cnnbrasil.com.br/internacional/eleicoes-nos-eua-2024/trump-chama-imigrantes-de-animais-e-intensifica-foco-em-imigracao-ilegal/]. Consultado: 4-4-2024.

CNN, AI "ressurrects" long dead dictator in murky new era of deepfake electioneering. Available at: [https://edition.cnn.com/2024/02/12/asia/suharto-deepfake-ai-scam-indonesia-

Coeckelbergh, Mark. The Political Philosophy of AI, Cambridge, UK: Polity Press, 2022.

UOL Confere. Vídeo é adulterado para sugerir que Lula estava bêbado em discurso na Bahia. *UOL*, 17 de febrero de 2023. Available at: [Vídeo é adulterado para sugerir que Lula estava bêbado em discurso (uol.com.br)]. Consultado: 25-03-2024.

Costa, Caio Túlio. Politização, polarização e o futuro do jornalismo profissional. In: Costa, Caio Túlio; Tardáguila, Cristina; Barreto, Luciana; Celestino, Helena; Amaral, Marina; Pereira, Merval; Bial, Pedro (eds.). *Tempestade perfeita: sete visões da crise do jornalismo profissional*. Rio de Janeiro: História Real, 2021.

Da Empoli, Giuliano. *Os engenheiros do caos*. São Paulo: Vestígio, 2020.

Del Rey Morató, Javier. *Comunicación política, Internet y campañas electorales. De la teledemocracia a la ciberdemocr@cia*. Madrid: Tecnos, 2007.

Del Rosario Rodríguez, Marcos. Nota introductoria. In: Manuel Acuña, Juan (ed.). *Invalidez de elecciones por violación de principios constitucionales*.

Ciudad de México: Tribunal Electoral del Poder Judicial de la Federación, 2019, p. 15–19.

Denemark, Jaroslav. The risk of artificial intelligence for democracy and the EU's first efforts tu regulate it. *The Lawyer Quarterly*, 14(1) (2024). Available at: The Lawyer Quarterly (cas.cz). Viewed: 22.03.2024.

Dias, Patrícia. From "infoxication" to "infosaturation": a theoretical overview of the cognitive and social effects of digital immersion. **Ambitos Comunicacion,** 2014. Available at:[https://repositorio.ucp.pt/bitstream/10400.14/14939/1/PD_Infoxication_2014.pdf]. Viewed: 20.03.2024.

Dörr, Konstantin Nicholas. Mapping the field of Algorithmic Journalism. *Digital Journalism*, 2015. Available at: [https://www.researchgate.net/publication/282642995_Mapping_the_field_of_Algorithmic_Journalism]. Consultado: 23-01-2024.

Dubini, Martina. Deepfake porn: Cuando la IA se utiliza para violentar a las mujeres en política. **Conversaciones de APC,** 16 August 2024. Available at: [https://www.apc.org/es/blog/deepfake-porn-cuando-la-ia-se-utiliza-para-violentar-las-mujeres-en-politica]. Viewed: 1.11.2024.

Eddy, Nathan. Deepfake Democracy: AI Technology Complicates Election Security. **Dark Reading,** 9 February 2024. Available at: [https://www.darkreading.com/application-security/deepfake-democracy-ai-technology-election-security]. Viewed: 14.02.2024.

Elliot, Victoria; Kelly, Makena. The Biden deepfake robocall is only the beginning. *Wired*, 24 de enero de 2024. Available at: [https://www.wired.com/story/biden-robocall-deepfake-danger/]. Consultado: 24-01-204.

Emcke, Carolin. *Contra o ódio*. Belo Horizonte: Âyiné, 2020.

Erikson, Edward. "Millary is my friend": MySpace and Political Fandom. *Rocky Mountain. Communication Review*, 4(2) (2008), p. 3–16.

Fabbro, Gabriela; Pittaro, Esteban. El algoritmo GPT-3 y su aplicación al periodismo: una experiencia de IA en la radio argentina. In: Vázquez-Barrio, Tamara; Salazar García, Idoia (eds.). *Inteligencia artificial, periodismo y democracia*. Valencia: Tirant lo Blanch, 2023, p. 113–127.

Fachin, Edson. Manifestação sobre o golpe em Mianmar. **Estado de Minas,** 2 February 2021. Available at: [https://www.em.com.br/app/noticia/internacional/2021/02/02/interna_internacional,1234501/sobre-golpe-em-mianmar-fachin-alerta-sobre-perversa-desmoralizacao-de-ele.shtml]. Viewed: 05.04.2024.

Filimowicz, Michael. Introduction. In: Filimowicz, Michael (ed.). *Deep fakes. Algorithms and society*. New York: Routledge, 2022, p. X–XI.

Fisher, Max. The Chaos Machine: The Inside Story of How Social Media Rewired Our Minds and Our World. New York: Little, Brown and Company, 2022.

Forti, Steven. *Extrema derecha 2.0. Qué es y cómo combatirla*. Madrid: Siglo XXI, 2021.

Frazão, Ana. A democracia na era digital: os riscos da política movida a dados. In: Branco, Paulo Gustavo Gonet; da Fonseca, Reynaldo Soares; de Branco, Pedro Henrique, Moura Gonet; Velloso, João Carlos Banhos; da Fonseca, Gabriel Campos Soares (eds.). *Eleições e democracia na era digital*. Brasília: Almedina, 2022, p. 69–84.

Frazão, Ana. TSE e as regras para o uso de inteligência artificial nas eleições. *Jota*, 13 de marzo de 2024. Available at: [TSE e as regras para o uso de inteligência artificial nas eleições (jota.info)]. Consultado: 31-03-2024.

Funk, Allie; Shahbaz, Adrian; Vesteinsson, Kian. Freedom on the Net 2023: The Repressive Power of Artificial Intelligence, 13th Edition. Freedom House, 2023. Available at: [https://freedomhouse.org/report/freedom-net/2023/repressive-power-artificial-intelligence]. Viewed: 30.10.2024.

Gandhi, Milan. *Terrorism, extremism, disinformation and artificial intelligence*. London: Institute for Strategic Dialogue, 2024.

García-Orosa, Berta. Desinformación, redes sociales, bots y astroturfing: la cuarta ola de la democracia digital. *Profesional de la Información*, 30(6) (2021), p. 1–10.

Garcés, Marina. *Novo esclarecimento radical*. Belo Horizonte: Âyiné, 2017.

Garriga, Miriam; Ruiz-Incertis, Raquel; Magallón-Rosa, Raúl. Inteligencia artificial, desinformación y propuestas de alfabetización en torno a los deepfakes. *OBS Journal* (2024, Special Issue,), p. 175–194. Available at: https://obs.obercom.pt/index.php/obs/article/view/2445/188188321. Viewed: 14.02.2024.

Gelev, Igor; Popovska, Bilijana. Fake news as part of the information operations. *Journal of European and Balkan Perspectives*, III(2) (2020), p. 55–71.

George, Cherian. The scourge of disinformation-assisted hate propaganda. In: Zimdars, Melissa; McLeod, Kembrew (eds.). *Fake news: understanding media and misinformation in the digital age*. Cambridge: The MIT Press, 2020, p. 145–152.

Gillespie, Tarleton. Platforms throw content moderation at every problem. In: Zimdars, Melissa; McLeod, Kembrew (eds.). *Fake news: understanding media and misinformation in the digital age*. Cambridge: The MIT Press, 2020, p. 329–239.

Giusti, Serena; Piras, Elisa. Introduction: in search of paradigms: disinformation, fake news and post-truth politics. In: Giusti, Serena; Piras, Elisa (eds.). *Democracy and fake news. Information manipulation and posttruth politics*. New York: Routledge, 2021, p. 1–16.

Góes, Bruno. Fake news 2.0: pré-campanha já tem suspeita de adulteração de áudios com uso de inteligência artificial em três estados. *O Globo*, 14 de enero de 2024. Available at: [Fake news 2.0: pré-campanha já tem suspeita de adulteração de áudios com uso de inteligência artificial em três estados (globo.com)]. Consultado: 01-04-2024.

Goltzman, Elder Maia; Soares, Rafael Rodrigues. Uso de IA nas campanhas: segurança jurídica nas eleições. *Conjur*, 2024. Available at: [Uso de IA nas campanhas: segurança jurídica nas eleições (conjur.com.br)]. Consultado: 01-04-2024.

Gómez De Ágreda, Ángel. La paz es la víctima última de la mentira. Desinformación con base tecnológica en la guerra. In: Vázquez-Barrio, Tamara; Salazar García, Idoia (eds.). *Inteligencia artificial, periodismo y democracia*. Valencia: Tirant lo Blanch, 2023, p. 205–226.

González-Torre, Ángel Pelayo. TIC, inteligencia artificial y crisis de la democracia. In: Cayón, Solar; Ignacio, José (eds.). *Dimensiones éticas y jurídicas de la inteligencia artificial en el marco del Estado de Derecho*. Madrid: Universidad de Alcalá, 2020, p. 55–78.

Guadián, Carlos. Cómo va a afectar la Inteligencia Artificial las elecciones en 2024. *CludPad*, 25 de enero de 2024. Available at: [https://carlosguadian.substack.com/p/como-va-a-afectar-la-inteligencia?utm_source=post-email-title&publication_id=259698&post_id=141029307&utm_campaign=email-post-title&isFreemail=true&r=4n36v&utm_medium=email]. Consultado: 25-01-2024.

Guarrigues Walker, Antonio; De La Garza, González; Miguel, Luis. *El derecho a no ser engañado. Y cómo nos engañan y nos autoengañamos*. Navarra: Arazandi, 2020.

Heller, Brittan. Enlisting useful idiots: the ties between online harassment and disinformation. *Colorado Technology Law Journal*, 19 (2021), p. 19–41.

Hernández Peña, Juan Carlos. El marco jurídico de la inteligencia artificial. In: *Principios, procedimientos y estructuras de gobernanza*. Madri: Aranzadi, 2022.

Hernández Ramos, Mario. Los retos constitucionales de la inteligencia artificial. Prospectivas, debilidades y fortalezas de los paradigmas vigentes. In: Soberanes Díez, José María; Garduño Domínguez, Gustavo (eds.). (coords.)*La interacción de las redes sociales, la tecnología y los derechos humanos*. Pamplona: EUNSA, 2023, p. 215–234.

Hsu, Tiffany; Thompson, Stuart A. ChatGPT será maior espalhador de desinformação que já existiu, diz pesquisador. **Folha de São Paulo,** 9 February 2023. Available at:[https://www1.folha.uol.com.br/tec/2023/02/chatgpt-sera-maior-espalhador-de-desinformacao-que-ja-existiu-diz-pesquisador.shtml]. Viewed: 25.03.2024.

Hui, Yuk. *Tecnodiversidade*. São Paulo: Ubu, 2020.

Hussain, Mumtaz; Soomro, Tariq Rahim. Social media: an exploratory study of information, disinformation, and malinformation. *Applied Computer Systems*, 28(1) (2023), p. 13–20.

Innerarity, Daniel. El año de la volatilidad. **El País,** 30 December 2018. Available at: [https://elpais.com/elpais/2018/12/28/opinion/1546021545_365361.html]. Viewed: 19.12.2023.

International Panel on the Information Environment [I. Trauthig, P. N. Howard, S. Valenzuela (eds.)]. *The Role of Generative AI Use in 2024 Elections Worldwide*. Zurich, Switzerland: IPIE, 2025. Technical Paper, TP2025.2,. DOI: 10.61452/HZUE9853.

Kahneman, Daniel; Sibony, Olivier; Sunstein, Carl. *Noise*. Nueva York: Little, Brown & Company, 2021.

Kakutani, Michiko. *A morte da verdade: Notas sobre a mentira na era Trump*. Rio de Janeiro: Intrínseca, 2018.

Kertysova, Katarina. Artificial Intelligence and Disinformation: How AI changes the way disinformation is produced, disseminated, and can be countered. *Security and Human Rights*, 29 (2018), p. 55–81.

Kiffer, Ana. O ódio e o desafio da relação: escritas dos corpos e afecções políticas. In: Kiffer, Ana; Giorgi, Gabriel (eds.). *Ódios políticos e política do ódio. Lutas, gestos e escritas do presente*. Rio de Janeiro: Bazar do Tempo, 2019, p. 35–78.

Kiffer, Ana; Giorgi, Gabriel. Apresentação: pensar o desafio presente. In: Kiffer, Ana; Giorgi, Gabriel (eds.). *Ódios políticos e política do ódio. Lutas, gestos e escritas do presente*. Rio de Janeiro: Bazar do Tempo, 2019, p. 9–20.

Kim, Eusong. On the depth of fakeness. In: Filimowicz, Michael (ed.). *Deep Fakes. Algorithms and Society*. New York: Routledge, 2022, p. 50–70.

Kissinger, Henry A; Schmidt, Eric; Huttenlocher, Daniel. The Age of AI: And Our Human Future. New York: Little, Brown and Company, 2021.

Klein, Ezra. Why We're Polarized. New York: Avid Reader Press / Simon Schuster, 2020.

Klein, Naomi. *On fire: The (burning) case for a green new deal*. New York: Simon & Schuster, 2020.

Kofi Annan Foundation. *Protecting electoral integrity in the Digital Age*. Geneve: *Koffi Annan Foundation*, 2020.

Kovic, Marko; Rauchfleisch, Adrian; Sele, Marc; Caspar, Christian. Digital astroturfing in politics: definition, typology, and countermeasures. *Studies in Communication Sciences*, 18(1) (2018), p. 69–85.

Labuz, Mateusz; Nehring, Christopher. On the way to deep fake democracy? Deep fakes in election campaigns in 2023. *European Political Science*, 23 (2024), p. 454–473. Available at: [https://doi.org/10.1057/s41304–024–00482–9]. Viewed: 29.04.2024

Lassalle, José María. Contra el populismo: Cartografía de un totalitarismo postmoderno. Debate, 2017.

Lassalle, José María. *Liberalismo herido. Reivindicación de la libertad frente a la nostalgia del autoritarismo*. Barcelona: Arpa & Arfil, 2021.

Lee, Kai-Fu. *Inteligência artificial. Como os robôs estão mudando o mundo, a forma como amamos, nos relacionamos, trabalhamos e vivemos*. 4.ª reimp. Rio de Janeiro: Globo Livros, 2021.

Leslie, David. Understanding artificial intelligence ethics and safety. The Alan Turing Institute, 2021. Available at: [https://zenodo.org/records/3240529]. Viewed: 22.04.2024.

Levitsky, Steven; Ziblatt, Daniel. *How democracies die*. New York: Broadway Books, 2023.

Lins, Rodrigo Martiniano Ayres. Abuso de poder algorítmico: considerações iniciais. In: Lins, Rodrigo Martiniano Ayres; Castro, Kamile Moreira (eds.). *O futuro das eleições e as eleições do futuro*. Belo Horizonte: Fórum, 2023, p. 289–306.

Loor-Cabal, Ginger Nicole; Gutiérrez-Izquierdo, María Alejandra; Alarcón-Chávez, Betty Elizabeth; Cedeño-Barreto, Mercedes de los Ángeles. La infoxicación digital y su incidencia en los niveles de estresse de los docentes de la Carrera de Trabajo Social de la Facultad de Ciencias Humanísticas y Sociales de la Universidad Técnica de Manabí. *Polo del Conocimiento*, 7(3) (2022), p. 356–370.

Loveless, Matthew. Information and democracy. Fake news as an emotional weapon. In: Giusti, Serena; Piras, Elisa (eds.). *Democracy and fake news. Information manipulation and post-truth politics*. New York: Routledge, 2021, p. 64–76.

Magallón Rosa, Raúl. *Updating news. Información y democracia*. Madrid: Ediciones Pirámide, 2023.

Maher, Sean. Deep fakes: seeing and not believing. In: Filimowicz, Michael (ed.). *Deep fakes. Algorithms and society*. New York: Routledge, 2022, p. 1–22.

Manfredi Sánchez, Juan-Luis; Ufarte Ruiz, María José; Gómez-Iniesta, Pablo. Desorden informativo y diplomacia en las relaciones internacionales. In:

Vázquez-Barrio, Tamara; Salazar García, Idoia (eds.). *Inteligencia artificial, periodismo y democracia*. Valencia: Tirant lo Blanch, 2023, p. 187–203.

Maradeo, Julián. *Fake news: cómo se fabrican en la Argentina y en el mundo*. Buenos Aires: Penguin Random House, 2021.

Maranhão, Juliano. A inteligência artificial não é a vilã das eleições. *Folha de São Paulo*, 6 de febrero de 2024. Available at: [https://www1.folha.uol.com.br/opiniao/2024/02/inteligencia-artificial-nao-e-a-vila-das-eleicoes.shtml?utm_source=whatsapp&utm_medium=social&utm_campaign=compwa]. Consultado: 07-02-204.

Margolis, M; Resnick, D. *Politics as usual: the cyberspace "revolution"*. London: Sage, 2000.

Marques, Maria Aldina. A verdade dos outros. Questões de responsabilidade enunciativa. In: Curcino, Luzmara; Sargentini, Vanice; Piovezani, Carlos (eds.). *Discurso e (pós)verdade*. São Paulo: Parábola, 2021, p. 135–153.

Martín Guardado, Sergio. Polarización, ruptura de la convivencia y crisis del sistema constitucional. In: Figueruelo Burrieza, Ángela (ed.). (dir.); Martín Guardado, Sergio (coords.)*Desinformación, odio y polarización*. Vol. I. Navarra: Arazandi, 2023, p. 211–231.

Mateos Crespo, José Luis. La desinformación como fenómeno creciente en las campañas de la Era Digital. In: Carratalá, Adolfo; Iranzo-Cabrera, María; López-García, Guillermo (eds.). *De la desinformación a la conspiración: política y comunicación ante un escenario híbrido*. Valencia: Tirant lo Blanch, 2023, p. 233–257.

Matoruga, Emir. El papel de la IA em la lucha contra las operaciones de influencia encubierta. **Ecommerce Hub.** Available at: [https://www.hulkapps.com/es/blogs/ecommerce-hub/el-papel-de-la-ia-en-la-lucha-contra-las-operaciones-de-influencia-encubierta]. Viewed: 10.8.2024.

Meaker, Morgan. Slovakia's Election deepfakes show AI is a danger to democracy. Fact-checkers scrambled to deal with fake audio recordings released days before a tight election, in warning for other countries with looming votes. *Wired*, 3 de octubre de 2023. Available at: [https://www.wired.co.uk/article/slovakia-election-deepfakes]. Consultado: 18-12-2023.

Mello, Patrícia Campos. *A máquina do ódio. Notas de uma repórter sobre fake news e violência digital*. São Paulo: Companhia das Letras, 2020.

Merenda, Federica. Reading Arendt to rethink truth, science and politics in the era of fake news. In: Giusti, Serena; Piras, Elisa (eds.). *Democracy and fake news. Information manipulation and post-truth politics*. New York: Routledge, 2021, p. 19–29.

Metaxas, Panagiotis Takis. Technology, propaganda and the limits of human intellect. In: Zimdars, Melissa; McLeod, Kembrew (eds.). *Fake news: understanding media and misinformation in the digital age*. Cambridge: The MIT Press, 2020, p. 245–256.

Minow, Martha. O ecossistema de notícias em mudança e os desafios para a Liberdade de imprensa. In: Branco, Paulo Gustavo Gonet; da Fonseca, Reynaldo Soares; Branco, Pedro Henrique de Moura Gonet; Velloso, João Carlos Banhos; da Fonseca, Gabriel Campos Soares (eds.). *Eleições e democracia na era digital*. Brasília: Almedina, 2022, p. 355–388.

Minsky, Marvin. *The society of mind*. New York: Simon & Schuster, 2021.

Mitchell, G Duncan. *Novo dicionário de Sociologia*. Porto: Rés Editora, 1970.

Montagut, Marta; Willem, Cilia; Carrillo, Nereida. Influencers y desorden informativo alrededor de la "ley trans". In: Carratalá, Adolfo; Iranzo-Cabrera, María; López-García, Guillermo (eds.). *De la desinformación a la conspiración: política y comunicación ante un escenario híbrido*. Valencia: Tirant lo Blanch, 2023, p. 255–275.

Moore, Martin. Democracy Hacked: Political Turmoil and Information Warfare in the Digital Age, London: Oneworld Publications, 2018.

Moreira Coellho, Gabriela Elizabeth; Castro Montenegro, Josua Isaí. *Cortinas de humo y su presencia en la Gestión gubernamental del Ecuador en 2021-2022*. Licenciatura en Comunicação. Guayaquil: Universidad Politécnica Salesiana, 2023. Available at: [UPS-GT004595. pdf]. Consultado: 20-03-2024

Morozov, Evgeny. *Capitalismo Big Tech: ¿Welfare o neofeudalismo digital?* Madrid: Enclave de Libros, 2018.

Mudde, Cas. *La ultraderecha hoy*. Barcelona: Paidós, 2019.

Mukherjee, Mitali. AI deepfakes, bad laws – and a big fat Indian election. **Reuters Institute,** 19 March 2024. Available at: [https://reutersinstitute.politics.ox.ac.uk/news/ai-deepfakes-bad-laws-and-big-fat-indianelection]. Viewed: 08.05.2025.

Mulhall, Stephen. *Heidegger and being and time*. New York: Routledge, 2021.

Mulhall, Stephen. *Heidegger's existentialism*. New York: Routledge, 2022.

Muñoz, Katja. The AI election year: how to counter the impact of artificial intelligence. DGAP Memo. Berlin: Forschungsinstitut der Deutschen Gesellschaft für Auswärtige Politik e.V., 2024, n. 1, pp. 1–5. Available at: [https://nbn-resolving.org/urn:nbn:de:0168-ssoar-92636-3]. Viewed: 08.05.2025.

Muñoz Vela, José Manuel. *Retos, riesgos, responsabilidad y regulación de la inteligencia artificial. Un enfoque de seguridad física, lógica, moral y jurídica*. Navarra: Thomson Reuters, Aranzadi, 2022.

Murgia, Michela. *Instrucciones para convertirse en fascista*. Barcelona: Seix Barral, 2019.

Murphy, Hannah. The rising threat to democracy of AI-powered disinformation. Experts fear 2024 could be the year a viral undetectable deepfake has a catastrophic impact on election. *Financial Times*, 11 de Enero de 2024. Available at: [https://www.ft.com/content/16f23c01-fa51-408e-acf5-0d30a5a1ebf2]. Viewed: 15-01-2024.

Niyazov, S. *(*2019*), The Real AI Threat to Democracy, Towards Data Science, 15 de noviembre.* Available at: [https://towardsdatascience.com/democracys-unsettling-future-in-the-age-of-ai-c47b1096746e]

Nunes, Felipe; Traumann, Thomas. *Biografia do abismo. Como a polarização divide famílias, desafia empresas e compromete o futuro do Brasil*. Rio de Janeiro: Harper Collins, 2023.

O'Connor, Cailin; Weatherall, James Owen. *The misinformation age. How false beliefs spread*. New Haven: Yale University Press, 2019.

Ochoa, Edurne. Ciber violencia política, el código que nos despolitiza. Ellas al Poder, 24 August 2023. Available at: [https://edurneochoa.com/ellas-al-poder/ciber-violencia-politica-el-codigo-que-nos-despolitiza]. Viewed: 25.03.2024.

Oliveira, Juliana Michelli; Siqueira, Rogério de Almeida. Três faces do ChatGPT: imaginários de uma máquina de linguagem. *Educação, Comunicação e Tecnologia*, 5(2) (2023), p. 104–123.

Osorio, Aline. *Direito Eleitoral e liberdade de expressão*. 2. ed. Belo Horizonte: Fórum, 2022.

Oyserman, Daphna; Dawson, Andrew. Your fake news, our facts. Identity-based motivation shapes what we believe, share, and accept. In: Greifeneder, Rainer; Jaffé, Mariela E; Newman, Eryn J; Schwarz, Norbert (eds.). *The Psychology of Fake News*. New York: Routledge, 2021, p. 173–195.

Pariser, E. The filter bubble. London: Viking, 2011.

Peirano, Marta. *El enemigo conoce el sistema*. Barcelona: Debate, 2019.

Pérez-Francesch, Joan Lluís. *Terrorismo*. In: Pereda, Carlos (ed.). Aragón Ribera, Álvaro; Delgado Parra, Concepción; Vega, Julieta Marcone; Leroux, Sergio Ortiz; Sermeño Quezada, Ángel (coords.). In: *Diccionario de injusticias*. Ciudad de México: *Siglo XXI, UNAM*, 2022, p. 718–724.

Pérez-Seijo, Sara; Vaz-Álvarez, Martín. Inteligencia artificial y periodismo: transformaciones, oportunidades y retos. In: Vázquez-Barrio, Tamara; Salazar García, Idoia (eds.). *Inteligencia artificial, periodismo y democracia*. Valencia: Tirant lo Blanch, 2023, p. 61–79.

Pomerantsev, Peter. *This is not propaganda: adventures in the war against reality*. New York: *Public Affairs*, 2020.

Prado, Magaly. *Fake news e inteligência artificial: o poder dos algoritmos na guerra da desinformação*. São Paulo: Edições 70, 2022.

Prado, Michele. *Tempestade ideológica. Bolsonarismo: a altright e o populismo iliberal no Brasil*. São Paulo: Todos Livros, 2023.

Przeworski, Adam. *Crises da democracia*. Rio de Janeiro: Zahar, 2020.

Rais, Diogo; Falcão, Daniel; Giachetta, André Zonaro; Meneguetti, Pamela. *Direito Eleitoral Digital*. São Paulo: Revista dos Tribunais, 2018.

Rebollo Delgado, Lucrecio. *Inteligencia artificial y derechos fundamentales*. Madrid: Dykinson, 2023.

Rodríguez Ferrándiz, Raúl. *Máscaras de la mentira. El nuevo desorden de la posverdad*. Valencia: Pre-textos, 2018.

RTW. Indonesia Instagram Pranowo Widodo. Available at: [https://restofworld.org/2024/elections-ai-tracker/#/indonesia-instagram-pranowo-widodo]. Consultado: 24-04-2024.

Rubio Núñez, Rafael. Los efectos de la posverdad en la democracia. *Revista de Derecho Político*, (103). Madrid: UNED, (2018), p. 191–228.

Rubio Núñez, Rafael. *Los efectos de la posverdad en la democracia. Revista de Derecho Político*. Vol. 103. Madrid: UNED, 2018.

Rubio, Ricardo; Monteiro, Vitor de Andrade. Preserving trust in democracy: the Brazilian superior electoral Court's quest to tackle disinformation in elections. *South African Journal of International Affairs*, 30 (2023). DOI: 10.1080/10220461.2023.2274860.

Sakamoto, Leonardo. *O que aprendi sendo xingado na Internet*. São Paulo: Leya, 2016.

Salinas Olarte, Miguel. Teorías de la conspiración: un análisis socio-político. In: Figueruelo Burrieza, Ángela (ed.). (dir.); Martín Guardado, Sergio (coords.)*Desinformación, odio y polarización*. Vol. I. Navarra: Arazandi, 2023, p. 333–349.

Samoilenko, Sergie A; Suvorova, Inna. Artificial intelligence and deepfakes in strategic deception campaigns: the U.S. and Russian experiences. In: Pashentsev, E (ed.). *The Palgrave handbook of malicious use of AI and psychological security*. New York: Springer, 2023, p. 507–529.

Sánchez, Manfredi; Juan-Luis; Gómez-Iniesta, Pablo. Desorden informativo y diplomacia en las relaciones internacionales. In: Vázquez-Barrio, Tamara; Salazar García, Idoia (eds.). *Inteligencia artificial, periodismo y democracia*. Valencia: Tirant lo Blanch, 2023, p. 187–203.

Sánchez Martínez, María Olga. Desafíos democráticos en el ecosistema digital. In: *SOLAR CAYÓN, José Ignacio. Dimensiones éticas y jurídicas de la inteligencia artificial en el marco del Estado de Derecho*. Madrid: Universidad de Alcalá, 2020, p. 79–124.

Sánchez Muñoz, Óscar. *La regulación de las campañas electorales en la era digital. Desinformación y microsegmentación en las redes sociales con fines electorales*. Madrid, Valladolid: Centro de Estudios Políticos y Constitucionales, 2020.

Sandel, Michael J. *O descontentamento da democracia. Uma nova abordagem para tempos periculosos*. Rio de Janeiro: Civilização Brasileira, 2023.

Sarlet, Ingo Wolfgang; Siqueira, Andressa de Bittencourt. As novas dimensões da liberdade de expressão numa democracia: uma análise à luz dos desafios relativos às fake news nas redes sociais. In: Branco, Paulo Gustavo Gonet; Fonseca, Reynaldo Soares da; Branco, Pedro Henrique de Moura Gonet; Velloso, João Carlos Banhos; Fonseca, Gabriel Campos Soares da (eds.). *Eleições e democracia na era digital*. Brasília: Almedina, 2022, p. 167–197.

Schick, Nina. *Deep fakes and the infocalypse: what you urgently need to know*. New York: Monoray, 2020.

Schick, Nina. *Deep fakes: the new disinformation phenomenon*. New York: Monoray, 2021.

Schulte, Stephanie Ricker. Fixing fake news: self-regulation and technological solutionism. In: Zimdars, Melissa; McLeod, Kembrew (eds.). *Fake news: understanding media and misinformation in the digital age*. Cambridge: The MIT Press, 2020, p. 133–144.

Shoai, Andrés; López Molina, Adrián. Polarización e inteligencia artificial: una sistematización del conocimiento disponible. In: Vázquez-Barrio, Tamara; Salazar García, Idoia (eds.). *Inteligencia artificial, periodismo y democracia*. Valencia: Tirant lo Blanch, 2023, p. 247–264.

Siddiqua, Ayesha. Use of cyber hate in the electoral campaigns by the mainstream political parties of Pakistan. *Humanities and Social Sciences Reviews*, 9(2) (2021), p. 325–332.

Silva, Max Melquiades da; Cendón, Beatriz Valadares. Estratégia, método e conteúdo: três componentes para a compreensão das campanhas contemporâneas de desinformação. *Bibliocanto*, 8(1) (2022), p. 21–44.

Siverio Luis, Sergio. Bulos y desinformación sobre los efectos de la autodeterminación del género como mecanismo para reconocer las personas trans. In: Figueruelo Burrieza, Ángela (ed.). (dir.); Martín Guardado, Sergio (coords.)*Desinformación, odio y polarización*. Vol. I. Navarra: Arazandi, 2023, p. 361–377.

Snyder, Timothy. *Sobre a tirania*. Vinte lições do século XX para o presente. São Paulo: Companhia das Letras, 2017.

Soch. PTI's Sher Azfal Marwat did not call for 2024 election boycott; video is a deepfale. Available at: [https://www.sochfactcheck.com/general-elections-2024-sher-azfal-marwat-pti-did-not-call-for-2024-election-boycott-video-is-a-deepfake/]. Consultado: 24-04-2024.

Souza, Andrade de, Bruno Cezar. *Dados pessoais: LGPD e as eleições*. Belo Horizonte: D'Plácido, 2022.

Strobl, Natascha. *La nueva derecha. Un análisis del conservadorismo radicalizado*. Buenos Aires, Madrid: Katz, 2022.

Sunstein, Cass R. *La conformidad. El poder de las influencias sociales sobre nuestras decisiones*. Ciudad de México: Instituto Nacional Electoral, Grano de Sal, 2020.

Suying, Dymples Leong. Deep fakes and disinformation in Asia. In: Filimowicz, Michael (ed.). *Deep fakes. Algorithms and society*. New York: Routledge, 2022, p. 23–49.

Szabó, Ilona. *A defesa do espaço cívico*. São Paulo: Objetiva, 2020.

Tasioulas, John. First step towards an ethics of robots and artificial intelligence. *Journal of Practical Ethics*, 7(1) (2019), p. 61–95.

Thaler, R H; Sunstein, C R. Nudge: Improving Decisions about Health, Wealth, and Happiness. Revised edition. Londres, Penguin. [*Disponible en castellano como Un pequeño empujón (nudge): el impulso que necesitas para tomar las mejores decisiones en salud, dinero y felicidad (2017)*]: Barcelona, Taurus, 2009.

Torrecillas Lacave, Teresa; Fernández Martínez, Luis Manuel. Inteligencia artificial y periodismo: oportunidades para la lucha contra la desinformación en la red. In: Vázquez-Barrio, Tamara; Salazar García, Idoia (eds.). *Inteligencia artificial, periodismo y democracia*. Valencia: Tirant lo Blanch, 2023, p. 81–96.

Torres Soriano, Manuel R. Hackeando la democracia: operaciones de influencia en el ciberespacio. *Instituto Español de Estudios Estratégicos*, 66/2017. Available at: [https://www.newtral.es/wp-content/uploads/2023/10/Hackeando_democracia_MRTorres-copy-2.pdf?x50694]. Consultado: 02/04/2024.

Utami, Pratiwi. Hoax in modern politics: the meaning of hoax in indonesian politics and democracy. *Jurnal Ilmu Sosial dan Ilmu Politik*, 22(2) (2018), p. 85–97.

Vallespín, Fernando. *La sociedad de la intolerancia*. Barcelona: Galaxia Gutemberg, 2021.

Van Der Linden, Sander. Las fake news generadas por IA te acecharán en las próximas elecciones. **Wired,** 23 January 2024. Available at: [https://es.wired.com/articulos/fake-news-generadas-por-ia-te-acecharan-en-proximas-elecciones]. Viewed: 24.01.2024.

Venice Comission. Joint Report of the Venice Comission on Digital Technologies and Elections. Estrasburgo: Comissão de Veneza, 24 de junho de 2019. Disponível em: [https://www.venice.coe.int/webforms/documents/?pdf=CDL-AD(2019)016-e]. Viewed: 28.12.2023.

Viennot, Bérengère. *A língua de Trump*. Belo Horizonte: *Âyiné*, 2021.

Vivas Escribano, Guillermo. Desinformación y polarización en las redes sociales. Figueruelo Burrieza, Ángela (dir.). Martín Guardado, Sergio (coord.). In: *Desinformación, odio y polarización*. I. Navarra: Arazandi, 2023, p. 351–359.

Vlachos, Scott. The link between mis-, dis-, and malinformation and domestic extremism. Council for Emerging National Security Affairs, June 2022. Available at: [https://censa.net/wp-content/uploads/2022/06/MDM_22.6.17b.pdf]. Viewed: 25.03.2024.

Vosoughi, Soroush; Roy, Deb; Sinan, Aral. The spread of true and false newsonline. *Science*, 359(6.380) (2018), p. 1.146–1.151.

Wagner, Angelia. Tolerating the trolls? Gendered perceptions of online harassment of politicians in Canada. Feminist Media Studies, 22 (2020), pp. 32–47. Available at: [https://www.tandfonline.com/doi/full/10.1080/14680777.2020.1749691]. Viewed: 05.04.2024.

Wardle, Claire; Derakshan, Hossein. *Information disorder: toward an interdisciplinary framework for research and policy making*. Strasbourg: Council of Europe, 2017.

Williams, Matthew. *A ciência do ódio*. Rio de Janeiro: Globo Livros, 2021.

Williams, Rhiannon. Humans may be more likely to believe disinformation generated by AI. **MIT Technology Review,** 28 June 2023. Available at: [https://www.technologyreview.com/2023/06/28/1075683/humans-may-be-more-likely-to-believe-disinformation-generated-by-ai]. Viewed: 20.04.2024.

World Economic Forum. The Global Risks Report 2024: Insight Report, 19th Edition. January 2024.

Wrong, Dennis H. Radicalismo. In: Outhwaite, William; Bottomore, Tom (eds.). *Dicionário do pensamento social do século XX*. Rio de Janeiro: Zahar, 2000, p. 645–647.

Zayan, R. Radicalismo. In: Thines, Georges; Lempereur, Agnés (eds.). *Diccionario General de Ciencias Humanas*. Lisboa: Edições 70, 1984, p. 777.

Zhang, Jerry; Carpenter, Darrell; Ko, Myung. Online astroturfing: A theoretical perspective. Proceedings of the 19th Americas Conference on Information Systems, Chicago, 15–17 August 2013. (2013), pp. 1–7.

CHAPTER 3

The Regulatory Response

AI is advancing and renewing itself at a frenetic pace. Any attempt to close the issue, given this condition, would be erroneous. Despite an abundance of publications, the study of the subject is in its infancy, and we are facing the first election cycles in that will take place under the influence of AI. Technological progress is advancing at an exponential rate, bringing with it new – and unpredictable – challenges for electoral bodies, which is why definitive maxims in this context tend to be out of step and inaccurate (ARCHEGAS; MAIA, 2022). In fact, categorical conclusions might suggest the opposite of what is intended here: a contribution to a discussion which, by its nature, should be open and unfinished. For this reason, this final chapter does not seek so much a conclusion as a focus toward the future with the intention of identifying, from the viewpoint of the present, the possibilities and opportunities that the future holds, particularly in the regulatory field.

Ángel Gómez (GÓMEZ DE ÁGREDA, 2023, p. 222) explains that humanity has defined its stages according to the most significant technology: "mastering stone, bronze, iron, steam, internal combustion or computers changed how we live." However, in all cases history has continued to be "the history of humanity." However, AI and the integration of digital layers into our lives

AI and Electoral Campaigns, First Edition. Rafael Rubio Núñez, Frederico Franco Alvim and Vitor de Andrade Monteiro.
© 2026 John Wiley & Sons, Inc. Published 2026 by John Wiley & Sons, Inc.

are not only changing the way we do things but also, "for better or for worse," the number of things that affect us. As we have already seen, part of the replacement of humans by machines affects the functioning of elections. In its automated manifestation, the contest for political power decides the fate of humans, while also to some extent challenging their central role.

Communication is one of the processes most affected by this transformation: the generation of synthetic content blurs the distinction between reality and fiction; the distribution of information and opinions depends on inscrutable algorithms; intelligent systems seek out interests and habits to direct messages that are perfectly adapted to the customer's tastes (LEWANDOWSKY; SMILLIE, 2020). Everything converges to dissolve truth and the rational dimension of politics into an aggressive, obstinate, and disturbing mode of persuasion. Dehumanization occurs in a double sense: there is less humanity both in the production process and in the ideological fiber of the content that circulates.

Years ago, technical limitations and built-in costs made it difficult to "industrially manage" these processes, reducing their impact and allowing for damage control: the mechanization of politics, while making it difficult to maintain (relatively) transparent and fair communication, exhibited a manageable, albeit challenging, scenario. However, the automation and above all, the *algorithmization of the dialogical sphere* through the integration of AI solutions is reshaping the public sphere, resetting the very dynamics of interaction between political subjects. AI, with its possibilities for learning and language processing, enables the industrialization of persuasion, making accessible – from a technical and economic perspective – the mass tapping of social networks to discover and falsify codes and ways of *hacking* people's minds.

In this context, it is worth asking whether the social configuration dictated by the *platformization of life* is compatible with the rule of law, democracy, and in particular, fundamental rights and freedoms, given that this new technology tends to "de-structure the social and collective mechanisms that we have organized up until now." While it is true that the course of this century will be determined by technology, "it is also true that these transformations must take into account the human condition" and the pillars of its organization, including the legal system itself. For this reason, we urgently need to assess "whether the law will be affected or replaced by another type of code" in the future (REBOLLO DELGADO, 2023, p. 29).

Conscious of these new situations, some institutions are beginning to address the repercussions of AI for democracy. However, as Bender points out, although academics have warned of "algorithmic damage" in a range of areas – from the criminal justice system to the labor market, from industrial

competition to civil rights – much of this harm has taken root in the electoral process yet has not been debated with the necessary depth, calculation, and attention (Bender, 2022, p. 489).

3.1 PRELIMINARY RESPONSES

In his *Meditations on Technique*, Ortega y Gasset anticipated some important current reflections on the subject of AI: for the philosopher, the human being cannot be understood without technology and the possibilities that new techniques offer for the development of civilizations. It is clear that new tools often bring with them risks and challenges for societies. History is full of examples of new technology which, while ushering in progress and development, also has the potential to cause damage and destruction (Rivero Ortega, 2023, p. 9); examples such as fire, TNT (trinitrotoluene), fossil fuels, and nuclear energy illustrate the point very well.

Nevertheless, it is evident that the memory of catastrophic experiences linked to technological innovation does not prevent society from wanting to explore new possibilities and take advantage, as far as possible, of opportunities to expand the "toolbox" capable of strengthening survival on Earth (Rivero Ortega, 2023, p. 9). The challenge lies in the difficult task of finding peaceful coexistence between the individual freedom to benefit from new technology, on the one hand, and the state's duty to protect society from the risks that these innovations entail, on the other.

When it comes to digital advances, state intervention at the regulatory level often happens late (Kim, 2022, p. 52) and is frequently timid and imprecise in the face of all the complexity and urgency presented by the digital environment's accelerated dynamics. There are, of course, some reasons for this behavior.

The first is that the more complex the innovation, the more difficult it is to predict the inherent risks. Despite the vast literature and the various studies underway, the vertiginous vocation (with the constant superimposition and renewal of forms and means of intervention) and the ultra-specialized nature prevent an exact understanding of the risks that AI might generate for humanity in the medium or long term, including with regard to the appropriate development of electoral processes. Indeed, it is a fact that today:

> [W]e inhabit a context of permanent change. Not every year, but every month, a new possibility arises, an improvement in applicability is identified in a circumstance that had not been foreseen. Technical possibilities grow exponentially in terms of their creation

> and applicability over time. On the other hand, the law, as we know, is slow to respond, and requires the coordinated action of many actors (legislators, jurisprudence, public administration, control bodies, etc.), which is compounded by the need to act at various levels, such as the regional or international. However far-sighted legal systems may be, AI possesses a strong element of unpredictability, of results or collateral consequences that are not anticipated,[1] and which, in their very creation and application, require a process of permanent control and readaptation. This implies an additional difficulty for to law as the response, and constitutes the most substantial threat posed by AI.
>
> (Rebollo Delgado, 2023, p. 54)

Another factor that works against a global movement to regulate AI gaining traction is the fact that regulation implies, in essence, limiting in some sense the freedom of economic exploitation within the scope of such innovations. Ethical and principled proposals aimed at preventing abuses or curbing excesses, however relevant they may be, may not be well received when they are presented as a form of opposing global technological progress.[2]

On the other hand, it is important to realize that the regulatory task is not simple, since it requires appropriate technical knowledge to avoid two counterposed risks: regulatory passivity and regulatory excess. The fact is that the absence of a disciplinary framework to curb abuses in the digital surge could be as harmful to social development as excessive intervention

[1] "The increase in its capacity for self-learning and its degree of autonomy brings with it a relative unpredictability in its relationship and interaction with its environment, context and people, which should be scarce or nil if we really subject AI to the intended security, supervision and human control throughout its life cycle and, in any case, susceptible to the annulment and reversibility of its decisions, actions or omissions when this is possible in view of the context of these, in accordance with the majority of ethical frameworks that are currently the subject of an intended consensus at European and international level" (Muñoz Vela, 2022, p. 26).

[2] From this perspective, it is argued that restriction is not a viable path and that it would be more appropriate to invest in "transparency mechanisms related to the use of AI by candidates in their communications, such as the introduction of digital markers." Having said that, we need to avoid alarmism and more specifically, "prejudiced approaches that identify technology as [synonymous with] fraud" (Maranhão, 2024).

based on a paralyzing logic rooted in a misconception of necessary precaution (Rivero Ortega, 2023, p. 11). To sum up:

> Let us recall that the law has considerable intrinsic disadvantages when it comes to regulating technology. It is eminently statist, and we have seen that technological innovations have broken through the barriers of space and time. Another characteristic of the law is that it arises to resolve an existing conflict, its predictive capacity is limited and ridden with errors and loopholes, and in any case, it is usually generated *a posteriori*, once the need has been identified. Finally, its elaboration is mediated by a multitude of actors and interests, which compounds its difficult application.
>
> (Rebollo delgado, 2023, p. 16)

Despite the issues raised, today's democracy requires a conscious citizenry capable of deciding freely and as such, AI can be a useful tool for parties and candidates to effectively fulfill their information functions (Vázquez-Barrio, 2023, p. 21). For this same reason, the regulatory proposals that have been created work on the premise that AI cannot – or should not – be excluded from the electoral sphere, but that it needs a series of legal outlines to ensure that its use is compatible with the values and principles that constitute the notion of electoral integrity, such as freedom, equality, civic responsibility, transparency, tolerance, subjection to control, and respect for rules and *fair play*.

As Hernández Ramos (2023, p. 216) points out, "there are currently no all-encompassing legal regulations for this type of technology, with ethical commitments and private self-regulation codes taking precedence." In more general terms, academics share the view that, to deal with current challenges, the legal experience needs to be reinforced with adjustments made at a supra-legal level, through technological measures which, among other things, guarantee greater transparency in relation to the use of data by digital platforms[3]

[3] Dolores Montero believes that the uncontrolled use of AI poses very serious problems in the area of behavior control. For this reason, she argues that algorithmic transparency should be considered a necessity based on the democratic principle, above all as a "response to the possible abuse of this technology." In her opinion, accountability for decisions made by technological entities "is crucial so that citizens can really see that decisions that directly affect their lives can be reviewed and, if necessary, withdrawn or modified." The author understands that "for this reason, it is essential that transparency, one of the pillars of open government, also manifests itself in environments controlled by artificial intelligence" (Montero Caro, 2023, p. 192).

as well as stricter protocols for the registration of accounts and profiles on social networks – with the elimination of anonymity and the prospect of accountability in view. In terms of the law, transformation efforts should pursue stricter provisions against disinformation, along with mechanisms capable of increasing the responsibility of relevant actors, all in light of the negative potential derived from AI's increasingly ubiquitous presence (TASIOULAS, 2019, pp. 87–88).

In line with this spirit, the regulatory framework recently promoted by the Brazilian Superior Electoral Court (TSE) seems to indicate that, from a democratic perspective, not deceiving should already be seen as a fundamental duty (GARRIGUES WALKER; GONZÁLEZ DE LA GARZA, 2020, p. 153ff; TASIOULAS, 2019, p. 87) and that the protection of truth should be seen as a governing value for legislative activity (ALVIM *et al.*, 2023, p. 80). As such, it becomes "fundamental to adopt a regulatory structure that establishes the distribution of responsibilities between all the actors in the digital ecosystem," putting in place "parameters and guarantees to minimize the risks of artificial intelligence," and starting from the inescapable premise that, in the electoral sphere, it is necessary to undertake "rapid actions to prevent attacks that endanger the electoral process" (BIONI *et al.*, 2024).

That said, for didactic purposes, AI can be understood from two different conceptions: *AI as a (simple) tool* and *AI as the subject of* communication itself. From this perspective, Shoai and López explain that synthetic intelligence can act both as a vehicle in the process of mediated communication (to generate, modify, or increase the reach) of the messages published by its users and alternatively, as a means of introducing and operating communications agents that interact directly with the public, through bots, chatbots,[4] robocalls,[5] etc. (SHOAI; LÓPEZ MOLINA, 2023, p. 249).

---

[4]According to Hampton (2019, p. 12), "the technology used to create chatbots has the potential to exploit weaknesses in the communication architecture and obstruct political processes." "AI can play a role in this by filtering and regulating information and in the form of bots, which influence political communication and potentially also voter preferences" (COECKELBERGH, 2022, p. 121). An example is Hello-vote, a chatbot created in the United States to help people register to vote as well as to provide information about the date and location of polling stations. These chatbots can be used to obtain information about voters or verify the information they provide (HAMPTON, ibid.).

[5]Before it was banned by the FCC, this practice had been used in the United States ever since the Democratic candidate for congress in Pennsylvania, Shamaine Daniel, inaugurated

Moreover, until now, regulatory projects have generally covered aspects related to both conceptions, which is illustrated by the scarcity of legislation in various places regulating issues related to the dissemination of false or harmful content generated or distributed with the participation of synthetic intelligence solutions. In this direction, the ban on the use of AI to produce materially dishonest messages, confirmed in Brazil and currently under discussion in the US Congress (Klobuchar, 2023), is joined by the prohibition of automated calls (also in the United States) and various other related measures.

The fact is that, to date, National Conference of State Legislatures regulatory projects have focused to a large extent on the visible layer of AI, as a general rule neglecting equally important issues related to the discreet, silent, and ruthless functioning of "algorithmic message selection" (Gerlitz; Helmond, 2013). In this setting, political pluralism ends up being constrained by bubble algorithms, and equal opportunities are threatened by a lack of transparency which, according to Óscar Sánchez, extends to numerous aspects, from the identity of those who pay to the amount of money spent, from the way in which information is filtered to the feverish way in which disinformation narratives circulate (Sánchez Muñoz, 2020, pp. 87–88), in many cases with the assistance of automated accounts of anonymous origin. Furthermore, the systematic sharing of personal data on preferences, habits, and opinions underlines "the risk that this information will be manipulated, which [often] becomes power in the hands of those who possess it," in the wake of a "new cycle [...] that challenges the foundations of democracy" (Lins, 2023, p. 289). In effect:

> If we consider the contexts in which AI is used today, [...] the issues of privacy and data protection are becoming increasingly problematic. It is relatively easy to respect these values and rights when researching as a social scientist: you can inform the interviewees and explicitly ask for their consent, and it is relatively clear what will happen with the data. But the environment in which AI and data science are used today is usually very different. Think of social networks: although privacy information is available and applications ask users for their consent, it is not clear what will happen to the

---

the Ashley bot (developed by Civox), which was capable of making automatic calls thanks to GenAI. The technology makes it possible to make thousands of personal calls a day, adapting each dialog to the characteristics and interests of the recipient. In addition, the functionality can work effectively in more than 20 languages (Tong; Coster, 2023).

> data provided or even which parts of it will be stored. Moreover, in order to use the application and enjoy its benefits, there is little alternative than to give consent. Users are frequently unaware that AI is driving the application they use. Often the data provided in one context is transferred to another domain and used for a different purpose [...].
>
> This latter phenomenon also points to the risk of users being manipulated and exploited. AI is used to manipulate what we buy, the news we follow, the opinions we trust, etc. Critical theory researchers have highlighted the capitalist context in which the use of social media takes place. It can be said that social media users are free digital "manpower" that produces data for companies. [...] The danger here is that, even in today's democracies, AI can lead to new forms of manipulation, surveillance and totalitarianism, not necessarily as authoritarian policies, but in a more hidden and highly effective way: by altering the economy in such a way that it turns us all into a herd of smart phones ordered by our data.[6]
>
> (Coeckelbergh, 2022, pp. 94–95)

In this framework, despite important and significant advances, there is still a significant regulatory vacuum that needs to be filled without further delay if elections are to take place under fairer and more balanced conditions. As Óscar Sánchez points out, in a reflection that is applicable to AI, computerized communication raises concerns among citizens and entails a "tangible risk" of undermining trust in the electoral process, insofar as some practices make it possible to attack "basic principles on which its integrity rests: equality of arms between electoral competitors and voters' freedom to make their own decisions through a process of public communication that is free and open, and without improper influences." In this context, the lack of a modern framework, adapted and capable of guaranteeing the full validity

[6] "But AI can also be used to manipulate politics in a more direct way, for example by analyzing data from social networks to help political campaigns (as in the case of Cambridge Analytica, which used Facebook users' data without their consent for political purposes during the 2016 US elections), or by having bots publish political messages on social networks based on an analysis of people's data regarding their political preferences in order to influence the vote. Some also worry that AI, by taking over the cognitive tasks of humans, will infantilize its users, 'making them less able to think for themselves or decide for themselves what to do'" (Coeckelbergh, 2022, pp. 95–96).

of these principles in the face of new challenges, "is a fact about which there is general consensus" (Sánchez Muñoz, 2020, p. 21).[7]

In this panorama, democratic organization requires a solid and complete legal framework that takes into account the electoral moment and its different phases and adopts clear and assertive positions on: (a) enhanced privacy protection; (b) the mitigation of anonymity;[8] (c) the necessary level of transparency (with measures such as preventing bots and fake accounts from masquerading as people and identifying content generated by machines); (d) accountability for content produced with AI; (e) advertising models in which AI plays a key role, to prevent the technology from affecting freedom of decision-making, increasing transparency (with open-access libraries of politicals) and taking a clear stance on the use of microtargeting, setting limits on its use or even proposing a ban on the use of these techniques in electoral processes (based on the premise that constitutions do not protect machine-generated speech); (f) the use of this technology to combat disinformation (to ensure that it is truly independent) as well as respect for procedural guarantees when moderation and content elimination measures are left in the hands of AI; and (g) the accountability of the platforms, facilitating access for researchers, fact-checking initiatives, and civil society organizations to evaluate the impact of AI on online political campaigns.

Regardless of the particularities of specific agreements, in a macro sense regulation must prioritize electoral integrity, and not AI per se. Based on this vision, the main task will be to guarantee, on the one hand, the elimination

[7] "This is made very clear, for example, by the Council of Europe's study on the use of the Internet in election campaigns, which notes the 'inability of regulation to guarantee a level playing field for political competition and to limit the role of money in elections'. For this reason, from different institutional and academic spheres, legal reforms of a varying nature and scope are being proposed in order to curb these practices and preserve the open and egalitarian nature of the communication process prior to the electoral decision" (Sánchez Muñoz, 2020, pp. 21–22).

[8] It should be born in mind that, on many occasions, the anonymous nature of social media users or message senders on chat platforms is a major factor in fomenting conflict and tension in demonstrations in the public sphere since irresponsibility eliminates ties and barriers. That said: "The danger that is readily apparent to us in this context is that it can become a suitable field for fake news to be accepted and expanded, for 'hate speech' to proliferate, or for the 'post-truth' to triumph, at least within certain user segments. This occurs especially when the recipients always receive news with the same bias (in the setting of a so-called 'filter bubble'), which helps to make the user think that this is the most truthful opinion, or even the socially majoritarian opinion" (González-Torre, 2020, pp. 70–71).

of manipulation, abuse, and deception and on the other, respect for freedom of expression (ensuring that any restriction respects the principles of legality, legitimate purpose, and necessity), personal privacy, the freedom to vote, and the right to participate in public affairs in an informed manner and under equal conditions. As a result, the emergence of an "artificial democracy, forged, created [...] as a result of automated accounts and debates maximized by the action of robots" (Leal; Moraes Filho, 2019, p. 354), while at the same time guaranteeing a certain degree of control over the functioning of algorithms, assuming the fact that beyond all doubt, their opacity undermines "the transparency and *accountability* of the electoral process," "limiting the capacity of individuals" (Lins, 2023, p. 303) and tipping the balance in favor of certain forces or ideologies, catapulted voluntarily or accidentally by the architecture of social media.

3.1.1 The First Regulations

To date, the regulatory mosaic has been fundamentally limited to instruments of self-regulation and *soft law* – which have proved insufficient, especially with regard to democracy, the rule of law, and human rights (Council of Europe, 2020, §§ 18–20). However, as we have seen, various international institutions and organizations have been working to establish frameworks for democratic protection with a general scope, adapted to this new context.

This is a good approach, since the regulatory response to the use of AI in electoral processes cannot proceed exclusively from electoral regulation, nor can it be relegated – without reservation – to the technology companies' self-regulation.[9] As a result, the impending treatment of the issue in the specific field of electoral law does not abrogate with the need to design general solutions to the challenges of AI, especially on a regulatory level that supersedes national borders.

Ultimately, a general regulatory framework is considered necessary, with principles, content rules, and an institutional structure capable of addressing the impact of AI not only on fundamental rights but also on political and social organization itself (Montilla Martos, 2023, p. 167). It should also be borne in mind that addressing the dilemmas of the digital context involves

[9] At the end of the day, following the natural order of modern societies, "let it be new technology that adapts to the traditional dogma of the law and not legal dogma that is adapted to the demands of new technology" (Coello de Portugal, 2014, p. 70).

collisions with fields adjacent to the world of law, such as software engineering and the internal governance of digital platforms.[10]

Ideally, projects of this scope should address, among other issues: data governance; the right to dignity, privacy, and control over one's own image (deepfakes); the issue of anonymity; hate speech, harassment, and political violence; the protection of vulnerable groups; the mitigation of systemic risks and cross-border accountability as well as a specific section on electoral communication with effective rules to guarantee transparency, accountability, and enhanced protections in the area of the extraction of personal information and the individualization of influential content (microtargeting), along with other measures to preserve equality between competitors. The primary objective is, as we know, technological neutrality and to guarantee this, we must invest in openness, controllability, and explainability as well as ensuring human participation in the design, setting of objectives, application, and supervision of the entire chain of activities involving AI.

3.1.1.1 The European Union's Response

In the European Union (EU), there were until recently no specific regulations relating to election campaigns. However, the General Data Protection Regulation (GDPR) applies to election campaigns, affecting the "processing of personal data in the context of an election" by establishing a series of provisions and limitations such as explicit consent, the processing of data to fulfill a legal obligation or in the public interest, including the obligations of those responsible for data processing. According to the GDPR, political actors "are obliged to inform people about the processing of their data and especially about who is responsible for this, the purposes of the processing, the sources of the data, the recipients, etc. It should be noted that in the context

[10] "It seems clear that the new digital society cannot be shaped solely by the law, bearing in mind that it involves a new horizontal configuration of the social body, and that it affects the media, the economy, all areas of society, even the very foundations of social and political organization. For this reason, De la Quadra Salcedo's proposal seems very apt, when he uses the concept of the holistic solution and specifically states that 'This affectation of all the fundamental elements that structure and inform our societies makes it obligatory to adopt a holistic perspective in dealing with the challenges posed by the digital society'. Viewing law as the only way to order and regulate the digital society is a serious mistake. Law must be, as it has always been, a way of resolving social conflicts with a common good perspective, but in all cases it needs the collaboration of other areas of knowledge, of all the elements that make up the social structure" (Rebollo Delgado, 2023, p. 52).

of an election, much of the data collected and processed belongs to the sensitive category, so both those in charge and those responsible for processing have to apply appropriate measures to guarantee the level of security required in proportion to the risks entailed" (García Mahamut, 2023).

Directive 2022/2555 of the European Parliament and of the Council of 14 December 2022 on measures to ensure a high common level of cybersecurity throughout the EU also applies, which underscores the fact that entities belonging to the digital infrastructure sector are essentially based on network and information systems and underlines the need for the physical security of these systems as part of its measures for the management of cybersecurity risks since, as we have seen, their use during elections can compromise the integrity of the process and engender distrust in its functioning.

In addition, the recent approval of a regulation on AI, which establishes harmonized rules on AI and modifies prior community rules, also affects election campaigns and has been accompanied by the adoption of other specific measures for this defining moment in democracy. This regulation, as Simón Castellano (2023, p. 26) points out, "establishes a regulatory model based on the proactive responsibility of the different actors involved in the development, implementation and commercialization of AI-based tools, and does so with an approach that starts from the delimitation of levels of risk depending on the technology employed and its possible uses."

To deal with these specific risks, the European Artificial Intelligence Regulation foresees four modalities,[11] each with a guideline that is described below:

[11] "To go deeper into risk-based regulation, we need to go back a few steps to explore the notion [...]. Gellert (2017) conceptualizes risk as a tool that helps the decision-making process, directing its analysis not to its existence – (presumed) – but to how much risk a given agent can or is willing to take and how much it is capable of mitigating. Similarly, Hood *et al*. (2001) define risk as a probability of adverse consequences, with risk regulation being a governmental interference in the market or in social processes to control possible adverse consequences. [...] Hood *et al*. point out that human activity and technology in modern times have the collateral effect of risks that depend on experts to be evaluated and recognized, are collective, global and irreversible in their impact, giving rise to a 'risk society' that is distinct from previous historical periods. Therefore, risk can be understood as 'the ability to define what might happen in the future and to choose between alternatives', functioning as a tool for decision-making insofar as it makes what is uncertain certain. Its constituent elements are two distinct but linked operations: predicting the future (with the help of numbers) and making decisions based on it. Thus, risk, although associated with something more quantifiable, can also be understood as a qualitative and evaluative element that needs to be assessed from different perspectives" (Bioni *et al*., 2023, p. 25).

> *Minimal or non-existent risks*: most AI systems with insignificant risks can continue to operate without regulation;
>
> *Limited risks*: AI systems with acceptable risks are subject to minor transparency obligations so that users can make informed decisions;
>
> *High risks*: a wide range of high-risk AI systems will be authorized, but with strict requirements and obligations to access the EU market;[12]
>
> *Unacceptable risks*: with limited exceptions, systems that contain risks considered unacceptable will be prohibited, such as cognitive manipulation, predictive police surveillance, emotion recognition in workplaces and schools, social punctuation and certain at distance biometric identification systems.
>
> (European Parliament, 2023; Revoredo, 2023; Ryan-Mosley, 2024)[13]

In the electoral sphere, firstly the new regulation prohibits *unacceptable risks,* such as those that threaten the security or rights and freedoms of citizens, including the prerogatives related to free, conscious, and self-determined political participation. This category includes applications capable of manipulating human behavior[14] and identifying or providing information on the vulnerabilities of certain groups as well as special circumstances involving biometric categorization or mass video surveillance by the authorities in public spaces.

High-risk systems also affect the electoral process; these threaten critical infrastructure and can interfere with people's rights, such as those related to

[12]Systems capable of negatively affecting safety or fundamental rights are considered high risk. These include: (1) AI systems subject to product safety regulations (cars, aviation systems, medical devices, elevators, etc.); and (2) AI systems related to the management of critical infrastructure, immigration management, access to essential public services, the management of welfare benefits, etc.

[13]As for compliance deadlines, the prohibitive rules are expected to come into force at the end of 2024, while the rules imposing obligations on companies that develop "foundational models" (models that serve as the basis for other AI products, such as GPT-4) will have to comply with the law within one year. The other obligations arising from the new legislation must be met within two years (Ryan-Mosley, 2024).

[14]Examples of these practices are subliminal marketing techniques and advertising, which are introduced into people's consciousness to substantially alter their behavior in a way that is potentially harmful to their interests (Sotero, 2023, p. 162).

democratic rights (management of the electoral register, signature recognition in postal voting, and biometric identification systems for access to voting). All systems using these techniques must guarantee: (1) data governance, so that quality standards are maintained and systems are kept free from bias and discrimination; (2) security and human supervision in all cycles;[15] (3) compliance with transparency duties regarding the system's operation; (4) registration in a database at community-wide level; and (5) successful completion of the compliance test, with a view to obtaining the corresponding certification.

Finally, medium or low risk systems, such as virtual assistants and chatbots, which do not directly affect privacy, initially do not pose significant risks to rights and freedoms, although they may eventually be used in voting recommendation applications (which are indeed becoming increasingly common). In this dimension, basic measures guarantee transparency so that users can understand how these systems work and what their main features are as well as avoid ideological or partisan prejudices.

In addition, recently and with a view to the 2024 European Parliament elections, a series of measures have been adopted which, although they do not deal specifically with AI, contain concrete references to its possible application in election campaigns. The measures derive from the "European Democracy Action Plan" (2020), which seeks to "empower citizens and increase democratic resilience across the Union by promoting free and fair elections, strengthening media freedom and combating disinformation." The plan was developed in phases, with milestones in November 2021 and December 2023, and relied on the cooperation of the European Parliament, particularly in the fight against foreign interference and disinformation through the work of successive special commissions on the subject. Among the measures observed, the following should be highlighted:

a. The Strengthened Code of Practice on Disinformation, adopted in 2022 (European Commission, 2022), replaced and strengthened the previous code (European Commission, 2018) and was signed by 34 bodies, including online platforms, advertising agencies, fact-checkers, academic institutions, and civil society organizations. Commitments which would apply to the use of AI include: demonetizing the spread

[15] According to Lage (2022, p. 62), "the life cycle of AI technically included the following phases: 1. design, data and modeling (planning, data collection and model construction); 2. development and validation (training and testing); 3. deployment; 4. supervision and perfecting (solving any problems that arise)."

of disinformation; guaranteeing the transparency of political advertising; reducing inauthentic behavior used to spread disinformation; cooperating with fact-checkers; and providing researchers with access to data. Within the working group created to supervise its application, of particular note was the creation of a transparency center,[16] which collates reports from participating platforms.[17]

b. The Regulation (EU) 2024/900 on transparency and segmentation in political advertising[18] seeks to support free and fair elections and also affects activities for which AI is being used in the following ways: (1) it broadens the concept of political advertising; (2) it reduces legal fragmentation and eliminates the usual obstacles to cross-border services; (3) it increases the transparency obligations of political advertising (defining it as that carried out by political parties but also including thematic advertisements), which must be identified as such and offer basic information about the sponsor, the election to which it is linked, the amount spent, and the targeting techniques used; and (4) it restricts the use of microtargeting and amplification techniques for this type of political advertising, which from now on can only be carried out with data collected directly from the subject, with their explicit consent for this particular use. Furthermore, (5) it prohibits, in any instance, microtargeting based on personal data related to race, ethnicity, or political opinions. Finally, in an effort to avoid external pressure, (6) it prohibits the contracting of advertising by organizations from third-party countries during the three months prior to the vote, in line with "the Judgement of the General Court of 25 November 2020 in the case T-107/19, in which it was affirmed that a party from a non-EU member state is not included in the definition of 'political party' if it does not involve EU citizens and is not recognized by the legal system of at least one member state, or established in accordance with the latter" and Opinion 01/2022 of the European Court of Auditors concerning the Commission's proposal for a regulation on the

[16] Available at: [https://disinfocode.eu]. Viewed: 11.04.2024.

[17] Available at: [https://disinfocode.eu/reports-archive/?years=2024]. Viewed: 11.04.2024.

[18] Available at: [https://data.consilium.europa.eu/doc/document/PE-90-2023-INIT/es/pdf]. Viewed: 11.04.2024. For a detailed study of the text see García Mahamut, 2023.

statute and funding of European political parties and European political foundations (García Mahamut, 2023; p. 90).

c. The Digital Services Act (DSA) also has a direct effect on elections, since it regulates the moderation of online content and harmonizes national regulations on illegal content, advertising, and disinformation, all of which are commonplace, as we have seen, in election campaigns. In particular: (1) it establishes mechanisms that allow users to report this type of content; (2) it guarantees that decisions made by platform moderators can be challenged; and (3) it strengthens the transparency of platforms, including, for example, the transparency of the algorithms used for recommendations. For very large platforms, which reach more than 10% of the EU population,[19] the obligation to mitigate systemic risks related to electoral processes is explicitly defined, which has given rise to a series of specific recommendations for the upcoming European Parliament elections which include: (4) reinforcing internal processes in function of the potential risk; (5) improving their capacity to respond to these behaviors; (6) the implementation of risk mitigation measures; (7) promoting official information on electoral processes, for example with media literacy initiatives; and (8) adapting their recommendation systems to empower users and reduce the monetization and virality of content that threatens the integrity of electoral processes. In addition, according to the new regulation on transparency and segmentation in political advertising mentioned above, political advertising must be clearly labeled as such and explicitly state when GenAI has been used. Cooperation is also encouraged between national and EU authorities, independent experts, and civil society organizations, especially the Working Group of the European Digital Media Observatory (EDMO). Finally, the adoption of a mechanism to respond to incidents that may have a significant impact on the results or electoral participation is recommended, as is the evaluation of the effectiveness of the measures through post-election analysis, which encourages third-party supervision to ensure that the measures applied are effective and respect fundamental rights.

d. Finally, the Recommendation on Inclusive and Resilient Electoral Processes in the EU addresses cybersecurity and the protection of

[19] Available at: [https://digital-strategy.ec.europa.eu/en/policies/list-designated-vlops-and-vloses]. Viewed: 11.04.2024.

election-related infrastructure, which should be classified as critical, as well as databases and processes, in line with the proposed Cyber Resilience Law. It proposes measures to minimize the risks of third-country interference through controlling the financing of political parties and their campaigns and reinforces the transparency of parties and candidates. Among the measures included are: the adoption of codes of conduct that facilitate electoral integrity and fair campaigns, promoting inclusive political discourse and prohibiting manipulative behavior, such as the generation or dissemination of falsehoods (including deepfakes), or messages that incite hatred as well as rejecting the use of inauthentic tactics, techniques, and procedures to disseminate or amplify political messages where AI plays an increasingly important role; and the promotion of independent control mechanisms for the fulfillment of acquired commitments. Finally, states are encouraged to protect the information environment and ensure that voters receive correct information, by promoting media awareness and literacy projects to combat information manipulation, interference, and disinformation related to the elections as well as strengthening rapid responses, both *pre-bunking* and *debunking*.

In short, the EU has created a regulatory framework which aims to prevent disinformation and foreign interference and which, within its comprehensive proposal, seeks to respond to the use of AI in electoral campaigns. The whole world will have the opportunity to observe the effectiveness of a legal framework – represented by this ambitious set of measures – which, although arriving somewhat late and newly applied in 2025, will form part of the regulatory–protective framework for the elections that will be held from then onwards in member states and in all likelihood in many others too.

Moreover, according to the new Regulation on Transparency and Guidance on Political Advertising mentioned earlier, political advertising must be clearly labeled as such, explicitly recording when AI has been used.

The European Commission has taken concrete measures in applying the DSA in the context of elections and information integrity.[20] The Commission's

[20]In April 2024, on the eve of the European elections, the Commission published guidelines for VLOPSE providers, after a public consultation, in which it recommended risk mitigation measures to protect electoral integrity (Commission Guidelines for providers of Very Large Online Platforms and Very Large Online Search Engines on the mitigation of systemic risks for electoral processes pursuant to Article 35(3) of Regulation (EU) 2022/2065). The Commission organized a stress test in the form of a simulation exercise

initial assessment, as well as the initial reports from the EDMO working group on the 2024 European elections,[21] agree that while false information and disinformation circulated around the elections, no major or systematic disinformation incidents capable of disrupting the elections occurred. However, more than 50 requests for information (RFIs) were sent to VLOPSEs (very large online platforms and search engines), some of which related to electoral risk mitigation measures and GenAI (such as so-called "hallucinations," where generative AI provides false information, the viral spread of deepfakes, and the automated manipulation of services that may mislead voters).[22] Among these was a request for information from Microsoft about the risks posed by its Bing Copilot AI chatbot and from Google Search, expressing concern about "the automated manipulation of services that can deceive voters." Information was also requested from Meta, X, Snapchat, TikTok, and YouTube for not doing enough to protect users from disinformation campaigns.[23] Lastly, the Commission has opened formal proceedings against X[24] and

to prepare the VLOPSE, civil society organizations, digital service coordinators, and European institutions for scenarios involving information manipulation and to practice coordinated responses within the framework of the DSA and electoral guidelines. (Commission stress test platforms' election readiness under the Digital Services Act – European Commission), prompting a series of dialogs with the actors involved (DSA Election Readiness – Roundtable with Platforms, Search Engines, and Digital Service Coordinators – European Commission) as well as an ad hoc working group. (Report on the European Elections Digital Services Act and Code of Practice on Disinformation. European Board for Digital Services, 2024. pp. 11–12). Available at: [https://ec.europa.eu/newsroom/dae/redirection/document/107587]. Viewed: 17.09.2024.

[21] Available at: [https://edmo.eu/blog/eu-elections-2024-the-battle-against-disinformation-was-won-but-the-attrition-war-is-far-from-over]. Viewed: 17.09.2024.

[22] Available at: [https://digital-strategy.ec.europa.eu/en/news/commission-sends-requests-information-generative-ai-risks-6-very-large-online-platforms-and-2-very]. Viewed: 17.09.2024.

[23] Report on the European Elections Digital Services Act and Code of Practice on Disinformation. European Board for Digital Services, 2024. Available at: [https://ec.europa.eu/newsroom/dae/redirection/document/107587]. Viewed: 17.09.2024.

[24] Available at: [https://ec.europa.eu/commission/presscorner/detail/en/ip_23_6709]. Viewed: 17.09.2024. Available at: [https://ec.europa.eu/commission/presscorner/detail/en/IP_24_3761]. Viewed: 17.09.2024.

Meta[25] for violations of the DSA during the elections. At the national level, many national digital service coordinators (DSCs) developed specific actions in this area.[26]

3.1.1.2 The Council of Europe

The Council of Europe Convention on AI and Human Rights, Democracy, and the Rule of Law (Council of Europe, 2024) seeks to establish a global legal framework to ensure that activities related to AI systems respect human rights, democracy, and the rule of law. Aimed at action by the signatory states, it includes principles on human dignity, transparency, accountability, privacy, and non-discrimination. It defines "AI systems" as "machine-based systems that, with explicit or implicit objectives, infer from the data they receive how to generate outputs such as predictions, content, recommendations, or decisions that can influence physical or virtual environments." Article 5 establishes the obligation of the signatories to adopt "measures ensuring that AI systems are not used to undermine the integrity, independence, and effectiveness of democratic institutions and processes, including the principle of separation of powers, respect for judicial independence, and access to justice" and to "protect their democratic processes in the context of activities related to AI systems, including fair access and participation in public debate and individuals' ability to freely form opinions." Among the responses, Article 16 establishes a risk and impact management framework for AI systems. Each party must take measures to identify, assess, prevent, and mitigate the risks that these systems may pose to human rights, democracy, and the rule of law. These measures should be proportionate to the context, severity, and likelihood of impacts. They should also include continuous monitoring and testing prior to use. It also provides for the possibility of imposing moratoriums or prohibitions on uses of AI incompatible with these values.

[25] Available at: [https://ec.europa.eu/commission/presscorner/detail/en/ip_24_2373]. Viewed: 17.09.2024.

[26] Report on the European Elections Digital Services Act and Code of Practice on Disinformation. European Board for Digital Services, 2024. pp. 13–15. Available at: [https://ec.europa.eu/newsroom/dae/redirection/document/107587]. Viewed: 17.09.2024.

3.1.1.3 North America

Within the general approach, Canada has regulated automated decision-making since 2019[27] to reduce the risks related to errors or discrimination. This regulation establishes the principle of transparency, with the obligation to publish in a prominent place that a decision will be made by an automated system; it also establishes the need to make public any source code held by the government as well as a series of preventive measures, such as prior testing to detect unintentional bias in the data, processes for monitoring the results of these decision-making systems, and the guarantee of human intervention as well as the right to appeal this type of decision.

In the United States, in addition to the different regulations in the various states,[28] in late 2023, the *Executive Order on the Safe, Secure and Trustworthy Development and Use of Artificial Intelligence* was published, which obliges the main AI developers and cloud-computing providers to share evidence and information related to national security issues as well as other matters considered crucial to government affairs.

While rescinded by Donald Trump within hours of his assuming office on January 20, 2025, it was conceived with the aim of reconciling the positive and negative aspects of AI in the face of the "rapid advance of its capabilities." The decree sought to "confront technological challenges and contribute to the prosperity and security of the people," considering that the "irresponsible use" of the technology in vogue "could exacerbate social damage such as fraud, discriminatory treatment and disinformation," imposing threats "to free competition, the labor market and national security" (United States, 2023). According to the executive order, development and responsible use policy implies that the use of AI tools must:

1. *Guarantee protection and security*, which requires standardized, assessments, policies, institutions, and robust, repeatable, and reliable mechanisms to mitigate risks prior to use. These requirements fall mainly on AI applications in fields such as biotechnology,

[27] Directive on automated decision-making of 1 April 2019. Available at: [https://www.tbs-sct.canada.ca/pol/doc-eng.aspx?id=32592]. Viewed: 4.5.2024.

[28] National Conference of States Legislatures (2022). Legislation Related to Artificial Intelligence. Available at [https://www.ncsl.org/technology-and-communication/legislation-related-to-artificial-intelligence]. Viewed: 4.4.2024; U.S. Chamber of Commerce (2022). State-by-State Artificial Intelligence Legislation Tracker. An interactive map showing states' action to legislate on artificial intelligence. Available at: [https://www.ncsl.org/technology-and-communication/artificial-intelligence-2025-legislation]. Viewed: 4.4.2024.

cybersecurity, critical infrastructure (such as energy supply), and other issues affecting national security. To this end, trials, assessments, and post-launch performance tests can provide a solid basis for addressing the risks of AI without sacrificing its benefits.

2. Promote innovation and responsible competition to make the United States a leader in AI.
3. Support American workers by creating new jobs and industries and incorporating collective bargaining to ensure that the working class is able to benefit from these opportunities. In this respect, the decree mentions the provision of vocational training to support the supply of skilled labor. It also aims to ensure that the emergence of AI neither violates rights or worsens the quality of work, nor does it encourage undue workplace surveillance or introduce new health and safety risks.
4. Explore how AI might illegally discriminate and facilitate the administration of state programs and benefits (at the federal level) to promote equity and civil rights.
5. Protect the interests of consumers by applying current legislation and enacting new safeguards against fraud, unintentional damage, discrimination, violation of privacy, and other possible AI-derived harm.
6. Protect privacy and civil liberties by ensuring that the collection, use, and storage of data is carried out legally and securely, mitigating risks to privacy and confidentiality.
7. Manage the risks arising from the use of AI by the federal government, strengthening its internal capacity to regulate, govern, and support the responsible use of AI.
8. Lead global social, economic, and technological progress, including effective leadership in pioneering systems and safeguards for the responsible deployment of AI. This action includes committing to the development of a framework with global allies and partners that mitigates the risks of AI, unlocks its positive potential, and unites to overcome shared challenges (Dougall; Ostrowski, 2024).

With specific regard to the use of AI in elections, following the use of fake audios purporting to be President Biden in robocalls during the New Hampshire primaries, the FCC has declared robocalls using GenAI-generated voices to be illegal.[29] In addition, at least four bills have been introduced in the US Congress that specifically address the use of deepfakes and other

[29] Available at: [https://docs.fcc.gov/public/attachments/FCC-24-17A1.pdf]. Viewed: 4.4.2024.

manipulated content in federal elections and at least four others that address this content more broadly. At a state level, new laws have been passed in recent years that prohibit or otherwise restrict deepfakes and other deceptive media in electoral advertising and political messages in states as ideologically diverse as California,[30] Minnesota,[31] Texas,[32] Washington,[33] and Florida.[34]

California, having ensured that election-related proceedings have priority on the judicial agenda, prohibits the distribution of deceptive audiovisual media affecting a candidate within the 60 days prior to an election, unless a warning is included indicating that the media has been manipulated, and establishes legal procedures for candidates whose images or voices are used in a misleading way. However, it provides exemptions for certain media.

Minnesota and Texas have established the offense of using deepfake technology to influence an election, in the 90 days prior to the election in Minnesota and 30 days in Texas. In both cases, establishing the intent to harm a candidate or influence the outcome of an election is required, and different penalties are set according to the seriousness of the offense.

Washington, with perhaps the most comprehensive regulations, in addition to requiring the courts to respond quickly, makes it easier for the victim to access precautionary measures or other types of equitable relief to prohibit the publication of such synthetic media or to bring an action for general or specific damages against the sponsor. It establishes the obligation to label content. Finally, it establishes the liability of the sponsors of electoral communications that contain synthetic media but not of the media that disseminates them, except in certain circumstances; it exempts the providers and users of interactive computing services from being treated as the publisher or spokesperson of any information provided by

[30] Available at: [https://leginfo.legislature.ca.gov/faces/billTextClient.xhtml?bill_id=201920200AB730]. Viewed: 4.4.2024.

[31] Available at: [www.revisor.mn.gov/bills/text.php?number=HF1370&type=bill&version=3&session=ls93&session_year=2023&session_number=0]. Viewed: 4.4.2024.

[32] Available at: [https://capitol.texas.gov/tlodocs/86R/billtext/html/SB00751F.htm]. Viewed: 4.4.2024.

[33] Available at: [https://lawfilesext.leg.wa.gov/biennium/2023-24/Pdf/Bills/Session%20Laws/Senate/5152-S.SL.pdf?q=20231003125542]. Viewed: 4.4.2024.

[34] Available at: [https://www.flsenate.gov/Session/Bill/2024/919/BillText/er/PDF]. Viewed: 4.4.2024.

another information content provider but allows for them to be liable in certain circumstances.

Finally, Florida has established requirements to label any content that uses GenAI, whether in whole or in part. The law further specifies that when such technology is used to create synthetic representations of real people, particularly in content intended to discredit candidates or mislead voters on electoral issues, detailed warning requirements must be met across all media formats, including written text, television, video, Internet, audio, and other forms of graphic communication. Any person identified as responsible for a political advertisement that does not include the required discharge of responsibility commits a first-degree misdemeanor, punishable by law.

3.1.2 Self-Regulation

Although "(as a) general rule, ethical commitments are limited to making a self-declaration, which describes an internal structure with an ethics committee and an advisory board, but their function and powers remain unclear" (Hernández Ramos, 2023, p. 216), in the context of a regulatory vacuum, two initiatives by different technology companies related to the use of technology in elections have been launched in recent months: (a) the *Tech Accord to Combat Deceptive Use of AI in the 2024 Elections*;[35] and (b) the *Voluntary Electoral Integrity Guidelines for Technology Companies*, stipulated by the International Foundation for Electoral Systems (IFES).[36]

The Tech Accord includes a series of actions and commitments to neutralize the dissemination of AI-produced content produced that could "compromise the integrity of electoral processes:" material such as images, audio, and video generated with AI-based tools "that falsify or misleadingly alter

[35] Announced in February 2024, the agreement in question was signed by Adobe, Amazon, Anthropic, ARM, ElevenLabs, Google, IBM, Inflection AI, LinkedIn, McAfee, Meta, Microsoft, Nota, OpenAI, Snap, Stability AI, TikTok, TrendMicro, TruePic, and X. Its full text can be accessed at the following address: [https://www.aielectionsaccord.com].

[36] The guidelines were launched in March 2024 and have been endorsed by the electoral bodies of Australia, Mauritania, Haiti, the Philippines, the Dominican Republic, Romania, and El Salvador as well as by experts from different countries, international organizations such as the Organization of American States (OAS), and civil society organizations such as Freedom House, the Atlantic Council, the Carter Center, or the German Marshall Fund, and IDEA. In addition, some technology companies such as Google, Meta, Microsoft, Snap, and TikTok have adopted the principles established. The full document can be consulted at the following address: [https://electionsandtech.org].

the appearance, voice or actions of political candidates, electoral officials and other stakeholders." These commitments are:

1. To develop and apply technology to mitigate the risks related to deceptive electoral content created with AI systems, including open-source ones.
2. To evaluate AI models within the scope of this agreement to understand the risks they may pose in relation to the production of misleading electoral content.
3. To control and detect the distribution of such material on their platforms.
4. To take measures to deal appropriately with the misleading information distributed by their services (labeled).
5. To encourage adaptation among sectors liable to harmful electoral content.
6. To provide the public with information that makes mitigation measures transparent.
7. To collaborate with academic and civil society organizations to develop follow-up plans.
8. To support efforts to promote public awareness, media literacy, and resilience throughout society.

The Voluntary Guidelines, on the other hand, establish a framework for: (a) improving the relationship between electoral authorities and technology companies; (b) establishing clear policies and processes; (c) information sharing; (d) ensuring that voters have access to high-quality information; (e) improving communication channels between technology companies and electoral authorities; and (f) democratization, making it easier for companies of all types and sizes to contribute to electoral integrity. These commitments are:

1. To determine how to prioritize participation in elections around the world through a process that takes into account a range of factors, including human rights and democratic principles, the relevant use of the company's products and services, and resource considerations.
2. To consult with civil society worldwide, when necessary and appropriate, through the established channels or events, in order to inform

companies about the understanding of the national electoral context and the commitment to the electoral authorities, as appropriate.

3. To establish and publicize clear policies and processes on content, activities, civil unrest, and violence related to the elections.
4. As far as possible, these policies and processes should be made accessible in an appropriately localized way.
5. To centralize information on policies, products and services that could be useful for electoral authorities and civil society.
6. To establish planning processes that take into account the deadlines and capacity of the electoral authorities, including before the electoral period and after the elections.
7. Identifying bilateral contact points at an early stage, organizing coordination and taking local capacity into account.
8. To allow access to reliable information on elections and voters, as appropriate.
9. To establish a strategy to tackle disinformation and misinformation about electoral participation.
10. These policies should be publicized and made known to electoral authorities and other interested parties.
11. To establish and make available to the electoral authorities communication channels that can be used to deal with critical incidents during the electoral period.
12. To provide public information on paid political and/or electoral content worldwide.
13. To provide information that can facilitate investigative efforts into issues related to paid political and/or electoral content.
14. To maintain appropriate coordination mechanisms and operations beyond the period immediately following an election.
15. To support post-election engagement with stakeholders in the election.
16. To support post-election analysis carried out by the electoral authorities and other interested parties, as appropriate.

In both cases, the principles and commitments are generic and do not explicitly contemplate, for example, the prohibition or elimination of deepfakes. Nor do they include details on how the resolutions will be applied or even a timetable for action.

For its part, as we have seen, Meta[37] has committed to labeling AI-generated images on all its platforms and has also declared that it will prohibit political campaigns from using its new AI-generated advertising products and will require political advertisers to reveal when they use AI tools to modify or create ads on Facebook and Instagram. Its oversight board[38] – a body created by the company to review Meta's decisions on content moderation – in a document full of warnings about the role of the platform in electoral processes, points out the need to establish clear standards for content generated by GenAI and fakes (deep or cheap).

Google, for its part, in addition to labeling this type of content on YouTube,[39] has restricted election-related answers on its Gemini chatbot since early 2024, offering the answer "I'm still learning how to answer this question. Meanwhile, you can consult Google Search." The company also restricts the generative search experience related to elections, applying these restrictions from the beginning of 2024.

Finally, OpenAI,[40] the creator of ChatGPT and DALL-E, is committed to preventing abuse, promoting transparency, and ensuring the integrity of elections around the world. It seeks to anticipate and prevent the potential misuse of its tools, especially in an electoral context. To achieve this, it has put in place a series of measures to detect and address abuse, such as identifying AI-generated content and preventing the creation of chatbots that imitate candidates. To this end, a credential and digital watermark system from the Coalition for Content Provenance and Authenticity (C2PA) will be introduced to identify AI-generated images, and the launch of a provenance classifier to detect DALL-E-generated images, even if they have been modified, has been announced. In addition, a "reporting" function will be implemented for users, allowing them to report possible violations in the use of personalized GPTs.

[37] Available at: [https://about.fb.com/news/2024/02/labeling-ai-generated-images-on-facebook-instagram-and-threads]. Viewed: 4.5.2024.

[38] Available at: [https://www.oversightboard.com/wp-content/uploads/2024/04/Oversight-Board-Elections-Paper-May-2024FINAL.pdf]. Viewed: 4.5.2024.

[39] Available at: [https://www.npr.org/2023/11/14/1212986395/youtube-will-label-ai-generated-videos-that-look-real]. Viewed: 4.5.2024.

[40] Available at: [https://openai.com/blog/how-openai-is-approaching-2024-worldwide-elections]. Viewed: 4.5.2025.

3.2 THE RESPONSE OF ELECTORAL BODIES (BRAZIL)

In the absence of a regulatory response, some electoral bodies have begun to take steps to provide one. Among them, Brazil stands out for its comprehensive and regulatory character: in the absence of a regulatory framework – and in a scenario of imminent threats – on 27 February 2024, the TSE took advantage of the prerogative included in Article 57-J of the Elections Law (Law No. 9.504/97[41]) to approve a set of rules aimed at imposing discipline on the use of AI in campaigns, specifically in art. 9°-B to 9°-H of Resolution No. 23.610/2019, updated by Resolution No. 23.372/2024. In the process, the court took an important step forward in the active defense of the process of political renovation and of Brazilian democracy itself.[42] Due to its novelty, its pioneering nature, and its relevance to the study of AI's impact on election campaigns, a detailed explanation of its content is merited.

According to Ana Frazão's (2024) summary, the regulatory text, as approved, establishes the following main objectives:

1. The prohibition of deepfakes.
2. The inclusion of warnings when AI has been used in election material.
3. Restrictions on the use of robots as mediators in contact with voters, prohibiting the simulation of a dialog with a candidate or any other person.

[41] "Article 57-J. The Superior Electoral Court will regulate the provisions of Articles 57-A to 57-I of this law in accordance with the context and technological tools existing at the time of each election and will promote, for vehicles, parties and other interested entities, the formulation and widespread dissemination of rules of good practice regarding election campaigns on the Internet."

[42] Faced with the traumatic experience of January 8, it was "more than urgent to think about regulating the use of artificial intelligence for political and partisan purposes. If there was already considerable concern about the manipulation of the information flow after the election of Trump and Bolsonaro [...], such fears have become even more relevant with the growing advances of artificial intelligence, especially in its generative mode, which even makes so-called 'synthetic propaganda' possible. Faced with these risks, the TSE recently issued a series of resolutions to try to prevent the creation and dissemination of false content during the elections, as well as imposing transparency obligations, so that the recipients of content produced with artificial intelligence are alerted to this circumstance" (Frazão, 2024).

4. Holding the big tech companies responsible if they do not immediately remove content containing disinformation, hate speech, Nazi and fascist ideology as well as anti-democratic, racist, and homophobic content.[43]

3.2.1 Authorization to Use Artificial Intelligence

To begin with, Article 9-B[44] expressly authorizes the use of GenAI for content production[45] as well as the use of synthetic intelligence solutions for the improvement, editing, or adaptation of communication material. GenAI is been expressly authorized for the purposes of creation (complete conception),

[43] In an alternative summary specifically focused on the production of pieces with AI, the resolution establishes three types of use guided by different rules: "The first group includes ordinary uses, for example, to make adjustments to images and sound or to create vignettes and other graphic elements. The use of this technology is free and does not require identification. The rule makes sense, since nowadays any photograph taken with a mobile phone, for example, automatically goes through a graphic adjustment process using AI tools, something that does not arouse any significant reactions from the public. The creation of synthetic/artificial content or more advanced adjustments, on the other hand, requires the campaign to clearly inform the electorate that the material has been produced using AI. There is an intermediate level of regulation here, which assumes that the information transmitted to the recipients of the message will allow them to make a more realistic judgment of what is being presented. Finally, there are uses that are strictly forbidden, which can constitute abuse of economic power or improper use of the media, which can lead to the annulment of candidacies and mandates, as well as the imposition of the penalty of ineligibility for eight years. These include deepfakes, regardless of their purpose – whether positive or aimed at attacking rival candidates – as well as the dissemination of disinformation boosted by the use of AI. Also prohibited are chatbots and the like that pretend to be the candidates themselves, giving the impression to the electorate that they have entered into direct communication with that person" (Neisser; Mattiuzzo, 2024).

[44] "Article 9-B. The use in electoral publicity, in any of its forms, of synthetic multimedia content generated through artificial intelligence to create, replace, omit, merge or alter the speed or superimpose images or sounds imposes on the person responsible for the publicity the duty to inform, in an explicit, prominent and accessible manner, that the content has been manufactured or manipulated and that the technology has been used."

[45] GenAI can be defined as "a methodology used [...] to describe any type of artificial intelligence that contains unsupervised learning algorithms used to create new digital images, video, audio, text or code. The purpose is to generate *synthetic data* that can pass a Turing test." Within this strand, the evolutionary leap is made with the introduction of generative adversarial networks (or generative antagonistic networks) – GANs – conceptualized as "a type of artificial intelligence algorithm that works like a game in which one neural network competes with another to generate new images or content" (Moreno, 2023, pp. 36, 60).

substitution (modification of aesthetic, sound, or content components), omission (deletion of lines, participations, visual elements, or adjacent details), blending (audio mixing, joining of sequences), alteration of speed (reduction of viewing time), superimposition of images (for example, so that a studio recording is projected over an external recording), or sounds (for example, so that the voice of one speaker prevails over that of another when played simultaneously).

However, as a general rule, the existence of (totally or partially) synthetic content must be indicated by a disclaimer, the purpose of which is to ensure that the audience is not deceived or misled, that is to say that the general public will have no doubts about the origin or nature of the information material. In particular, the TSE takes on board the spirit of the recommendation contained in a report drawn up by the European Commission, which stresses the importance of transparency as a way of reducing the possibilities of manipulation (Denemark, 2024, p. 130).[46]

According to §1 of the Article under discussion: (1) the warning must precede the main message in audio pieces (given that listeners may not listen to the piece in its entirety); (2) in static images (photographic records, realistic representations, graphic arts) the warning should form part of the visual composition, with the use of a label or watermark (taking into account the possibility that people may not read subtitles) as well as being reinforced, when published in digital media, by audio-description mechanisms;[47] (3) in audio and video media (video with audio), the requisites are combined to cumulatively require a warning at the beginning of the piece, the use of a label or watermark, and an audio description; and (4) in the case of printed material (such as guides, proposal booklets, or booklets), the warning must

[46] However, contrary to the ambitions of the European scenario, the TSE missed the opportunity to also require the presence of legal warnings in messages with personalized political content based on data-mining activities and the elaboration of psychometric profiles. According to the experts, this regulation would be one of the most important to prevent AI techniques from leading to harmful speech and disinformation (Denemark, ibidem).

[47] Audio description is a technological tool capable of translating images into words and is necessary, in particular, to meet the needs of visually impaired people, whether they are blind or have reduced vision. In technical terms, audio description can be implemented, among other ways, by: (1) merging with the original audio, offering it as an option to the listener or viewer; (2) making it available on headphones (as is usual, for example, in museums); (3) embedding it in hidden texts that can be accessed, optionally, on reading devices or software (commonly called "alt text"); and (4) through hashtags that are ever more present on social networks, such as #visuallyimpaired.

be reproduced on all pages (taking into account the possibility that readers will not read all the material).

However, the Resolution itself stipulates cases in which requirement for the warning is waived (Article 2),[48] starting with neutral aesthetic changes, aimed, for example, at eliminating noise, correcting light, increasing pixelation, etc. It is understood, however, that the flexibility in question, given the outstanding importance of image as a cognitive determinant of the vote (Alvim, 2019, p. 258), does not extend to adjustments aimed at positively reconstructing the image of a candidate, which can be appreciated, for example, in techniques for de-aging (artificial rejuvenation) or makeovers (beautification or increasing likability or attractiveness).

The need to inform is also waived in the case of electoral symbols, such as logos or vignettes, which is justified by the fact that these elements, in a strict sense, fulfill the simple purpose of identification, more or less lacking relevance to the question of information tactics, persuasion, mobilization, and convincing voters.

Finally, AI warnings are not necessary for customary montages, such as those that bring together candidates with sponsors or political sympathizers in a single shot, insert a false background over chrome, or alter a studio photo to place the candidate in any setting.

Section 3 of Article 9-B[49] expressly allows the use of avatars[50] or chatbots (conversational applications) in Brazilian campaigns, provided that they are accompanied by a clear warning to users, one which eliminates any doubts

[48] "Paragraph 2 of this article does not apply to: I – adjustments intended to improve the quality of the image or sound; II – the production of graphic elements of visual identity, signs and logos; III – marketing resources that are commonly used in campaigns, such as montages of images in which candidates and supporters appear in a single photographic record which are used in the production of printed and digital advertising material."

[49] "Paragraph 3. The use of chatbots, avatars and synthetic content as an artifice to mediate campaign communication with physical persons is subject to the provisions of the head of this article, and any simulation of a dialog with a candidate or other real person is prohibited."

[50] Digital avatars are nothing more than fictional characters, humanoid or otherwise, created digitally for communicative purposes which interact with users through text or more often than not, some form of speech, especially in virtual media. In the electoral sphere, this strategy gained visibility with the presentation of iXóchitl, the virtual spokesperson of Xóchitl Gálvez, potential candidate for the presidency of Mexico, in December 2023 (Urquijo, 2023).

about the synthetic origin of interactions that will ensue.[51] As a result, it is up to campaigns who use chatbots to make it evident to users that the messages and answers offered by the tool are derived from automatic commands and that, therefore, there is no direct dialog with the candidate, campaign staff, or any physical person.[52] In short, the regulations prohibit the simulation of human interactions but authorizes the use of this tool so long as transparency is maintained.

For those responsible, failure to comply with these requirements entails the possibility of content being removed or the communication service being made unavailable, not only as a consequence of a court order but also on the platforms' own initiative, in the manner set out in Article 9-B, §4 of the Resolution. In addition, the sanction of withdrawal or disconnection may be accompanied by other penalties provided for in the legislation so that the covert use of avatars to disseminate illegal messages (defamatory, slanderous, harmful, hateful, or disinformation) may result in the fine applicable to cases of anonymity, according to Article 57-D, §2, of the Electoral Law.

3.2.2 Prohibitions on the Electoral Use of Artificial Intelligence

Article 9-C of the regulations prohibits the use in advertising of artificially produced or manipulated content to disseminate disinformation narratives capable of tipping the balance or compromising the integrity of the election.[53] The rule explicitly prohibits disinformation that targets electoral institutions

[51] "[T]he use of chatbots, avatars and synthetic content as an artifice to mediate in campaign communication with physical persons prohibits any simulation of interlocution with a candidate or other real person. This brings us to the recent decision by the Federal Communications Commission, the regulatory body for the telecommunications sector in the United States, which declared phone calls with voice features generated by Artificial Intelligence illegal. In a statement, the FCC stressed that voice cloning technology has been used in automated calls to extort vulnerable family members, impersonate famous people and to misinform voters" (De Teffé, 2024).

[52] Given that chatbots can be used – and are often used – to collect voters' personal data, it follows that offering this type of service entails the implicit obligation to comply with the rules applicable to data protection in an electoral context.

[53] "Article 9-C. The use in electoral publicity, in whatever form or mode, of content that is produced or manipulated in order to disseminate notoriously false or decontextualized information with the potential to damage the balance of an election or the integrity of the electoral process is prohibited."

and, consequently, aligns with Article 2 of Resolution No. 23.714/2022,[54] which regulates the exercise of political power in the event of false or decontextualized content that undermines public confidence in the proceedings in question. The combination of both provisions shows that: (1) anti-system disinformation can be halted both through administrative orders based on police power, regardless of the provocation, and through legal proceedings for illegal propaganda which originate in claims brought by actors qualified to do so; and (2) disinformation against candidates or parties, in principle,[55] falls outside the scope of police power and should be dealt with mainly in the context of legal proceedings for illegal propaganda, based on Article 96 of the Electoral Law.

It should be noted that the regulations prohibit false or decontextualized content that generates an imbalance in the electoral contest, thus invoking an expression with a double meaning. Firstly, an imbalance can be understood as something that undermines the equal playing field in an election, generating or increasing an advantage that ultimately determines the participants' chances of victory. In this sense, the provision specifically prohibits fake news that has the potential to damage the reputation of any competitor (disinformation targeting candidates and parties). However, the concept of imbalance can also be associated with the notion of agitation, lack of control,

[54] "Article 2. In the terms of the Electoral Code, the dissemination or sharing of notoriously false or seriously decontextualized information which affects the integrity of the electoral process, including the processes of voting, vote-counting and tallying the votes, is prohibited. § Paragraph 1: Once the occurrence provided for in the preamble has been verified, the TSE, in a motivated decision, will order the platforms to immediately remove the URL, URI or URN, under penalty of a fine of between R$ 100 000 (one hundred thousand reales) and R$ 150 000 (one hundred and fifty thousand reales) per hour of non-compliance, from the end of the second hour after receipt of the notification. § Paragraph 2: Between the day before and the three days following an election, the fine in Paragraph 1 will apply from the end of the first hour after the notification is received."

[55] Exceptionally, however, disinformation against candidates or parties can be the object of spontaneous objection, in the event of the occurrence described in Article 4 of TSE Resolution No. 23.714/2022: "Article 4. The systematic production of disinformation, characterized by the constant publication of false or decontextualized information about the electoral process, warrants the temporary suspension of profiles, accounts or channels maintained on social networks, subject to the requirements, deadlines and consequences set out in Article 2. Sole paragraph. The order to which the preamble refers will include the suspension of the registration of new profiles, accounts or channels by those responsible or under their control, as well as the use of previously registered profiles, accounts or contingency channels, under penalty of the offense provided for in the Electoral Code."

or instability. From this point of view, Article 9-C also protects the normalcy of the process by prohibiting controversial false content aimed at creating or reinforcing inter-group hostility, "moralistic violence" (Aguado Terrón; Villaplana Jiménez, 2023, p. 206), and digital lynching (Vallespín, 2021, p. 78) that foster division, conflict, and social animosity (Lardiez, 2021, p. 15) toward supervisory and control bodies (disinformation targeting institutions).

Section 1 of Article 9-C[56] establishes that synthetic audiovisual content cannot use the biometric records (voice or image) of human figures, whether living, deceased, or invented,[57] even if, in the first two cases, authorization is obtained from the cloned person. The wording, however, is dubious, since it is not clear whether the prohibition of cloning with AI is absolute, applying to all cases (a reading of §1 in isolation) or whether, in the opposite sense, it would only apply when the message that favors or benefits a candidate contains disinformation elements (a reading of §1 in relation to the Article's heading).

The issue is quite important, since the first interpretation leads to the conclusion that human presence in synthetic videos is only allowed in neutral content pieces (for example, for the dissemination of general information, such as campaign schedules) but not in persuasive products, which ultimately constitute the really valuable videos for campaigns. As a result, generative techniques would be restricted even in animated videos, since the rule analyzed, read in isolation, textually prohibits not only the synthetic representation of living or dead people, whether famous, anonymous, or historical but also the act of "creating [...] a fictitious [...] person."

However, the alternative reading invites this for two reasons: firstly because the topological context, at least in theory, leads to the interpretation of §1 in the light of the rule that opens the Article (the preamble); secondly because §1 ends with an explicit reference to the expression "deep fake" (deepfake). If, according to this line of thinking, §1 is considered to be an

[56] "Section 1. Prohibits the use, to harm or favor a candidacy, of synthetic content in audio or video format, or a combination of both, that has been generated or digitally manipulated, even with authorization, to create, replace or alter the image or voice of a living, deceased or fictitious person (deep fake)."

[57] It should be noted that the use of GenAI to generate realistic videos that depict deceased political leaders has become commonplace around the world, and has proliferated as a campaign strategy in several countries, such as India (Mukherjee, 2024) and in Argentina with Sergio Masa's advertising pieces during the presidential elections (Durães, 2023).

order that interacts with the preamble according to a principle of continuity, the conclusion would be that synthetic content can count on the presence of humans as long as it does not transmit disinformation. However, this conclusion is flawed, especially given that disinformation content is already in itself prohibited, regardless of any other aspect of form or background (Article 9-C, *caput*). In this reading, the rule would be superfluous, innocuous, and redundant.

As a result, it is understood that for the moment, the TSE, in a conservative position, has opted to prohibit the synthetic creation of human representations to prevent voters from being misled into interpreting the artificial characters as sympathizers, supporters, or spokespeople of flesh and blood. In other words, within persuasive communication, GenAI is unchained exclusively for clearly fictional content, with hyper-realism excluded as a measure to avoid fraud or self-induced errors, which denotes an implicit fear of a possible overcoming of a modern version of the Turing Test, historically related to the discovery of the ability of machines to act like humans without being detected.[58]

Although the Resolution (Article 9-C, §2[59]) establishes that non-compliance with this rule "constitutes" a supposition of abuse of power or improper use of communications media, it is argued that the nullification of the order or the declaration of ineligibility must be interpreted as a possibility rather than as a deterministic imposition. This is because setting a maximum gravity by absolute presumption would be unconstitutional because it would violate the principles of the proportionality, reasonableness, and individualization of penalties as well as entering in conflict with

[58] Created by Alan Turing, a prominent researcher in the fields of computer science and AI, the test is similar to an imitation game with three participants: two humans and a computer. "The evaluator, a human, asks open questions to the other two (a human and a computer) in order to determine which of them is the human. If the evaluator cannot make the distinction, it is assumed that the computer is intelligent. [...] The brilliance of this concept is that there is no need to check whether the machine really knows anything, whether it is aware of itself or even whether it is correct. Instead, the Turing test indicates that a machine can process large amounts of information, interpret speech and communicate with human beings [without being discovered]" (Taulli, 2019, pp. 18–19).

[59] "Section 2. Failure to comply with the provisions of the preamble and §1 of this article constitutes abuse of political power and improper use of the media, giving rise to the annulment of registration or the mandate, and imposes the declaration of responsibility in the terms of §1 of Article 323 of the Electoral Code, without prejudice to the application of other measures that may proceed in relation to the irregularity of the publicity and the illegality of the content."

the provisions of Article 22, XVI of the Complementary Law No. 64/1990,[60] which imposes, for these cases, the need for a case-by-case examination governed by a precise assessment of the "gravity of the circumstances."[61]

3.2.3 The Platforms' Obligations

Article 9-D[62] imposes on the platforms[63] the obligation to adopt and publish measures to prevent or reduce the circulation of disinformation narratives that could jeopardize the smooth running of elections. Although the term is not used, in this sense the Resolution establishes a duty of care,[64]

[60] "Article 22. Any political party, coalition, candidate or Public Electoral Ministry may make a representation to the Electoral Justice, directly to the General or Regional Magistrate, reporting facts and indicating evidence, signs and circumstances, and requesting the opening of a judicial investigation to investigate the improper use, misuse or abuse of economic power or power of authority, or the improper use of vehicles or means of communication, in favor of a candidate or political party, in accordance with the following procedure: [....] XVI – for the configuration of the abusive act, the potential of the act to alter the result of the election will not be considered, but only the seriousness of the circumstances that characterize it."

[61] Furthermore, the possibility of a criminal offense, despite the express mention, is not always contemplated, since Article 323 of the Electoral Code has publicity "in relation to parties or candidates" as an element of the type. Therefore, disinformation against electoral institutions can give rise to an investigation into abuse of power as well as being the subject of a representation requesting removal from office. However, it cannot be criminally investigated from the perspective of this crime.

[62] "Article 9-D. It is the duty of the provider of Internet applications that allow the dissemination of political-electoral content to adopt and publicize measures to avoid or reduce the circulation of notoriously false or seriously decontextualized information that may affect the integrity of the electoral process, including: [...]."

[63] In the strictest sense, the obligation falls on any platform "that allows the dissemination of electoral-political content." In practice, however, there are no reports of platforms banning debates on these topics, so despite the potentially restrictive signaling, the rule has general application.

[64] The duty of care consists of a "legal obligation which establishes that a person or organization is responsible for their actions or omissions which cause damage to third parties." Considered a "fundamental principle of civil liability," the duty of care "requires people and organizations to adopt reasonable measures to avoid damage to third parties, and applies to anyone who may be sufficiently affected by the actions or omissions of the person or organization in question" (Campos *et al.*, 2023). As a result of the debates related to the positivization of the duty of care in various contexts, such as in the United States and Europe, for some years now technology companies have been introducing changes to the way certain content is presented to minimize the impact of disinformation, especially

determining the adoption of a proactive stance aimed at preventing, or at least acting upon, the presence of disinformation in the digital environment targeted at electoral institutions. This clearly has the objective of confronting a new surge of anti-democratic movements, taking into account the recent memory of the traumatic coup attempt that took place on 8 January 2023. Among other issues, these measures point to a debate centered on improving the platforms' community standards – in particular the dynamics of moderating messages. In this regard, it should be noted that:

> All the major social-media companies have created specific content-moderation systems. The majority adopt a customer-service approach: users (and, increasingly, computer programs) are asked to identify problematic content or behavior; the platform moderators then carry out a behind the scenes procedural review, and decide whether or not to delete the flagged posts based on their own guidelines and judgments. It can be said that these procedures, developed over more than a decade, have been able to meet the needs of the platforms, while maintaining the promise of a sufficiently secure community for a satisfactory number of users [...].
>
> That would be the best version of the story. Another version states that a toxic culture of hatred, especially against women and minorities, seems to have taken root on social networks, one that is blithely tolerated by platform managers, eager to promote their own ideas of freedom of expression and benefit from the data they collect along the way; that legislators in Europe and elsewhere demand stricter intervention by the platforms against hate speech and terrorist propaganda; that an avalanche of fraudulent news and conspiracy

---

by adapting the algorithms that select and order the information viewed by users. In this respect: "YouTube [...] has used algorithms since the end of 2016 that decide which content is automatically suggested after a viewing, adding a new criterion of social responsibility. In this way, it has addressed the need to correct the tendency to favor the recommendation of videos with extreme content in order to maximize views, something of which the company had often been accused. It has also eliminated millions of channels for violating its guidelines and has begun to show more content from authorized sources and belonging to traditional media in top search positions. In addition, [...] initiatives have been launched to promote quality journalism and fact-checking. [...] Facebook has taken different initiatives to reduce the spread of fake news and inauthentic content in general, adapting its internal algorithms so that content from suspicious sites has less presence in users' news feeds" (Sánchez Muñoz, 2020, p. 115).

> theories erodes trust and influences voters' way of thinking, to a point that is often enough to decide national elections. Even a high level of success in large-scale moderation continues to allow hundreds of thousands of mistakes and hundreds of thousands of oversights, and each one of them represents an offended, misled or unprotected user.
>
> (Gillespie, 2020, pp. 329–330)

Although well intentioned, the regulation is controversial, among other factors because: (1) it establishes an express duty outside any legal precept (the prerogative inscribed in Article 57-J of the Electoral Law must be read in consonance with Article 105 of the same law);[65] (2) regulatory power exists to develop, and not to contradict, current statutes,[66] with the rule apparently clashing with the liability regime for intermediaries provided for

[65] "Article 105. Until March 5 of the year of the election, the Superior Electoral Court, taking into account the regulatory nature and without restricting rights or establishing sanctions other than those provided for in this law, may dictate all the necessary instructions for its faithful fulfillment, previously listening to, in a public hearing, the delegates or representatives of the political parties."

[66] As the doctrine explains, the role of the TSE in the exercise of its regulatory function is reduced, since it is subject to the following limitations: "(a) the limitation of material character: since it has the status of an ordinary law, it cannot deal with matters that the constitution reserves for a complementary law, such as those relating to ineligibility and the organization of the Electoral Justice system; (b) the limitation of logical character: while it may have force of law, its regulatory character does not permit it to contradict primary normative acts; (c) the limitation of political character: it is prohibited from assuming the functions that the constitution allots to the legislative power, under penalty of violating the republican principle; and (d) the time limitation: it must rule by the maximum limit of 5 March of the electoral year" (Alvim, 2016, p. 70). In summary: "The Resolution can be understood as a secondary formal source, which is responsible for interpreting and regulating primary sources, such as the constitution and federal laws. It cannot innovate in the legal system or restrict rights or establish sanctions other than those provided for by law. [...] In this case, the law expressly authorizes the regulation of the matter by resolution. According to Article 57-J of Law 9.504/97, the TSE will regulate the provisions of Articles 57-A to 57-I of this Law (chapter on electoral publicity on the Internet) according to the circumstances and technological tools existing at each electoral moment and will promote, for vehicles, parties and other interested bodies, the formulation and widespread dissemination of rules of good practice relating to electoral campaigns on the Internet. However, as has been pointed out, it is important that the proposed rules speak directly to the rules in force in the legal system" (De Teffé, 2024).

in Article 19,[67] *caput*, of the Civil Internet Framework[68] and reflected in the electoral legislation itself in Article 57-F[69] of the Electoral Law;[70] (3) the duty to act to "prevent or diminish" is extremely generic, and it should be recalled that the principle of *lex certa is* a basic requirement of disciplinary law so that jurisdictions are guaranteed a normative predetermination that makes it clear what the offending conducts are; and (4) from a technical point of view, the rule is difficult to apply given the huge number of borderline cases and the unequivocal complexity of semantic analysis in cases involving, for example, "silence techniques" (Grijelmo, 2012) and

[67] "Article 19. In order to guarantee freedom of expression and avoid censorship, the Internet application provider may only be held civilly liable for damages and losses arising from content generated by third parties if, in the presence of a prior and express judicial order, it does not adopt measures to make the content indicated as in infringement unavailable within the scope and technical limits of its service and within the indicated timeframe, without prejudice to the legal provisions to the contrary."

[68] As a result of this new rule, the Internet application platform can only be held liable in the civil sphere "if, following a specific court order, it fails to adopt the necessary measures so that the content identified as harmful is no longer available within the scope and technical limits of its service and within the specified period." This rule resolves, at least in part, the divergence between court rulings that convict or do not convict Internet service providers for the content of offensive pages on their websites and social networks. The aim is to guarantee full freedom of expression in the use of the Internet, preventing any kind of censorship. In this way, it is clear that the Civil Framework did not intend to establish objective liability (risk theory) for providers for the actions of third parties, and any liability should be determined in the light of subjective liability (culpability theory)" (Teixeira, 2016, p. 110).

[69] "Article 57-F. The content and multimedia service provider that hosts the dissemination of the electoral publicity of a candidate, party or coalition will be liable to the sanctions provided for in this law if, within the period determined by the Electoral Justice, counted from the notification of the decision on the existence of irregular publicity, it does not take measures to cease such dissemination. Sole paragraph. The content or multimedia service provider will only be held responsible for the dissemination of publicity if it is proven that the publication of the material was with their prior knowledge."

[70] According to Article 57-F, platforms that host electoral publicity are subject to the sanctions provided for in the Electoral Law if, within the time limit set by the Electoral Court, they do not adopt measures to stop the dissemination of illegal content. However, the law establishes different rules for the actions of content providers (search engines, social networks, video platforms, blogs, etc.) and information providers (news portals in general). In the first case, liability for third-party content is only possible when prior knowledge is demonstrated; in the second, prior knowledge is evident, so that the sanction is independent of prior notification to remove the irregular material (Alvim, 2016, p. 326).

"countervailing truths" (MACDONALD, 2019), linguistic practices that make it extremely difficult to identify disinformation.

In any case, we understand that the imposition of a legal duty leads to the application, by analogy, of Article 13, §2, letters "a" and "c" of the Penal Code,[71] so that any negligence on the part of platforms in the face of illicit results can be viewed as a "relevant omission" for the purpose of its penalization (with the imposition of a fine) in the area of representations.

In this sense, the regulations establish an illustrative list of measures designed to signal compliance with the duty of diligence. According to the six sections of Article 9-D,[72] the regulatory objectives can be achieved, in principle, through: (a) adjustments to the wording and application of terms of use and content policies; (b) the implementation of effective notification instruments and reporting channels open to all users and to public or private institutions or bodies; (c) the planning and execution of corrective and preventive actions, including the improvement of content recommendation systems; (d) transparency in terms of the results achieved by the aforementioned improvement of recommendation algorithms; (e) the preparation of specific impact reports in electoral years, not only in terms of reducing

[71] "Article 13. The result, on which the existence of the crime depends, can only be attributed to the person who caused it. The action or omission without which the result would not have been produced is considered a cause. Paragraph 2. The omission is penally relevant when they who caused the omission should and could have acted to prevent the result. The duty to act is incumbent on those who: (a) have a legal obligation of care, protection or vigilance; (b) have otherwise assumed responsibility for avoiding the result; (c) by their previous behavior, created the risk that the result would occur."

[72] "Article 9-D. It is the duty of the Internet application provider which allows the dissemination of political-electoral content to adopt and publicize measures to prevent or reduce the circulation of notoriously false or seriously decontextualized facts that may affect the integrity of the electoral process, including: I – the development and application of conditions of use and content policies compatible with this objective; II – the implementation of notification tools and effective reporting channels, accessible to users and to public and private institutions and bodies; III – the planning and implementation of corrective and preventive actions, including the improvement of content recommendation systems; IV – the transparency of the results achieved by the actions mentioned in point III of the heading of this article; V – the preparation, in an electoral year, of an evaluation of the impact of its services on the integrity of the electoral process, with the aim of applying effective and proportionate measures to mitigate the risks identified, including with regard to political gender violence, and the application of the measures provided for in this article; VI – the improvement of its technological and operational capabilities, giving priority to tools and functionalities that contribute to achieving the objective set out in the preamble of this article."

disinformation related to elections but also in terms of mitigating the associated risks; and (f) the improvement of technological and operational capacities to prioritize tools and functionalities that contribute to the objectives of the Article.

According to point I, the adaptation of the conditions of use and community policies to reduce and prevent anti-system disinformation is an indication that the platform is acting to comply with the obligation stipulated in the heading. It should be noted, however, that the mention of "*observance*" leaves it clear that the regulation will not make do with hypothetical predictions, therefore compliance with the duty of diligence must also be analyzed from the perspective of *observance,* reflected in the analysis of the effectiveness of moderation processes. In other words, the adoption of appropriate policies is a necessary but insufficient condition; in addition, the consequences provided for in the moderation plan (labeling, marking, reducing visibility, eliminating content, suspending accounts, interrupting monetization, or prohibition) must actually be applied so that a broad deterrent effect is fostered within virtual communities, one which favors information integrity. This addresses the need to avoid the practice of ethics washing (Muñoz Vela, 2022, p. 61), whereby some companies adopt positive principles as a public relations measure, without intending to apply them rigidly.

In relation to the previous point, point II stipulates that opening up channels for comments is an additional parameter of duty of care, important insofar as it introduces the prospect of purifying public debate at the request of the public authorities or social organizations, thus seeking a form of "inclusive governance" (Coeckelbergh, 2022, p. 157). The requirement is justified by the fact that the "cure" in a strict, internal, and isolated context, even if it is proactive, tends to be less complete than moderation in a context driven by multilateral, multitudinous, and participatory control. In any event, the rule requires "effective instruments," so complaints must have significant effects. This means that the opening of complaint channels alone does not fulfill the order: Complaints need to be processed quickly and efficiently and culminate in a timely resolution so that the moderation cycle can take place properly.

Point III imposes a certain degree of professionalism, coupled with the need for the fight against false narratives to receive, in addition to a programmatic outline, a model of anticipatory action which serves for preventive purposes. *Preventive measures* can take the form, for example, of broad awareness or immunization campaigns as well as the hiring of professional fact-checking agencies to help with moderation. The use of AI tools for the automated

analysis of large amounts of information is also contemplated. In parallel, adapting the platform's rules of use to the national context could be seen as a precautionary measure. Adjustments can also be made to the product, for example, by adopting preventive measures that make the viral spread of disinformation or harmful content more difficult. On the other hand, *corrective measures* can be applied; for example, revising registration protocols to make it compulsory to enter personal data that creates effective obstacles to anonymity. Adjustments could even be made to community policies and conditions of use, such as preventing the monetization of disinformation channels, expressly prohibiting disinformation related to elections in the case of platforms that do not already prohibit this, or increasing the sanctions for this type of practice in the case of platforms that already prohibit false narratives related to elections in some way. Furthermore, the platforms themselves could fund, promote, or facilitate research in collaboration with universities or research laboratories to test the effectiveness of their internal governance, highlight its weak points, and help to revise protocols in order to correct the latter.

In point IV, the imperative of transparency emerges as an instrument of supervision and control so that the adoption of corrective and preventive measures does not depend exclusively on the self-declarations of *big tech companies*. In this context, the measures adopted will be judged in the light of their results, according to quantitative and qualitative analyses that should show a positive transformation in the willingness of each platform to make its environment less favorable and exposed to the circulation of disinformation. In principle, and by way of illustration, from a *quantitative* point of view it would be possible to report, for example, the number of publications banned, accounts suspended, content tagged or marked, and channels demonetized as well as the number of irregular paid advertisements detected and blocked before their publication; from a *qualitative* point of view, the same information could be provided as from a quantitative point of view: the number of prohibited publications, suspended accounts, tagged or marked content, demonetized channels as well as the number of irregular paid advertisements detected and blocked before their publication and the nature, frequency, and scope of awareness campaigns. In addition to this analysis, information could be included on any agreements signed with public institutions (the judiciary, the prosecutor's office, etc.) or associations whose aim is to disseminate appropriate information on the functioning of elections, adjustments to regulations or moderation practices, the development of new products such as the insertion of labels or tags, adjustments to hinder the viral spread of illegal content, or to eliminate anonymous, false, or automated accounts, etc. In short, impact assessment is a governance mechanism

capable of showing the results of an action that is expected to be positive. In this context, its objective is to make the benefits obtained thanks to the measures applied "demonstrable" and "scalable."

The regulation demonstrates that Brazilian electoral justice expects improvements not only in what specifically concerns the mitigation of disinformation and the neutralization of its effects but also in regard to harmful content or abuses of language generally associated with the dissemination of fake news with a hostile or pejorative DNA. Therefore, these actions can – and should – encompass other phenomena, such as gender-based political violence, hate speech,[73] cyber bullying, and narratives that incite extremism and radicalization.

In the light of point IV, technological and operational innovations are encouraged, which were previously contemplated as parameters that demonstrate compliance with the duty of care. Other hypothetical examples of movement in this direction would be the development and implementation of new technology to detect coordinated disinformation campaigns or robot accounts, measures that enable message tracing and the preservation of temporary or self-deleting messages as well as the prohibition of hidden forms of advertising and the creation of reinforced layers in the protection of personal data. Furthermore, compliance with the duty of care does not have a defined shape, and may take the form of an open-ended list of measures not mentioned in the Resolution which, without claiming to be definitive, might involve, among others: (1) the promotion of media literacy initiatives as well as initiatives relating to information literacy, algorithmic literacy, and an introduction to AI; (2) the adoption of internal governance measures that foster content moderation in the face of disinformation and harmful messages; (3) the implementation of independent review and audit processes to strengthen *compliance* and ensure the effectiveness of internal procedures, particularly with regard to their prior and ongoing *implementation;* (4) the standardization of basic procedures for reviewing complaints, notes, and related procedures, preferably to reduce response and action times; and (5) the introduction of elements of transparency, for example, structuring libraries, enabling APIs, and providing information and explanations about

[73] Óscar Sánchez (Sánchez Muñoz, 2020, p. 33) takes care to signal the connections between disinformation and incitement to hatred: "These are two distinct phenomena, but there is common ground between them, because, as has been shown in recent years, the calculated amplification of hate speech has become one of the favorite strategies of disinformation campaigns, for which these campaigns rely on the existing social dynamics in societies, stirring up internal divisions and conflicts."

internal flows and algorithmic programming (MUÑOZ VELA, 2022, p. 95–96), with additions and adaptations.

Going even further, §1 of Article 9-D[74] expressly prohibits the promotion and paid prioritization of disinformation that threatens the integrity of the electoral process in the results obtained from user searches, both on search engines (Google, Bing, or others) and on other social-media platforms (for example, searches on video-hosting platforms such as YouTube, or social networks such as Facebook, X, or Instagram). Read in conjunction with §2,[75] it is clear that the obligation of §1 has an objective nature, strictly linked to the risks of the business which are undeniably high if one considers their possible impact on the electoral landscape.[76]

[74] "Section 1. Providers of applications that commercialize any form of content enhancement, including the prioritization of search results, are prohibited from making this service available for the dissemination of notoriously false or seriously decontextualized facts that could affect the integrity of the electoral process."

[75] "Section 2. The provider of the application that detects illegal content of the nature referred to in the heading of this article or is notified of its circulation by users shall adopt immediate and effective measures to stop the boosting of, monetization of and access to said content and will promote the internal investigation of the event and of the profiles and accounts involved in order to prevent the further circulation of said content and inhibit illegal conduct, including making the boosting service or monetization unavailable."

[76] "Social networks [...] play an increasingly influential role in the electoral process. Their ability to connect voters, candidates and news organizations in real time has fundamentally changed the way campaigns are carried out and how voters receive information. Twitter (now X), for example, has become a key forum for political discourse. Politicians, analysts and ordinary citizens use the platform to share their opinions, debate issues and react to events in real time. It offers candidates a direct and immediate way to communicate with voters, but it has also been used to spread false information and inflame divisive rhetoric. Facebook, with its huge user base and sophisticated advertising targeting capabilities, offers a powerful platform for spreading campaign messages. It allows candidates to reach voters in a very personalized way, but has been criticized for its lack of transparency and its potential for abuse. Instagram, especially popular with young people, is often used to humanize candidates and attract voters in a more visual and emotional way. However, it has also been used to spread disinformation through memes and other shareable content. TikTok, although more recent, has already demonstrated significant potential for influencing electoral discourse. Its short, attractive videos offer a unique way of attracting voters, especially the younger generation, but these also raise concerns about the spread of disinformation and manipulative content. In short, social networks have the power to significantly influence the electoral process, both positively and negatively. They offer new avenues for candidates to connect with voters and for voters to get involved in politics, but they also present new risks and challenges that require appropriate attention and regulation. It is crucial that voters, candidates and regulators understand this dynamic and work to ensure that social media is used in a way that supports, rather than undermines, the democratic process" (Lins, 2023, pp. 292–294).

However, the regulation is not limited to prohibiting the deliberate commercialization of services related to artificially increasing the visibility of false narratives targeting the electoral process. On the contrary, non-compliance will occur whenever the platforms, even due to carelessness or negligence, allow the privileged exposure of false content in exchange for remuneration. From the combination of both regulations, it can be deduced that platforms are obliged to: (1) not commercialize the boosting or prioritization of disinformation content that threatens electoral integrity (§1); (2) not allow the evasion of the aforementioned prohibition (§1); (3) on their own initiative, ensure the immediate removal of content that has managed to evade the impediments (§2, first part); and (4) in the event of a failure, take measures (promote changes) so that similar events are not repeated (§2, final part), perfecting automated-detection algorithms and strengthening the human review process as appropriate.

Furthermore, §2 broadens the scope of prohibitions on monetization, creating a rule that financially de-incentivizes disinformation activities. Thus, it is the duty of the platforms not only to block the promotion, or prioritization, of paid disinformation content but also to prevent the remuneration of monetized profiles or channels dedicated to this type of antisocial activity. According to the text, the end of such payments is imposed *ex vi legis,* since platforms are obliged to take such measures proactively, regardless of whether there is a social or official complaint. In this sense, notification by users only indicates inertia, which verifies – but does not "inaugurate" – the context of violation of the imposed rule. In direct terms, platforms have a duty of care in terms of social impact, and this prevents them from financing participants in the disinformation industry.

Non-compliance entails the obligation to restore the information ecosystem by promoting, free of charge, messages that refute fake news or clarify it (for example, by adding information that was maliciously omitted or other contextual elements).[77] By way of illustration, reparation – derived from a kind of institutional right of reply enshrined in police power – can occur through clarifying notes, fact-checking articles, specialized studies, official documents, or public reports of any type, at the discretion of the judicial authorities.

Unfortunately, on this specific point the Resolution did not take into account that disinformation, due to its emotional component, has a natural

[77] "Section 3. The electoral court can order the application provider to disseminate, via boosting and free of charge, informative content which explains notoriously false or seriously decontextualized content previously boosted in an irregular way, in the same form and with the same reach as in the contract."

tendency toward viral dispersion (Gillespie, 2020, p. 324), which is why as a general rule, its real reach far exceeds the limits of boosting, multiplied by innumerable multiplatform social distribution (forwarding) reactions. In this sense, the opportunity has been missed to guarantee clarifying messages that warn against disinformation with a visibility closer to that which disinformation itself actually obtains, making an analogy with Article 58, IV, b, of the Electoral Law.[78] In this vein, it is worth noting that, aware of this asymmetry, the ordinary legislator provides, in the aforementioned Article, that in cases of offense or the dissemination of untrue facts on the Internet, the reparative response will be available to users for no less than twice the amount of time that the offensive message was available for. In conclusion, the rule ends up giving a digital phenomenon a parallel treatment to analog injustices, when it would be preferable to adopt the logic of abuse of communication freedoms in the virtual arena. However, there are no obstacles to making this correction in the field of judicial interpretation.

3.2.4 Types of Behavior Targeted by Regulation

The Resolution transfers the logic of asymmetric regulation to the field of elections, based on the magnitude of the risks; as we have already seen, this is very typical of the approach toward regulating (or intending to regulate) AI around the world. In this sense, application providers will be "jointly liable, civilly and administratively, when they do not promote the immediate unavailability of content and accounts, during the electoral period, in [...] cases of risk" listed in the five sections of Article 9-E, which specifically enumerate: (1) antidemocratic conduct, information, and acts that typify the crimes provided for in Articles 296, sole paragraph (falsification of a public seal or sign),[79]

[78] "Article 58. Once the candidates have been chosen at a convention, the right of reply is guaranteed to any candidate, party or coalition affected, even if indirectly, by an insulting, defamatory, libelous or knowingly false concept, image or statement disseminated by any means of communication. [...] IV – in electoral publicity on the Internet: [...] b) the reply will be available for access by users of the Internet service for no less than twice the time that the message considered offensive was available for."

[79] "[Falsification of a seal or public sign] Article 296. Falsifying, producing or altering these: [...] §1 – the same penalties will be applied to: I – whomever makes use of the falsified seal or sign; II – whomever makes improper use of the true seal or sign causing harm to others or for self-benefit or the benefit of third parties; III – whomever alters, falsifies or makes improper use of hallmarks, logotypes, initials or any other symbols used by bodies of the state administration or which identify the latter."

359-L (violent overthrow of the democratic rule of law),[80] 359-M (coup d'état),[81] 359-N (interruption of the electoral process),[82] 359-P (political violence),[83] and 359-R (sabotage)[84] of the Brazilian Penal Code; (2) the dissemination or spreading of notoriously false or seriously decontextualized information that affects the integrity of the electoral process, including voting procedures and techniques and the counting and tallying of votes; (3) the serious, direct, and immediate threat of violence or incitement to violence against the physical integrity of officers and employees of the electoral justice system and the Public Electoral Ministry, or against the physical infrastructure of the judiciary, with the aim of restricting or impeding the exercise of constitutional powers or the violent abolition of the democratic rule of law; (4) the incitement to hatred, including the promotion of racism, homophobia, fascist ideologies, or hatred against people or groups, in any form of discrimination; and (5) the dissemination or exchange of content produced or manipulated, in part or in whole, by digital technology, including AI, outside the required *disclaimer*.

In general terms, the Resolution imposes on platforms the duty to summarily remove any content that fits the theory that describes a particular set of criminal behaviors linked to the protection of the democratic regime, among which are specifically included: (1) the *falsification of a public seal or sign*, which includes the improper use of brands, logos, acronyms, or any other identifying symbols of state bodies or institutions (which, in the field of disinformation, is generally produced with the creation of false accounts purporting to represent bodies that form part of electoral justice); (2) *the attempt to abolish the rule of law,* with the use of violence or serious threats that prevent or restrict the functioning of constitutional powers; (3) the

[80] "[Violent abolition of the democratic rule of law] Article 359-L. To attempt, with the use of violence or serious threat, the abolition of the democratic rule of law by preventing or restricting the exercise of constitutional powers."

[81] "[Coup d'état] Article 359-M. To attempt to depose, by means of violence or serious threat, the legitimately constituted government."

[82] "[Interruption of the electoral process] Article 359-N. Impeding or disrupting the election or the verification of its results, improperly violating the security mechanisms of the electronic voting system established by the Electoral Justice."

[83] "[Political violence] Article 359-P. Restricting, preventing or hindering, through the use of physical, sexual or psychological violence, the exercise of political rights to any person on the grounds of their sex, race, color, ethnicity, religion or national origin."

[84] "[Sabotage] Article 359-R. Destroying or rendering useless means of communication to the public, establishments, installations or services destined for national defense, with the aim of abolishing the democratic rule of law."

attempt to depose the elected government, through violence or serious threats; (4) *preventing or disturbing elections or the counting of votes,* by improperly violating the security mechanisms of the electronic voting system; (5) *restricting, preventing, or disturbing the exercise of political rights, with the use of* physical, sexual, or psychological violence, on grounds of sex, race, color, ethnicity, religion, or national origin; and (6) *destroying or rendering useless means of communication, installations, or services intended for national defense, with the* aim of abolishing the democratic rule of law.

Despite the laudable intention of safeguarding the system of public freedoms inherent to the constitutional state, the regulation possesses a very worrying aspect, insofar as it delegates to the platforms the duty to carry out an extremely technical scrutiny that is undeniably characteristic of the functions of the judiciary.[85] Moreover, it moves in a direction which is in principle contrary to the regime established by Article 19 of the Civil Internet Framework,[86] which presupposes a model of responsibility that is limited to

[85] It should be recalled that current legislation establishes that big tech companies, as a general rule, are not responsible for the content shared by users on social networks (Article 19 of the Brazilian Civil Framework of the Internet). According to this blueprint, the decision to delete content in the event of a breach of the terms of use or community rules is merely an option for the platforms (Lins, 2023, p. 299). However, Bill No. 2.630/2020 (the Fake News Bill) aims to approve a similar regime to that established by the TSE in a recent resolution. If the bill is approved, "social networks would be held responsible for content that falls under some of the crimes defined in Brazilian legislation, such as acts of terrorism, instigation to suicide or self-mutilation, crimes against the democratic rule of law, crimes against children and adolescents, violence against women, among others" (Lins, 2023, ibid).

[86] "The provision seems to be an exception to the rule of the Civil Framework of the Internet (Law No. 12.965/14) which, in its Article 19, establishes that, in order to guarantee freedom of expression and avoid censorship, the Internet application provider can only be held civilly liable for damages and losses arising from content generated by third parties if, after a specific court order, it does not adopt measures to ensure that the content indicated as in infringement is made unavailable within the scope and technical limits of its service and within the period indicated, without prejudice to the legal provisions to the contrary. In view of this, there are serious doubts as to whether a TSE Resolution could be the appropriate instrument for this, given its nature, and whether it would not be innovative and go beyond its possibilities" (De Teffé, 2024). It should be noted, however, that the constitutionality of Article 19 is being challenged in the context of Extraordinary Appeal No. 1.037.396 before the Brazilian Supreme Court.

resisting compliance with a court order.[87] Furthermore, it tends to be highly ineffective, given that most of the crimes mentioned include violence, serious threats, or violation of the security of the voting system as objective elements of the types of crimes, which makes them criminal *acts* and not *crimes of mere expression,* which can be committed by simply stating them. Strictly speaking, according to the literal wording of the text, the rule obliges platforms to remove content that corresponds to the offenses listed, but what is clear is that four of the six offenses could hardly be committed through manifestations on social networks (the exceptions would be the improper use of symbols and the disruption of the exercise of political power through psychological violence, potentially observable in hate speech and online hate campaigns). From the point of view of guarantee, in line with the principle of legal certainty, the requirement as it is worded[88] does not cover publications which, although they impact democratic institutions, lack any of the elements necessary to constitute one of the criminal offenses described.

Rather, the duty to withdraw should be directed at justifying or inciting these practices, and not at the practices themselves – and this could be the case if the interpretation leans toward the expression "cases of risk" contained in the heading, which would equate public incitement with the commission of the crimes themselves. In this way, the regulation would give effective

[87] It is worth highlighting, however, the important point made by Ana Frazão (2024): "[G]iven the current state of development of AI, past experiences and the risks that have already been identified, the TSE's initiative is essential to guarantee the legitimacy of the Brazilian electoral process. This is why even the argument of a supposed violation of Article 19 of the Civil Framework should be viewed with caution. Apart from the fact that this is a legal provision whose constitutionality is being discussed before the STF and the fact that it needs to be interpreted in accordance with electoral legislation, it is essential to understand that [...] Article 19 [...] is clearly aimed at third-party content, in relation to which platforms assume a position of absolute neutrality. Content over which platforms have an influence, often through management and promotion, obviously cannot be considered as mere third-party content, which is why it should not be subject to the restricted theory of liability provided for in Article 19. Consequently, there are excellent reasons to justify the compatibility of the TSE Resolutions with the Brazilian constitution and laws. However, the main reason that justifies them is the simple fact that Brazilian democracy cannot wait."

[88] "Article 9-E. Application providers will be jointly and severally liable, civilly and administratively, if they do not make content and accounts immediately unavailable during the electoral period, in the following cases of risk: I – of anti-democratic conduct, information and acts that characterize violations of Articles 296, sole paragraph; 359-L, 359-M, 359-N, 359-P and 359-R of the Penal Code."

coverage to pro-coup and insurrectionary publications, which have been cause for concern ever since "conspiratorial messages" began to "crush social consensus," generating an absurd "polyphony of discourses" which "undermined confidence in public institutions, the pillar of democratic societies" (Carratalá *et al.*, 2023, pp. 13–14). Nevertheless, in these circumstances a strict and non-expansive interpretation, compatible with the dogma of restrictive rules of law, indicates that platforms are not technically obliged to delete posts that do not fit the types enumerated, which exempts them from obligations with regard to posts especially related to crimes whose violence or serious threat determines the commission of a crime.

As for point III, the law is right to include the perspective of incitement to violence, broadening the scope beyond cases of violence per se, which are null and void. However, the wording significantly reduces the provision's scope by including the need for a subjective element linked to the intention to "restrict or impede the exercise of constitutional powers or violently overthrow the democratic rule of law." Ideally, fomenting violence against the organs of the judiciary should be prohibited, regardless of intent, to include calls for vandalism, which are more linked to a general feeling of unrest than to an articulated action aimed at deposing the elected government. According to the literal wording of the regulation, incitement to attack the headquarters of electoral offices would be difficult to include in this provision, so the regulatory mandate, in many circumstances, could become ineffective.

On the other hand, the crusade against hatred and prejudice in the flow of electoral conversations is urgent and beneficial and reflects the social purpose of "protecting certain collectives, fighting intolerance and discrimination," currently enshrined in international law (Alcácer Guirao, 2023, pp. 31–32).[89] However, point IV of Article 9 *sexies* raises important concerns in terms of legal certainty and the always complex limitation of freedom of expression, precisely at the point where it inserts, alongside legally determinable notions such as racism and homophobia, vague concepts that have historically been the subject of important

[89] For example, Recommendation No. 15 of the European Commission against Racism and Intolerance, published on 21 March 2016, promotes the fight against hate speech, which in this document is characterized as "the promotion or instigation, in any of its forms, of hatred, humiliation or disparagement of a person or group of people, as well as harassment, discrediting, the dissemination of negative stereotypes, the stigmatization or threatening of a particular person or group of people and the justification of these manifestations on the grounds of race, color, ancestry, national or ethnic origin, age, disability, language, religion or beliefs, sex, gender, gender identity, sexual orientation and other personal characteristics or conditions" (Alcácer Guirao, 2023, p. 32).

semantic disputes, such as Nazism and especially, fascism. In particular, the excess of abstraction forces subjective interpretations which, in turn, can give rise to cases of injustice, both if the lack of equality before the law arises voluntarily and if not. In addition, semantic indeterminacy would make automated action practically impossible, which is undoubtedly essential for large-scale defense actions.

All things considered, defining Nazi and fascist behavior is extremely complicated, even more so at a time when these terms are employed in an arbitrary and random way as rhetorical weapons designed to "have a mental significance" (Castells, 2009, p. 535). These aim to accumulate power and claim a monopoly on dignity in a battle of narratives guided by "processes of identification and moral confrontation" (Elorza Saravia, 2023, p. 58) in which the "emotive meaning of notions" makes certain words "part of the war" (Del Rey Morató, 2007, pp. 130–131).

In this sense, it should be recalled that "the realm of meanings is very complex," given that "meanings are varied, divergent and changing" and can "be imprecise [and] open to interpretation." In pluralist societies, in particular, "not only are they different, but they co-exist in tension, competition and conflict," which makes it impossible to aspire to a "meaningful totality", an unambiguous meaning for society as a whole. Ultimately, "every meaning is precarious and problematic" (Martínez García, 2020, p. 30), especially in fragmented and polarized environments, and can be questioned argumentatively, given that this type of stigmatizing categorization will always be the subject of discourses that mobilize opposing values, emotions, references, and theoretical and ideological currents (Del Amo Castro, 2023, pp. 63–64). It should therefore be noted that:

> The language games of magical engineering allow us to act on notions, broadening or narrowing their meaning. The field of a notion is expanded or narrowed so that it encompasses or does not encompass certain beings, certain things, certain ideas, certain behaviors or certain situations [...] There are notions that are more malleable than others, which allow for narrowing or broadening the field they encompass, lending themselves to confusion.
>
> (Del Rey Morató, 2007, pp. 129–130)

Despite these considerations, what is clear is that the approved regulation, as worded,[90] has the central objective of preventing incitement to hatred.

[90] "Article 9-E. Application providers will be jointly and severally liable, civilly and administratively, when they do not promote the immediate unavailability of content and accounts, during the electoral period, in the following cases of risk: [...] IV – of behavior

In this context, the associated phenomena – "racism," "homophobia," "Nazi, fascist or hate ideologies," "prejudices" – are mentioned from a clearly instrumental perspective. That is to say, these do not need to be monitored, analyzed, and excluded per se, but only when they are employed as *means* (instruments) for achieving a *goal*: the practice of hate speech, understood in a broad sense as any form of hostile or aggressive manifestation against minorities (Alcácer Guirao, 2023, p. 27), in line with the "herd majoritarianism" that pervades social networks (Fisher, 2022, p. 102). In any event, the semantic vagueness suggests a need to develop the concept further.

Hate speech is considered to be any linguistic practice aimed at spreading derogatory statements or opinions about vulnerable or minority peoples or groups (Alvim *et al.*, 2023, p. 182), precisely because of their vulnerable or minority status. In the context of elections, hate speech implies a "desire for exclusion," followed by an action aimed at causing personal distress, which induces, in addition to a feeling of not belonging, "the production (through incitement to or collusion with violence) of damage to opponents," in a movement that interferes with the normalcy and regularity of the election (Guaraty, 2023, p. 130). In other words:

> Hate speech is a communicative action that, based on social identities – often belonging to vulnerable minorities – seeks to attack people's equality in dignity and rights, mobilizing various manifestations of language to do so. When we talk about *hate speech*, we might think that this is merely about words, but that is not the case. Hate speech transcends verbal communication, whether written or spoken. It makes use of visual motifs, symbols, graphics, images and also gestures or body positions that become part of the composition, coding and transmission of the message, ways of using our language and our ideas, which are not verbal. All forms of communication can be used as a means of expressing hate speech.
>
> (García Fajardo, 2022, pp. 243–244)

Hate speech, as has been demonstrated, can be carried out in different ways, but in any case, the Resolution elects to highlight some of the supposedly more obvious or common ones. In the first place, the notion of *prejudice* denotes the idea of "beliefs, opinions or value judgments that lack adequate

---

that incites hatred or hate speech, including the promotion of racism, homophobia, Nazi ideologies, fascist ideologies or hatred against a person or group due to prejudices of origin, race, sex, color, age, religion and any other form of discrimination."

justification and tend to predispose people against certain groups," which "cause or perpetuate unjust social attitudes, situations or structures" (Stepanenko, 2022, p. 652). *Racism*, in parallel, implies "practices and discourses that distinguish, classify, discriminate against and treat human beings differently, based on the idea that there are biological or bodily differences between them associated with their origin, which serve to justify systems of political domination and economic or scientific exploitation" (Navarrete, 2022, p. 668).[91] On the other hand, *homophobia*[92] takes shape when "fear or rejection of homosexuals manifests itself through some form of violence" (Bustamante Tejada, 2022, p. 391).[93] Finally, *discrimination* emerges as the *leitmotif* of all the forms of linguistic abuse provided for in the Article. In conceptual terms, it can be understood as "a differentiated treatment that implies an adverse distinction derived from negative assessments made about particular people or groups," generating situations that are "in direct contradiction to the idea of equality and justice of constitutional democracies" (Luna Corvera, 2022, p. 238).

[91] "Racism consists of a behavior of hatred, rejection and contempt towards people who supposedly have different physical characteristics (skin color, hair type, etc.) and who are believed to form a distinct and evidently inferior race. Racism seeks to scientifically theorize the existence of these races that are considered to be clearly unequal. These races are conceived of as hierarchized animal species, whose features induce particular moral and cultural characteristics that control the actions of each individual included within each race" (Birnbaum, 2014, p. 254).

[92] "Disinformation linked to LGTBI people has experienced a notable increase all over the world in recent years, in countries such as the US, Poland, South Korea and Malaysia, and closely linked to public debate at election times, as a result of media treatment [...]. The circulation of these disinformation messages, which could well be considered hate speech insofar as they seek to cause harm to a vulnerable group, is greater in proportion to the extent they resonate with dominant moral values in a given sociocultural context and often serves to legitimize certain forms of government and the exercise of social control through the creation of 'sexual panics'" (Carratalá; Peris-Blanes, 2023, pp. 240–241).

[93] Despite not being explicitly mentioned, it is clear that transphobic content is covered by the prohibition laid down in the resolution. Having said that, it should be pointed out that "for trans and intersexual people, the experience of hatred and mistreatment is especially virulent. They are even more exposed to extreme discrimination and brutal violence than gays and lesbians. This is due, among other things, to the lack of public spaces where they can socialize and feel protected. In swimming pools, gymnasiums or public gyms, they run the risk of being excluded or attacked" (Emcke, 2021, pp. 134–135).

The regulation adds that the proactive supervision of activity on the platforms must pursue content that spreads hate speech involving "Nazi, fascist or hate ideologies against a person or group based on prejudices of origin, race,[94] sex, color, age, religion and any other form of discrimination," such as sexual orientation, gender identity, ethnic origin, disability, etc., to encompass all types of phobic discourse. However, when it comes to worldviews, it has already been made clear that it is not a question of defining what constitutes – or does not – a discourse that contains traces of destructive ideologies but of knowing whether, around any Nazi-fascist idea that is made public, there is an embedded discourse of hatred. Based on this premise, the reduction of complexity is clearly immense, which greatly facilitates the task of judicial agents.

Even so, there is no need to explain that the historical intersections between *fascism* and hate speech are less obvious than the intrinsic link between *Nazism*, racism, and discriminatory persecution (Borja, 1997, p. 704), especially since "the emphasis placed on racism and antisemitism" is precisely the "main difference" between the two worldviews (Costa, 2020, p. 293). However, in some instances, fascism, in its connections with extreme nationalism (Saccomani, 2009, p. 466), emanates from xenophobic impulses (Pérez-Curiel *et al.*, 2023, p. 29) and spreads religious intolerance. Of course in its postmodern version, it is recognized that fascism, with its tendency to single out enemies, tends to disseminate chains of digital attacks not only against foreigners and Islamists (or against adherents of religions of African origin, in the case of Brazil) but also against homosexuals (Carratalá; Peris-Blanes, 2023, p. 24), proponents of Islam (Carratalá; Peris-Blanes, 2023, p. 241), proponents of gender identity (Montagut

[94] "According Oracy Nogueira, racial prejudice is 'an unfavorable disposition (or attitude), culturally conditioned, towards members of a population who are perceived as stigmatized, whether because of their appearance or because of all or part of the ethnic ancestry that is attributed to them or recognized'. Based on this concept, he designates two types of prejudice: mark prejudice and prejudice of origin, using the examples of social relations in the US and Brazil to describe both types. If a prejudice is based on a person's physical features, gestures or accent, this is a mark prejudice, which varies subjectively depending on the characteristics of the person observing and the person identified. When the assumption that a person descends from a particular ethnic-racial group is enough to make them a victim of a prejudice, this is a prejudice of origin. In racial relations defined by mark prejudice, the criterion for defining the discriminating and discriminated against groups is racial appearance and in racial relations defined by prejudice of origin, regardless of the appearance or proportion of ancestry of the discriminating or discriminated against group, if there is interracial mixing there will be prejudice" (Braga, 2022).

et al., 2023, p. 256), and feminist activists (MURGIA, 2019, p. 49), all of which reflect xenophobic and Islamophobic agendas (EMCKE, 2021, p. 55; MULHALL, 2021, p. 25) or transphobic and misogynist agendas typical of alt-right movements, particularly in the field of the extreme radical right (MASSANARI, 2020, p. 179).[95]

3.2.5 Effects on Freedom of Expression

Having said this, the limiting effect that this regulation can have on the climate in which freedom of expression develops cannot be overlooked, given that, from the point of view of big business, laws of this scope encourage the elimination of all sorts of controversial content, including legitimate content, due to the fear of being sanctioned (LINS, 2023, p. 300). Of course, this can have repercussions both in the area of freedom of expression of opinions and ideas and hypothetically, in the equality of conditions between candidates and political parties, since poorly executed content moderation can generate imbalances. However, as Henry Kissinger and his co-authors ask: Within a constitutional society, is there a "right to read [produce or disseminate], or even a legitimate interest in reading, 'false' information generated by AI?" (KISSINGER *et al.*, 2021, p. 106).

Moreover, it is worth asking whether the requirement to eliminate artificially generated content that lacks a *disclaimer* about its synthetic origin is feasible from a technical point of view. This is due to the fact that the enormous flow of daily publications makes automated analysis essential, so that compliance with the standard would require a combination of semantic analysis tools (operating, as we have seen, on excessively vague terms) and image processing or visual computing software, possibly combined with the subsequent intervention of human reviewers to minimize the potential of the machine, for example, when terms related to false narratives are the subject of genuinely informative journalistic articles (which, for example, amplify the spread of condemnatory decisions) or content that reproduces these terms with the sole purpose of denying or criticizing them publicly.

[95] "Hierarchy, which can be related to race, ethnicity, sexual condition or political position, is always present in extreme right-wing ideologies, and radicalization, when taken to the extreme, produces an environment of persecution against those who, within their beliefs, are considered inferior human beings who should not enjoy the same rights and protections of individual freedoms. In general, the rhetoric of the extreme right makes explicit the tendency to dehumanize the groups it considers inferior" (PRADO, 2023, p. 80).

Finally, it is worth highlighting the efforts toward harmonization in this area. The Brazilian electoral justice system has more than 2 600 magistrates serving in the electoral zones. Considering the need for uniform action so that equivalent cases are treated equally, the collegiate decisions of the TSE have been established as binding in proceedings in lower courts, both in administrative proceedings based on police power and in the resolution of claims (the right of reply or irregular publicity).[96]

In this context, the Resolution stipulates that equivalent treatment covers not only cases of identical content but also those in which there is a state of "substantial similarity" between the original content and the approximate replica.[97] In fact, the rule addresses the fact that the concept of "sameness" is difficult to determine in practice. After all, in an almost infinite number of publications on the same subject, what can be considered, in a strict analysis, objectively identical? Obviously, in an independent interpretation, the expression "identical" clearly differs from what is "analogous," "corresponding," or simply "similar." However, taking into account the complexity and high degree of variation between cases, the TSE elected to equate "identical" with "substantially similar" to make the rule more operative.

Therefore, it is possible to deduce that the regulation should apply not only to 100% identical content (which implies the absolute absence of any difference between the original piece and the derived piece) but also – and as a minimum – to cases of: (1) partial reproduction, especially through the exhibition of isolated clippings or "high impact" fragments; (2) repetition via a different medium (broadcast of the audio only from a video considered to

[96] TSE Resolution No. 23.610/2019 – "Article 9-F. In the event that the electoral publicity on the Internet transmits notoriously false or seriously decontextualized information about the electronic voting system, the electoral process or the Electoral Justice, the judges mentioned in Article 8 of this Resolution will be adhere, in the exercise of police power and in representations, to the decisions of the plenary of the Superior Electoral Court on the same subject, in which the removal or maintenance of identical content has been ordered." Available at: [https://www.tse.jus.br/legislacao/compilada/res/2019/resolucao-no-23-610-de-18-de-dezembro-de-2019]. Viewed: 4.5.2025.

[97] TSE Resolution No. 23.610/2019 – Article 9-F. "Paragraph 1. The provisions of the main body of this article shall apply to cases in which, despite editing, restructuring, alteration of words or other artifices, methods or techniques to elude automatic duplicate content detection systems or to hinder human verification, there is substantial similarity between the content removed by order of the Superior Electoral Court and that disseminated in regional or municipal advertising." Available at: [https://www.tse.jus.br/legislacao/compilada/res/2019/resolucao-no-23-610-de-18-de-dezembro-de-2019]. Viewed: 4.5.2024.

be disinformation); (3) simple editing (acceleration of word speed, insertion of subtitles, etc.); (4) complex editing (taking advantage of key moments to create new content with the same symbolic basis, for example, with the addition of sound elements to make it more dramatic, entertaining); and (5) overlay (for example, in *reactive* content, where the glossed content is reproduced in the *background* while it is analyzed or reinforced by the *assessments* made by a commentator).

The issue, however, has other layers, and the supposition of *spin-offs* is particularly complex, for example, with the use of partial transcriptions such as hashtags, cards, photo footers, or memes that reproduce, in text, a central idea present in a piece considered to be disinformation. As an example: once the High Court has considered the official material of party X claiming that candidate Y is a murderer to be disinformation, can future publications with the same accusation be considered to be substantially similar in all cases? In addition, the complexity increases when we recall the importance of context, especially since in many cases the reproduction may be linked to a journalistic or satirical activity, etc., or even from a perspective that is critical of the original publication (reproduction for the purposes of condemnation, contradiction, raising awareness, etc.). Therefore, the comparison of videos or other pieces must take into account not only their sameness or similarity of subject matter but also their intentionality and contextual framework.

In short, the regulatory edict establishes the creation of a repository of collegiate decisions[98] created to facilitate the work of the zonal authorities as well as defining that orders to withdraw content based on the existence of a collective High Court ruling can indicate a deadline for compliance of less than 24 hours, depending on the gravity of the content in question and the context of the elections underway (Article 9-F, §§2–3). However, to prevent abuse and guarantee the preferential position that freedom of expression occupies in the constitutional framework, it creates a mechanism to promote judicial self-restraint, mitigating excesses in the exercise of police power

[98] TSE Resolution No. 23.610/2019 – "Article 9-G. The decisions of the Superior Electoral Court ordering the removal of content that transmits notoriously false or seriously decontextualized information that affects the integrity of the electoral process will be included in a repository made available for public consultation. § Paragraph 1. The repository will contain the case number and the full text of the decision, in which the electronic address where the content to be removed is hosted will be highlighted, along with a description of its essential elements for inclusion in a field provided by the Judicial Secretariat." Available at: [https://www.tse.jus.br/legislacao/compilada/res/2019/resolucao-no-23-610-de-18-de-dezembro-de-2019]. Viewed: 4.5.2024.

(Article 9-F, §4[99]). This rule opens up the possibility that illegal or disproportionate (abusive) decisions can be challenged through complaints, which can lead, cumulatively or alternatively, to the nullification or revocation of the order and referral to the respective department of internal affairs for the initiation of an investigation or disciplinary administrative procedure.

For decisions to remove disinformation to be taken objectively and with certainty, the judicial authorities must exercise moderation and refrain from ordering the removal of messages in the absence of certainty as to their veracity. It is also recommended that they act according to technical criteria, observing what the doctrine describes as the "legitimating conditions for the judicial restriction of content." On the one hand, this involves *negative requirements* (which cannot be present) and *positive requirements* (whose presence is necessary for legitimation). According to Frederico Alvim, Rodrigo Zilio, and Volgane Carvalho, the negative requirements are: (1) the absence of objective doubt or a state of indeterminacy; (2) the absence of statements that are not subject to the regime of factual truth (such as artistic and religious manifestations, for example); (3) the absence of statements that cannot be verified; and (4) the absence of transparency regarding the untruthful nature of the statements. The *positive* requirements, on the other hand, are: (1) awareness of the falsehood (factual or contextual); (2) expressive harmfulness; (3) relevant scope; and (4) harmful intent (which may eventually overcome doubts about the presence of the first requirement). Furthermore, the aforementioned authors state that:

> [I]n a scenario that is coherent with the guarantee of a free and plural public debate, there are no reasons for cutting off *harsh criticism based on correct premises,* nor for obstructing *indecisive erroneous statements* (with limited scope). Likewise, the intention to mislead should be evaluated, since in many circumstances mere *mockery* or *involuntary errors* do not merit judicial attention. [...] the curtailment of freedom of expression is not justified in the face of *innocuous accusations* or *banal thoughts, based on the mention of* trivial or innocuous facts or circumstances. Moral judgments naturally take part in

[99]TSE Resolution No. 23.610/2019 – Article 9-F. "Paragraph 4. Any exercise of police power that is at odds with or exceeds that provided for in Paragraph 1 of this article will permit the use of an administrative electoral complaint, subject to the provisions of Articles 29 and 30 of TSE Resolution No. 23.608/2019" Available at: [https://www.tse.jus.br/legislacao/compilada/res/2019/resolucao-no-23-610-de-18-de-dezembro-de-2019]. Viewed: 4.5.2024.

> political confrontations, and Lilliputian offenses, being innocuous and commonplace, should not be treated as anomalies.
>
> (Alvim *et al*., 2023, pp. 323–324)

In addition, it is stipulated that within the established deadline, digital platforms must not only comply with takedown orders but also, when specifically determined, contribute to supplying the repository, indicating: (1) the text, image, audio, or video file which is the subject of the takedown order; (2) screenshots containing all the comments available on the site hosting the content; (3) metadata relating to access, such as the Internet Protocol, port, date, and time of publication; and (4) metadata relating to publication's participation at the time of its withdrawal (Article 9-G, §2).

The order imposes additional burdens on digital platforms which, as a general rule, must not only comply with the orders but also report on their compliance. It also allows the judicial authorities, by means of an express order, to oblige Internet application providers to supply the repository with material that incorporates a considerable amount of information. The problem is that, taking into account the magnitude of elections in certain countries (remember that, for example, the Brazilian context involves more than 2 600 different authorities in action), the burdens appear to be disproportionate and perhaps counterproductive, since the platforms' attention is divided: Instead of focusing mainly on the rapid elimination of posts with harmful content, at the same time they will devote themselves to fulfilling (possibly slow) bureaucratic-administrative tasks.

To make matters worse, setting evidently short deadlines for supplying the repository with material could make it inapplicable, placing the platforms in a permanent state of involuntary defiance. In any case, it is to be hoped that relevant data, such as the comments thread (important for demonstrating the harmfulness and social repercussion of withdrawn messages) and information on engagement (useful for verifying the scope and degree of discomfort generated by withdrawn content) will be effectively used by control bodies and knowledge-producing institutions, thus making it possible to better understand and combat the phenomenon of disinformation. It should also not be forgotten that the revenues of the big tech giants are more than sufficient to strengthen and adapt their workforce; while the task may be complex and difficult to accomplish, is not at all disproportionate.

Article 9-H goes further and reproduces a regulatory guideline for the Brazilian legal system, establishing the logic of *non bis in idem,* explaining that the effective removal of content does not exonerate the users involved of their

responsibilities: These can subsequently be fined in the context of electoral representations. In this sense, the removal of content would not prevent the cumulative establishment of monetary sanctions, as long as the guarantee of an adversarial process and full defense in a subsequent trial is observed.

At a more advanced point, the TSE, very attentive to the profit orientation that guides the platforms' decisions and behavior – despite the image they project through incisive public relations activities, always characterized by "laudatory rhetoric" (Morozov, 2018, p. 48) – makes the profit motive conditional on an additional precautionary framework, which is especially valid for companies that profit from the elaboration of user profiles and the consequent offer of content enhancement. In this segment, it incorporates recommendations that are also included in a report drawn up by the Venice Commission (2020, p. 8) on the fundamental principles for preserving integrity in the context of new digital technology.

3.2.6 Transparency and Data Protection

Article 27-A[100] states that providers who exploit the market for the artificial amplification of messages in digital media, including through the prioritization of search results, must maintain a complete repository of advertisements, equipped with an effective consultation mechanism that allows for easy social supervision of the financing of the circulation of political content.[101] As a result, society, academic bodies, the press, electoral competitors,

[100] "Article 27-A. The application provider that provides the service of boosting political and electoral content, including in the form of prioritizing search results, must: I – maintain a repository of these advertisements in order to monitor, in real time, the content, the amounts, those responsible for payment and the characteristics of the population groups that make up the audience (profiling) of the contracted advertising; II – provide an accessible and easy-to-use consultation tool that allows advanced searches to be carried out on the data in the repository, containing, as a minimum: (a) ad searches based on keywords, terms of interest and advertiser names; (b) access to precise information on the amounts spent, when it was boosted, the number of people reached and the segmentation criteria defined by the advertiser at the time of delivering the ad; (c) systematic collection, through a dedicated interface (application programming interface – API), of the data of the advertisements, including their content, spending, reach, audience reached and those responsible for payment."

[101] "Paragraph 1. For the purposes of this article, political-electoral content, regardless of the classification made by the platform, is characterized as content that deals with elections, political parties, federations and coalitions, elected positions, people holding elective office, candidates, government proposals, draft laws, the exercise of the right to suffrage and other political rights or matters related to the electoral process."

and control bodies will be able to monitor, in real time, the actors, values, and data used in the distribution of profiled messages through simple scraping (searching for ads by keywords, terms of interest, or advertiser names) or in an automated and professional way through APIs (application programming interfaces).

From this it follows that the amounts invested, the period of time contracted, the number of users reached, and the psychometric or sociodemographic criteria used to define the target audience will be made public, which will give Brazilian elections a transparency that is probably without precedent anywhere else in the world.

These measures, which apply not only during the official campaign period but on a permanent basis (Article 27-A, II), must be applied within a maximum period of two months from the entry into force of the Resolution (in the case of platforms already established in Brazil) or from the start of service provision, in the case of future providers who begin to offer their services in the Brazilian market. Appropriate conduct, in particular, is seen by the regulations as an obligatory condition for accreditation before the electoral justice system and consequently, as an indispensable requirement for obtaining income for activities of this nature (Article 27-A, §4, of TSE Resolution No. 23.610/2019).

In addition, the changes introduced by Resolution No. 23.723/2024 in the Advertising Resolution (No. 23.610/2019) also represent a significant milestone in efforts related to the protection of personal data, with the aim of ensuring that the use of emerging technology takes place in a responsible and transparent manner in the context of electoral campaigns. This action is especially relevant given that neither Brazilian data protection legislation nor electoral legislation has expressly dealt with data processing in the context of elections (Souza, 2022, p. 78[102]). Although it would have been preferable for the matter to be regulated in detail by the legislative branch, because of the more democratic nature of its regulatory work, the Brazilian institutional model has endowed the TSE with the regulatory authority to issue resolutions aimed at safeguarding the smooth running of elections (Souza, 2022, pp. 81–82).

[102]The author mentions a few provisions that address the issue tangentially, such as Articles 57-E and 57-G of the Electoral Law (Law No. 9.504/97), which deal with the prohibition of the use and donation of electronic records and the right of the interested party to demand that political communications be stopped (Souza, 2022, pp. 78–79).

To this end, §4 of Article 10 establishes that data processing must be adjusted to the purpose for which the data was collected, and that the principles and rules of Law No. 13.709/2018 – the GDPR – must be observed.[103]

In accordance with this law, the Resolution (Article 10, §5) stipulates that electoral agents must provide clear information on the processing of personal data as well as creating channels for voters to exercise their individual rights, such as requesting deletion or removal from databases (Article 10, §5). This measure reinforces individuals' control over their information in the context of political competitions (informational self-determination). In the same sense, §9 of Article 28 states that electoral publicity involving the processing of sensitive personal data must comply with the conditions laid down in the GDPR.[104]

[103] When AI tools assume control over campaigns, they bring with them a set of ethical and normative postulates pertinent to conduct guided by the technology itself, in connection with the framework for dealing with the rights and duties of the digital society. This is what happens, for example, with the rules laid down in data protection and Internet application regulation laws. Furthermore, and as is obvious, they are subject to the structuring discipline of electoral law, which includes the inescapable subjection to a set of general and specific principles. Basically, these premises start from the basic idea that technological innovations, in principle, do not require a complete and foundational treatment, provided that statutory protections can be adjusted, by interpretative means, to maintain the coherence of the system and the fulfillment of its social purpose. As Lucrecio Rebollo (Rebollo Delgado, 2023, p. 52) summarizes, in the midst of this scenario "judicial protection and the means of guarantee already exist, what varies are the forms of infringement, so it is not necessary to create an *ex novo* legal system, it is necessary to adapt the existing one to new needs, and this is a task for both the legislator, jurisprudence and also doctrine."

[104] Law No. 13709/2018 – "Article 11. Processing of sensitive personal data. The processing of sensitive personal data can only take place in the following cases: I – when the data subject or their legal representative consents, in a specific and separate manner, for specific purposes; II – without providing the data subject's consent, in cases where it is indispensable for: (a) the fulfillment of a legal or regulatory obligation on the part of the person responsible for the processing; (b) the shared processing of data necessary for the execution, by the public administration, of public policies provided for in laws or regulations; (c) the carrying out of studies by a research body, guaranteeing, whenever possible, the anonymization of sensitive personal data; (d) the regular exercise of rights, including in a contract and in judicial, administrative and arbitration proceedings, the latter in lo/.307, 23 September 1996 (Arbitratrion Law); (e) the protection of the life or physical safety of the interested party or a third party; (f) the protection of health, exclusively in procedures carried out by health professionals, health services or health authorities; or (g) guaranteeing the prevention of fraud and the security of the data subject, in the processes of identification and authentication of registration in electronic systems, safeguarding the rights mentioned in Article 9 of this law and except in the case where the fundamental rights and freedoms of the data subject that require the protection of personal data prevail."

The regulation also establishes that in municipal elections in cities with fewer than 200 000 voters, political parties, and related bodies (coalitions and federations) as well as candidates, will be considered as small processing agents, taking into account the provisions of Resolution 2/2022 of the National Data Protection Agency (ANPD). Thus, in the terms of Article 10, §6-B: (1) they are exempt from designating the person responsible for the processing of personal data; and (2) they have the right to establish a simplified information security policy, "which must include the essential and necessary requirements for the processing of personal data, with the aim of protecting it from unauthorized access and from destruction, loss, alteration, communication or any form of inappropriate, accidental or unlawful processing."

The new Article 33-B assigns some specific obligations when processing data to application providers and electoral actors some specific obligations, namely: (1) guaranteeing easy access to processing information, especially to the data used to create user profiles, a measure aimed at micro-segmented advertising;[105] (2) guaranteeing those affected the rights provided for in Articles 17–20 of the GDPR (access to data, anonymization, revocation of consent, etc.); (3) the adoption of measures necessary to protect against unlawful discrimination and abuse; (4) to refrain from using the data for purposes other than those explained and consented to by the respective parties, in accordance with the principles of purpose, necessity, and appropriateness; (5) the application of security settings to protect against unauthorized access and avoid situations that give rise to leaks; and (6) notifying both stakeholders and the national protection authority in the event of an incident.

[105] According to the glossary included in Article 37 of the same resolution, profiling is understood as "the processing of multiple types of data on identified or identifiable individuals, generally carried out in an automated manner, with the aim of forming profiles based on patterns of behavior, tastes, habits and preferences and classifying these profiles into groups and sectors, using them for analysis or predictions of movements and trends of political and electoral interest" (point XXXII). Microtargeting, on the other hand, "is the segmentation strategy of electoral publicity or campaign communication that consists of selecting people, groups or sectors, classified through the elaboration of profiles, as the target audience of the messages, actions and political-electoral content elaborated on the basis of the profiled interests, with the aim of increasing influence on their behavior" (point XXXIII).

In addition, parties, federations, coalitions, and candidates are obliged to keep a record of the data-processing operations carried out, detailing the minimum information that must be included (Article 33-C).[106] Section 2 of this Article establishes the obligation to keep these records of operations throughout the electoral period, and the obligation is maintained in the event that a lawsuit is filed to investigate any irregularity or illegality in a campaign's data processing. The authorities may order the submission of these records of operations, duly accompanied by supporting documents (§3).

Another interesting point that reinforces data protection in the context of campaigns is the prohibition of the sale, donation, and transfer of customer data by legal entities. This is effected by Article 31, which also prohibits the sale of electronic records of addresses and databases of legal persons or legal entities. The only exception, introduced by paragraph 1-B of Article 31, concerns personal contact details obtained legitimately by a legal person. In these cases, their free transfer to political parties as well as their use is allowed provided there is prior consent from the recipients of the electoral publicity (the data owners).[107]

In addition to the specific rules, AI solutions, to be used legally in the quest for votes, are also subject to the observance of generally applicable campaign rules. In all cases, candidates are obliged to record their spending in accountability processes, and it is also desirable to adopt good practices in terms of transparency and good faith by indicating publicly and explaining which tools have been used and to what purpose.

---

[106] For the purposes of the present resolution, political parties, federations, coalitions, candidacies, and candidates must keep a record of personal data processing operations, containing, as a minimum: (1) the type of data and its origin; (2) the categories of data subjects; (3) a description of the process and purpose; (4) the legal basis; (5) the expected duration of the processing, in accordance with Law 13.709/2018; (6) the period during which the personal data will be stored; (7) a description of the flow of exchange of personal data, if applicable; (8) the contractual instruments specifying the role and responsibilities of the controllers and operators; and (9) the security measures used, including good practice and governance.

[107] "§1-B. The record of personal contact data, legitimately held by a physucal person, may be transferred free of charge to a political party, federation, coalition, candidate or candidate, with its lawful use in the campaign conditional on the prior, express and informed consent of the recipients in the first contact by message or other means."

3.3 A GLOBAL, COMPREHENSIVE, AND NECESSARY RESPONSE

The design of technology is also the design of society. Within this techno-social design, AI is taking on a growing role, opening up a field of opportunities in areas such as scientific research and business. It can also help societies to organize themselves more efficiently and provide better services to their citizens, without ruling out undesirable situations in which the rule of law may be replaced by the rule of the algorithm, thereby threatening human rights and fundamental freedoms and weakening democratic processes.

The use of AI in elections is situated in a general context in which we have moved from the utopian vision of placing all hopes for democratic regeneration in technology (due to its impetus for transparency, participation, and accountability) to an apocalyptic vision, according to which that all the evils that beset democracies stem from technology, with the former breathing its last gasps as a result of the latter. In these circumstances, electoral processes occupy a central place in the legitimacy of the democratic system, and the use of AI, due to its complexity and opacity, can contribute to this reactive rejection.

The automation of certain processes, through algorithms, and particularly through the use of AI systems in the electoral machinery, is transforming the dynamics of elections. In this way, different mechanisms have been incorporated into campaigns which, driven by AI, are increasing the reach and efficiency of electoral proselytizing.

This type of campaign affects the voter's "decision-making architecture," fostering an increasingly personalized opinion based on segmentation and micro-segmentation; content generation with GenAI that blurs the distinction between non-fiction and fiction, especially via the use of deepfakes, and their dissemination and re-dissemination through automated mechanisms. This was signaled by Sam Altman, executive director of OpenAI, during his appearance before the US Senate in May 2023; when asked if he thought AI-based language models, such as ChatGPT, could be used to make voters behave in a certain way, he showed his concern that these mechanisms could indeed be used to persuade, involve, and manipulate candidates' relationships with voters.

Despite these warnings, the tendency remains for AI to consolidate itself as an indispensable tool for the effective exercise of passive political rights, since underutilizing its potential will diminish the chances of success in electoral contests. In the very near future, campaigns without technology,

especially in large constituencies, will be increasingly expensive and less effective, since generalist communication seems less and less capable of mobilizing voters. Disregarding the gold mine of data and high-performance tools that is opening up on the horizon will be tantamount to challenging hordes of snipers blindly and without weapons.

The emerging model affects central aspects of democratic legitimacy. It seems indisputable that "the problems of information manipulation are systemic" and need "to be addressed by recognizing the role of the actors involved and their responsibilities in maintaining the rights of data subjects and democracy itself" (Frazão, 2022, p. 567). This situation is aggravated when the arsenal of truth deconstruction engulfs the core of the political agenda, taking on attacks from *accelerationist*[108] segments that intend to destabilize society and create the perfect conditions for authoritarian uprisings that use non-existent fraud in electoral processes as their pretext.

AI systems reconfigure the public sphere, diminishing its rationality and its free and informed nature. They resurrect deformations and inequalities, affecting the way in which information circulates and is accessed by the body politic.

Fung and Lessig have given this phenomenon a name: Clogger. The starting point is that the most decisive thing about the "new algorithmic order," of which AI is a part, is that it absolutely conditions our access to information, selectively predetermining it. As well as being able to manipulate search engine algorithms so that websites or news items containing false information appear first, AI entails, as we have seen, a gradual surge in the effectiveness of microtargeting mechanisms and influences on behavior. Thus, the dream of personalized messaging is turned into a feasible reality, thanks to AI's personalized, dynamic, and adaptive nature: (1) the generation, thanks to its language models, of messages in different formats that are truly personalized without any real limit as the entire process of segmentation, creation, and distribution of the message will be automated; (2) the use of reinforcement learning techniques, based on machine learning and trial and error, serves to generate a series of messages which, adapting to the responses, gradually increase their level of persuasion and their effectiveness; (3) this

[108] Within the literature on extremist movements, "accelerationism" characterizes groups that "seek to accelerate chaos" and employ various means to do so, including the use of digital networks to expand and consolidate the reach of diverse conspiracy theories (Prado, 2023, p. 226).

learning will also be fueled by the responses of other voters, through dynamic conversations in which the messages that secure the best persuasive results in similar communities at a specific time are increasingly refined and used for the rest of the community with the consequent adaptation to the particular elements of each one. In addition, authors also point out that this work of persuasion can be carried out even without using strictly political messages, diverting the attention of voters selected for their sympathy with the political opponent using messages about their hobbies, which are much more attractive than politics, or with unpleasant messages or ads which "hide" the political content. As we have also seen, this can also be achieved through persuasion of a voter's trust groups that are unconnected to politics, persuasion involving self-referential micro-communities, repositories of different truths shared only by equals. This occurs without those affected having any way of knowing the strategies used in this exercise in persuasion (Fung; Lessig, 2023).

The use of these techniques by one of the candidates, like any other technology that bursts onto the campaign scene, would generate an imitation effect, which would lead to their generalization in a short space of time and risk converting the campaign into a war of technologies, in which the most effective would triumph. This outcome would in practice signify the end of democracy, even while some vestiges of the electoral process were maintained – such as speeches, advertisements, messages, voting, and vote counting – but in a space where public opinion as well as being fragmented, would be induced, unreal, and falsified.

However, the most serious risk for democracy seems to be the erosion of trust: In the age of AI, fiduciary consensus is being undermined by high-yield fallacies, and the social fabric is falling apart in an environment of permanent suspicion. Distrust and division come together to increase hostility, insecurity, and polarization, revealing a world in which instability seeks to impose itself as the new normal.

Political hatred and radical polarization poison the democratic project, which is linked to the condition of pluralism. The validity of the ideals of respect, tolerance, and dialog would be the corollary of a model of society in which differences do not become ever greater but interact in the search for possible consensus (Frazão, 2022, p. 567). Normal and legitimate elections only thrive when democratic values prevail, which demands – from society as a whole – greater attention to the consequences of normalizing the use of technology from a social perspective.

The development of digital democracy – with AI at its core – changes social behavior, dictates a new regime of political information, and reconfigures the competitive conditions of the electoral arena. It entails a new set of

systemic challenges for electoral institutions and the substratum of citizenship and as a result, signals the need to revise agreements with a view to safeguarding the ideas of freedom, equality, integrity, transparency, and justice. Both understanding and demythologizing the threats and identifying and addressing vulnerabilities are essential steps to keeping popular sovereignty safe from fraud, cheating, and manipulation.

3.3.1 First Global Responses

The "Report on AI Governance for the Benefit of Humanity" focuses on the need to establish a global governance framework for these systems. The report, prepared by the High-Level Advisory Body on AI, explicitly refers to the risk of "the increased creation and dissemination of disinformation." The document highlights the difficulty of regulating AI under traditional regulatory frameworks and the gaps in the global governance of these techniques, with consequent problems of coordination, representation, and effective implementation when 118 countries are not participating in any AI governance initiative. It thus proposes the creation of a new social contract for AI, a comprehensive, inclusive, and collaborative governance that regulates AI globally to ensure that this technology is deployed for the benefit of all humanity without leaving anyone behind.

The Organisation for Economic Co-operation and Development (OECD) has also developed a recommendation on AI, approved in 2019 and updated in 2024, which, in its section on respect for human rights and democratic values, includes as risks: misleading information and disinformation amplified by AI.

The recommendation highlights the need to address these threats "while respecting freedom of expression and other rights and freedoms protected by applicable international law," such as the right to access information. This means that while it is necessary to take measures to limit the malicious use of AI in spreading misleading information, these actions must be proportionate and should not inhibit people's right to share opinions and express themselves freely. In this regard, it is recommended that AI actors implement "mechanisms and guarantees such as human oversight and intervention capabilities, addressing risks stemming from unintended or deliberate misuse, appropriately to the context and consistent with the current state of technology."

To address disinformation and other social risks posed by AI techniques, the recommendation calls for cooperation between governments, technology companies, and other stakeholders. This implies sharing best practices, collaborating on the creation of international standards regulating AI use in generating and disseminating information, and developing appropriate legal frameworks to address the problem.

Furthermore, the recommendation encourages governments to promote digital literacy, making people more aware of the existence of AI-generated disinformation and better able to detect misleading content. It also discusses supporting the creation of technological tools that help detect and automatically remove false or manipulated information while respecting fundamental rights.

Among the proposals relevant to this study are transparency and explainability, which direct AI actors to provide sufficient information so that affected parties can understand how AI systems work and how they impact specific interactions, thus enabling them to challenge outcomes if they are negatively affected. This transparency would include, according to the recommendation, "transparent and understandable information about data sources/inputs, factors, processes, and/or reasoning underlying predictions, content, recommendations, and decisions," which can help those affected by an AI system to understand the outcomes.

For electoral processes, it is also interesting to highlight the emphasis on robustness and security, stating that AI systems must be robust, secure, and protected, with the ability to correct or dismantle systems if they threaten to cause harm. It also points to the responsibility of the actors involved to respect the outlined principles, with mechanisms such as ensuring traceability and "permanently applying a systematic risk management approach at each phase of an AI system's life cycle."

Additionally, the potential harm mitigation principle applies to disinformation, meaning that when an AI system is detected as having been used to spread disinformation, AI actors should have the ability to correct or dismantle those systems. In our area, this could mean removing maliciously generated content, adjusting algorithms to prevent future errors, or creating more robust systems that are not easily prone to abuse.

3.3.2 The Foundations of Regulation

In the absence of specific regulations for elections, it is necessary to establish a clear and solid legal framework for the use of AI in campaigns, without delegating this task to companies in the sector. This framework would need a position on: (1) the protection of privacy; (2) responsibility for content produced with AI; (3) the advertising models in which AI plays a key role, to prevent it from affecting freedom of decision, increasing transparency (with open-access libraries of political ads), and taking a clear stance on the use of microtargeting by setting limits on its use or even calling for a ban on the use of these techniques in electoral processes (on the basis that freedom of expression does not protect machine speech); (4) respect for due process guarantees when AI is left in charge of moderating and eliminating content as well as the use of this technology to combat disinformation (to ensure that

it is truly independent); (5) the level of transparency required (with measures such as those that prevent bots from presenting themselves as people by identifying content generated by machines); and (6) accountability, facilitating access for researchers, data-verification initiatives, and civil society organizations to evaluate the impact of AI on online political campaigns.

In order to do so, legislators must answer a series of key questions: what is the purpose of regulation, what behaviors should be covered, how to regulate these behaviors, and who should be responsible, specifically whether regulation should target only those who create or disseminate *deepfakes* or also the online platforms that facilitate their transmission.

Regulating the use of AI in campaigning is crucial to fostering an informed electorate and protecting the integrity of the electoral process. It must therefore address the various democratic risks related to the role of AI in electoral processes, which have been analyzed in this work. Specifically:

1. *Regarding misuse*: limiting external interference, algorithmic pathologies, microdirected messages, and psychometric blackmail.
2. *Regarding the manipulation of public opinion*: confronting disinformation and conspiracy theories, discrediting inauthentic behavior,[109] and deepfakes.
3. *Regarding self-determined opinion*: reinforcing citizen autonomy, enabling them to critically evaluate communication and develop skills to distinguish facts from opinions, and to check the consistency of evidence and lines of argument (information education); avoiding exposure to false information, evaluating the credibility of sources, identifying the characteristics of fake news, and encouraging the ethical and responsible use of social networks (media education);[110] spreading constitutional values, based on co-responsibility between the political class and civil society and on the need for peace, tolerance, and respect between all people (education for democratic culture).

[109] The term denotes "techniques used to artificially publish, promote and disseminate organic participation on digital platforms through fake accounts that interact by validating comments, distributing likes or sharing posts between them" (Cardiel Soto *et al.*, 2022, p. 53). In this context, the need for new mechanisms to curb the acquisition of bot accounts is asserted as well as the setting up of click farms, troll farms, etc.

[110] "In an increasingly digital society, there is no equal access to information proceeding from information and communication technology, causing a digital divide and unequal opportunities for the most vulnerable groups." In these circumstances, the need arises for "new skills such as information selection, coding, searching for information and the ability to interpret it," so that individuals are less exposed to disinformation and infoxication (Morales Romo, 2023, p. 260).

4. *Regarding information disruption*: increase the supervision of online advertising, establish labeling systems, and verify the authorship of AI-generated content in political campaigns.
5. *Regarding the use of AI systems with a discriminatory bias*: regulating social-media platforms to ensure fair moderation and neutral algorithms. Algorithms have become the "disciplinary apparatus of our time" (Beiguelman, 2021, p. 46) and as such, it is necessary to reform their comprehensibility so that *stakeholders* can be clear and confident about their ethical, responsible, and politically neutral functioning.
6. *Regarding access to information*: guarantee online neutrality and an open Internet so that any access restrictions are based on legal provisions. We have seen that search engines and platforms: (a) impose opaque computational rules that selectively hide plural points of view and commodify user attention; (b) normalize hyper-personalized campaigns that isolate people and subject them to an arbitrary and restrictive information diet; (c) strengthen disinformation campaigns, amplifying hate speech and harmful content; (d) unbalance electoral processes, thus structuring an asymmetrical playing field in which access to data technology segregates the disadvantaged from the advantaged. In light of these issues, it is no exaggeration to state that "the digital communications revolution [...] has been mistreating our elections" (Moore, 2023, p. 11).
7. *Regarding AI cyber threats*: offer quick and efficient responses, relying on the platforms' collaboration, in a permanent balance so that the reduction in the circulation of messages does not degenerate into a mode of private censorship.[111]
8. As far as the *subjects* are concerned, most of the regulations should be aimed at candidates, political action committees (PACs), and others who use techniques that jeopardize the integrity of the elections, but they should not neglect specific regulations for online platforms and other intermediaries that are indispensable for the creation and dissemination of this type of content.

In short, there is a need for balanced regulation that addresses the risks posed by the use of these techniques without placing excessive restrictions on freedom of expression.

[111] In this context, there is a need for transparency, responsibility, and accountability mechanisms in relation to AI technological applications and Internet intermediaries in the context of electoral processes. There is also a need for the mandatory adoption of ethics codes and corporate social responsibility codes among digital service providers and in this case, developers of AI applications.

3.3.3 The Central Role of Electoral Bodies

To achieve these objectives, it is essential to prepare electoral bodies to deal with emergency situations by providing them with the necessary resources and specific training to enable the effective integration of information and communication technology, including AI, as well as the management of the risks associated with cybersecurity. This means strengthening their skills and increasing their material and human resources so that they are prepared to deal with the complexity that comes with increasing automation, always with the awareness that at the present time, it is not possible to deal with AI without using AI.[112]

Democratic resilience and the corresponding reorganization of electoral bodies must develop in parallel with the preservation of individual freedoms and in accordance with an understanding of their limits, which is fundamental if we want to face cognitive threats – which are becoming increasingly larger and more complex – in harmony with the dictates of democracy. Maintaining a "healthy virtual environment [...] can only be achieved by paying attention to a freedom of expression free from distortions" (Branco; Branco, 2022, p. 67).

AI introduces new tools that change the rules of the democratic game. By modifying the communicative sphere, AI is challenging traditional social structures and transforming the contemporary logic of electoral contests, as has already occurred many times throughout history. In this context, electoral organizations face a double challenge: (1) the incorporation of this technology with a view to expanding and protecting democratic principles and (2) the mitigation of its adverse effects on the integrity of electoral processes, achieved in an effective manner.

Nonetheless the response, which we have seen is more necessary than ever, must go far beyond the legal sphere and must also include technological, communicative, cultural, political, and educational components to be effective. This comprehensive vision must have a global perspective that

[112] "However, in 'the future' both the 'offensive' and the 'defensive' – both the dissemination of disinformation and efforts to combat it – will become increasingly automated and entrusted to AI. The language-generating AI GPT-3 has demonstrated its ability to create fictional characters, use them to produce a language characteristic of hate speech and engage in conversations with human users to instill prejudices and even incite violence. If such an AI were deployed to spread hatred and division on a large scale, human beings alone might not be able to combat it as a result. Unless AI is paused at an early stage of its take off, identifying and manually deactivating all of its content through research and individual decisions would be very difficult, even for the most sophisticated governments and network platform operators. For such a vast and arduous task, they would have to resort – as they already do – to AI content moderation algorithms. But who creates and supervises these, and how?" (Kissinger *et al.*, 2021, p. 105).

respects democratic principles, guaranteeing not only freedom of expression and the right to information but also privacy and the freedom to vote, the right to participate in public affairs as well as fairness and integrity in the processes, all principles that are affected by the use of AI in campaigns.

The guarantee of freedom of informative expression, although it "protects the articulation of opinions, convictions, comments, valuations or judgments on any matter, including questions of public interest," can – despite its protected position – be subject to restrictions,[113] provided that these are laid down in law (the principle of legality), pursue a legitimate aim, and that restrictions avoid a greater harm than they cause (necessity and proportionality). Freedom of expression is directly related to the right of access to adequate information, and it is the combination of both that establishes the criteria of proportionality, which in the electoral sphere must be judged with special attention. However, freedom of expression is not an obstacle to the right to information but rather its main guarantee. Hence, during elections, defending freedom of expression is the only way to "protect citizens' access to truthful information, so that they can make their personal, social or political decisions in a well-founded manner" (Mendes, 2022, pp. 70, 73).[114]

[113] "From this perspective, the fundamental rights of access to information and freedom of expression must be interpreted in the best possible way, in order to guarantee the preservation of their essential core. And false information can weaken these rights – so valuable to the construction of the democratic rule of law – to the extent that it is disseminated without any analysis of the veracity of its content. For this reason, fake news influences the formation of public opinion in such a way that it jeopardizes the formation of the democratic rule of law, since it eliminates the citizen's ability to follow acts of public life in a comprehensive and coherent manner" (Mendes, 2022, p. 75).

[114] "The image of a marketplace of ideas, in which the clash between them will make the most positive prevail and the truth emerge, is important in the sense of recognizing freedom of expression in the broad sense as a central element of democracy and individual autonomy. However, its uncritical adoption, without limitations to certain discourses, represents a risk for democracy itself. A free-market model without restrictions makes it possible for information that is far removed from the purpose of dissemination and that calls the democratic system into question to act in an unlimited manner and even prevail, in the face of the evident realization that people are not always capable of rationally identifying falsehood and the manipulation of the truth" (Biolcati, 2022, p. 127). Adapting Rafael Alcácer's graphic explanation of hate speech, in terms of constitutional interpretation disinformation, is seen as a negative delimiting element of the right to freedom of expression so that the relationship between disinformation content and freedom of expression will be that of "tangent circles:" what disinforms does not enter into freedom of expression, and vice versa (Alcácer, 2023, p. 41), adding that the same reasoning applies to attacks against vulnerable or minority groups (Piras, 2021, p. 32).

Attempting to pit the defense of the right to information against freedom of expression would be akin to cutting off the oxygen to prevent the spread of viruses, a sure guarantee of death by asphyxiation.

Safeguarding the integrity of campaigns means protecting electoral processes and democracy. To do this, it is important to understand the profound changes that new technology is causing in order to design proportionate and forceful responses that prevent public debate from descending into a kind of artificial rambling, in which reality is distorted, privacy is hacked to exert undue pressure, and the algorithms close the door to the plurality of the world. Ultimately, the challenge is to ensure that AI technology is applied in ways that are compatible with fairness and authenticity, with autonomy and the freedom to define the vote.

BIBLIOGRAPHY

Aguado Terrón, Juan Miguel; Villaplana Jiménez, F Ramón. Guerras culturales, desinformación y moralización del discurso público. In: Carratalá, Adolfo; Iranzo-Cabrera, María; López-García, Guillermo (eds.). *De la desinformación a la conspiración: política y comunicación ante un escenario híbrido*. Valencia: Tirant lo Blanch, 2023, p. 201–216.

Alcácer Guirao, Rafal. Delitos de odio y discurso del odio: clarificaciones conceptuales. In: Figueruelo Burrieza, Ángela (ed.). (dir.); Martín Guardado, Sergio (coord.)*Desinformación, odio y polarización*. Vol. I. Navarra: Arazandi, 2023, p. 23–44.

Alvim, Frederico Franco. *Curso de Direito Eleitoral*. 2. ed. Curitiba: Juruá, 2016.

Alvim, Frederico Franco. *Abuso de poder nas competições eleitorais*. Curitiba: Juruá, 2019.

Alvim, Frederico Franco; Zilio, Rodrigo López; Carvalho, Volgane Oliveira. *Guerras cognitivas na arena eleitoral: o controle judicial da desinformação*. Rio de Janeiro: Lumen Juris, 2023.

Archegas, João Victor; Maia, Gabriella. O que é a inteligência artificial (IA)? Análise em três atos de um conceito em desenvolvimento. In: Cadernos Adenauer XXIII - Inteligência Artificial: aplicações e implicações. Rio de Janeiro: Fundação Konrad Adenauer, (2022), pp. 9–28.

Beiguelman, Giselle. *Políticas da imagem: vigilância e resistência na dadosfera*. São Paulo: Ubu, 2021.

Bender, Sarah M L. Algorithmic elections. *Michigan Law Review*, 121(3) (2022), p. 489–522.

Biolcati, Fernando Henrique de Oliveira. Eleições e a importância do engajamento dos provedores de redes sociais no controle das fake news. In: Branco, Paulo Gustavo Gonet; Fonseca, Reynaldo Soares da; Branco, Pedro Henrique de Moura Gonet (eds.). *Eleições e democracia na era digital*. Brasília: Almedina, 2022, p. 127–144.

Bioni, Bruno; Almeida, Virgilio; Mendes, Laura Schertel. Inteligência artificial e ameaça a integridade das eleições. **Folha de São Paulo,** 17 February 2024. Available at: [https://www1.folha.uol.com.br/ilustrissima/2024/02/inteligencia-artificial-e-ameaca-a-integridade-de-eleicoes.shtml]. Viewed: 19.02.2024.

Bioni, Bruno; Garrote, Mariana; Guedes, Pedro. *Temas centrais na regulação de IA: o local, o regional e o global na busca da interoperabilidade regulatória*. São Paulo: Associação Data Privacy Brasil de Pesquisa, 2023.

Birnbaum, Pierre. Raça/racismo. In: Hermet, Guy; Badie, Betrand; Birnbaum, Pierre; Braud, Philippe (eds.). *Dicionário de Ciência Política e das instituições políticas*. Lisbon: Escolar, 2014, p. 254–255.

Borja, Rodrigo. *Enciclopedia de la Política*. Ciudad de México: Fondo de Cultura Económica, 1997.

Braga, Sérgio de Paula. Racismo, colorismo e afro-conveniência nas cotas raciais eleitorais. Conjur, 26 September 2022. Available at: [www.conjur.com.br/2022-set-26/direito-eleitoral-racismo-colorismo-afro-conveniencia-cotas-raciais-eleitorais]. Viewed: 11.04.2024.

Branco, Paulo Gustavo Gonet; Branco, Pedro Henrique de Moura Gonet. Fake News – desafios para a democracia. In: Branco, Paulo Gustavo Gonet; Fonseca, Reynaldo Soares da; Branco, Pedro Henrique de Moura Gonet; Velloso, João Carlos Banhos; Fonseca, Gabriel Campos Soares da (eds.). *Eleições e democracia na era digital*. Brasília: Almedina, 2022, p. 51–68.

Bustamante Tejada, Walter Alonso. Homofobia. Pereda, Carlos (ed.). Aragón Ribera, Álvaro; Delgado Parra, Concepción; Vega, Julieta Marcone; Leroux, Sergio Ortiz; Sermeño Quezada, Ángel (coords.). In: *Diccionario de injusticias*. Ciudad de México: Siglo XXI, UNAM, 2022, p. 391–397.

Campos, Rafael; Oliveira, Sabrina R. de; Santos, Célio Xavier dos. O conceito de dever de cuidado no âmbito das plataformas digitais. Conjur, 21 March 2023. Available at: [www.conjur.com.br/2023-mar-21/direito-digital-conceito-dever-cuidado-ambito-plataformas-digitais]. Viewed: 04.04.2024.

Cardiel Soto, Rodrigo Hidalgo; Alvim, Frederico Franco; Rondon, Thiago Berlitz. *Glosario contra la Desinformación*. Ciudad de México: Instituto Nacional Electoral, 2022.

Carratalá, Adolfo; Iranzo-Cabrera, María; López-García, Guillermo. Presentación. In: Carratalá, Adolfo; Iranzo-Cabrera, María; López-García, Guillermo (eds.). *De la desinformación a la conspiración: política y comunicación ante un escenario híbrido*. Valencia: Tirant lo Blanch, 2023, p. 13–24.

Carratalá, Adolfo; Peris-Blanes, Àlvar. (Des)'infroentetenimiento' en los magazines televisivos de actualidad: sesgos y bulos a propósito de la 'ley trans'. In: Carratalá, Adolfo; Iranzo-Cabrera, María; López-García, Guillermo (eds.). *De la desinformación a la conspiración: política y comunicación ante un escenario híbrido*. Valencia: Tirant lo Blanch, 2023, p. 237–254.

Castells, Manuel. *Comunicación y poder*. Madrid: Alianza, 2009.

Coeckelbergh, Mark. The Political Philosophy of AI, Cambridge, UK: Polity Press, 2022.

Coello De Portugal, José María. **Parlamento, derecho parlamentario y nuevas tecnologías:¿ una discusión nueva?** en Parlamentos Abiertos. Tics y Asambleas Legislativas. Congreso de los Diputados, 2014. p. 61–138.

Costa, Henrique. Extrema-direita. In: Souza, Cláudio André de; Alvim, Frederico Franco; Dantas, Humberto; Barreiros Neto, Jaime (eds.). (coords.). *Dicionário das Eleições*. Curitiba: Juruá, 2020, p. 293–294.

Council of Europe. Need for democratic governance of artificial intelligence. Committee on Political Affairs and Democracy, 23 September 2020. Available at: [https://assembly.coe.int/LifeRay/POL/Pdf/TextesProvisoires/2020/20200908-DemocraticAI-EN.pdf]. Viewed: 24.12.2023.

Council of Europe, 2024. Council of Europe Framework Convention on Artificial Intelligence and HumanRights, Democracy and the Rule of Law, Available at: [https://rm.coe.int/1680afae3c]. Viewed: 26.05.2025.

De Teffé, Chiara Spadaccini. Eleições, inteligência artificial e responsabilidade das plataformas digitais. **Migalhas,** 21 March 2024. Available at: [https://www.migalhas.com.br/coluna/migalhas-de-responsabilidade-civil/403881/eleicoes-ia-e-responsabilidade-de-plataformas-digitais]. Viewed: 22.03.2024.

Del Amo Castro, Ion Andoni. ¡Todo es mentira, borregos! Una aproximación a los discursos y bases sociales de la desconfianza en torno a la covid-19. In: Carratalá, Adolfo; Iranzo-Cabrera, María; López-García, Guillermo (eds.). *De la desinformación a la conspiración: política y comunicación ante un escenario híbrido*. Valencia: Tirant lo Blanch, 2023, p. 63–78.

Del Rey Morató, Javier. *Comunicación política, Internet y campañas electorales. De la teledemocracia a la ciberdemocr@cia*. Madrid: Tecnos, 2007.

Denemark, Jaroslav. The risk of artificial intelligence for democracy and the EU's first efforts to regulate it. *The Lawyer Quarterly*, 14(1) (2024). Available at: The Lawyer Quarterly (cas.cz). Viewed: 22.03.2024.

Dougall, Danielle; Ostrowski, James. What's in Biden's Executive Order on Artificial Intelligence? **Lawfare,** 3 January 2024. Available at: [https://www.lawfaremedia.org/article/what-s-in-biden-s-executive-order-on-artificial-intelligence]. Viewed: 05.01.2024.

Durães, Ulisses. Novo estágio das fake news: deepfake vira arma de campanha na Argentina. **UOL,** 18 November 2023. Available at: [https://noticias.uol.com.br/internacional/ultimas-noticias/2023/11/18/novo-estagio-das-fake-news-deepfake-vira-arma-de-campanha-na-argentina.htm]. Viewed: 26.04.2024.

Elorza Saravia, Juan Daniel. Tecnología de los valores: el caballo de Troya de la polarización en las democracias constitucionales. In: Figueruelo Burrieza, Ángela (ed.). (dir.); Martín Guardado, Sergio (coord.). *Desinformación, odio y polarización*. Vol. I. Navarra: Arazandi, 2023, p. 57–81.

Emcke, Carolin. *Contra o ódio*. Belo Horizonte: Âyiné, 2021.

European Comission, 2018. **Code of Practice on Disinformation**. Available at: [https://digital-strategy.ec.europa.eu/en/news/code-practice-disinformation]. Viewed: 26.04.2024.

European Comission. *2022 Strengthened Code of Practice on Disinformation*. Available at: [https://ec.europa.eu/newsroom/dae/redirection/document/87585]. Viewed: 26.04.2024.

United States, Executive Office of the President [Joe Biden]. Executive Order 14110: Safe, Secure, And Trustworthy Development And Use Of Artificial Intelligence. 30 October 2023. 88 FR 75191 (2023). Available at: [https://www.federalregister.gov/d/2023-24283]. Viewed: 05.01.2024.

Fisher, Max. The Chaos Machine: The Inside Story of How Social Media Rewired Our Minds and Our World. New York: Little, Brown and Company, 2022.

Frazão, Ana. A democracia na era digital: os riscos da política movida a dados. Branco, Paulo Gustavo Gonet; Fonseca, Reynaldo Soares da; Branco, Pedro Henrique de Moura Gonet; Velloso, João Carlos Banhos; Fonseca, Gabriel Campos Soares da. *Eleições e democracia na era digital. Brasília: Almedina*, 2022, p. 69–84.

Frazão, Ana. TSE e as regras para o uso de inteligência artificial nas eleições. **Jota,** 13 March 2024. Available at: [https://www.jota.info/opiniao-e-analise/colunas/constituicao-empresa-e-mercado/tse-e-as-regras-para-o-uso-de-inteligencia-artificial-nas-eleicoes]. Viewed: 31.03.2024.

Fung, Archon; Lessig, Lawrence. How AI could take over elections – and undermine democracy. **The Conversation,** 2 June 2023. Available at: [https://theconversation.com/how-ai-could-take-over-elections-and-undermine-democracy-206051]. Viewed: 18.12.2023.

García Fajardo, Adán. Discurso de odio. In: Pereda, Carlos (ed.); Aragón Ribera, Álvaro; Delgado Parra, Concepción; Vega, Julieta Marcone; Leroux, Sergio Ortiz; Sermeño Quezada, Ángel (coords.). *Diccionario de injusticias*. Ciudad de México: Siglo XXI, UNAM, 2022, p. 243–249.

García Mahamut, Raúl. Elecciones, protección de datos y transparencia en la publicidad política: la apuesta normativa de la UE y sus efectos en el ordenamiento español. *Revista Española de la Transparencia*, 17 (2023), p. 75–105.

Garrigues Walker, Antonio; González De La Garza, Luis Miguel. *El derecho a no ser engañado. Y cómo nos engañan y nos autoengañamos*. Navarra: Arazandi, 2020.

Gerlitz, Carolin; Helmond, Anne. The like economy: social buttons and the data-intensive web. *New Media and Society*, 15(8) (2013), p. 1.348–1.365.

Gillespie, Tarleton. Platforms throw content moderation at every problem. In: Zimdars, Melissa; McLeod, Kembrew (eds.). *Fake news: understanding media and misinformation in the digital age*. Cambridge: The MIT Press, 2020. https://doi.org/10.7551/mitpress/11807.003.0036

Gómez De Ágreda, Ángel. La paz es la víctima última de la mentira. Desinformación con base tecnológica en la guerra. In: Vázquez-Barrio, Tamara; Salazar García, Idoia (eds.). *Inteligencia artificial, periodismo y democracia*. Valencia: Tirant lo Blanch, 2023, p. 205–226.

González-Torre, Ángel Pelayo. TIC, inteligencia artificial y crisis de la democracia. En: Solar Cayón, José Ignacio. *Dimensiones éticas y jurídicas de la inteligencia artificial en el marco del Estado de Derecho*. Madrid: Universidad de Alcalá, 2020, p. 55–78.

Grijelmo, Álex. *La información del silencio. Cómo se miente contando hechos verdaderos*. Ciudad de México: Taurus, 2012.

Guaraty, Kaleo Dornaika. *Discurso de ódio no Direito Eleitoral. Conceito jurídico e hermenêutica*. Rio de Janeiro: Lumen Juris, 2023.

Hampton, Sarah. Parasite and Catalyst: The Polarizing Influence of Chatbots in Political Discourse. In: *Tesis de maestría*. Texas: University of Texas at Austin, 2019. Available at: [https://bit.ly/4a17MSp]. Viewed: 13.01.2024

Hernández Ramos, Mario. Los retos constitucionales de la inteligencia artificial. Prospectivas, debilidades y fortalezas de los paradigmas vigentes. In: Soberanes Díez, José María; Garduño Domínguez, Gustavo (eds.). (Coords.). *La interacción de las redes sociales, la tecnología y los derechos humanos*. Pamplona: EUNSA, 2023, p. 215–234.

Kim, Eusong. On the depth of fakeness. In: Filimowicz, Michael (ed.). *Deep fakes. Algorithms and society*. New York: Routledge, 2022, p. 50–70.

Kissinger, Henry A; Schmidt, Eric; Huttenlocher, Daniel. The Age of AI: And Our Human Future. New York: Little, Brown and Company, 2021.

Klobuchar, Amy. Klobuchar, Hawley, Coons, Collins introduce bipartisan legislation to ban the use of materially deceptive AI-generated content in elections. Available at: [https://www.klobuchar.senate.gov/public/index.cfm/2023/9/klobuchar-hawley-coons-collins-introduce-bipartisan-legislation-to-ban-the-use-of-materially-deceptive-ai-generated-content-in-elections#:~:text=This%20ban%20extends%20to%20a,generated%20content%20in%20news%20broadcasts]. Viewed: 11.03.2023.

Lage, Fernanda de Carvalho. *Manual de Inteligência Artificial no Direito Brasileiro*. 2. ed. Salvador: Jus Podivm, 2022.

Lardiez, Adrián. *La seducción de la extrema derecha: un ensayo sobre el comportamiento electoral y la psicología tras el voto populista*. Jaén: Editorial Libros.com, 2021.

Leal, Luziane de Figueiredo Simão; Moraes Filho, José Filomeno. Inteligência Artificial e democracia: os algoritmos podem influenciar uma campanha eleitoral? Uma análise do julgamento sobre o impulsionamento de propaganda eleitoral na Internet do Tribunal Superior Eleitoral. *Direitos Fundamentais e Justiça*, 13(41) (2019), p. 343–356.

Lewandowsky, Stephan; Smillie, Laura (coords.); Garcia, David; Hertwig, Ralph; Weatherall, Jim; Egidy, Stefanie; Robertson, Ronald E. (lead authors); O'connor, Cailin; Kozyreva, Anastasia; Lorenz-Spreen, Philipp; Blaschke, Yannic; Leiser, Mark (contribuiting authors). *Technology and democracy: Understanding the influence of online technologies on political behaviour and decision-making*. Brussels: European Commission, 2020.

Lins, Rodrigo Martiniano Ayres. Abuso de poder algorítmico: considerações iniciais. In: Lins, Rodrigo Martiniano Ayres; Castro, Kamile Moreira (eds.). (orgs.). *O futuro das eleições e as eleições do futuro*. Belo Horizonte: Fórum, 2023, p. 289–306.

Luna Corvera, Teresa González. Discriminación. In: Pereda, Carlos (ed.); Aragón Ribera, Álvaro; Delgado Parra, Concepción; Vega, Julieta Marcone; Leroux, Sergio Ortiz; Sermeño Quezada, Ángel (coords.). *Diccionario de injusticias*. Ciudad de México: Siglo XXI, UNAM, 2022, p. 238–243.

Macdonald, Hector. *Verdade. 13 motivos para duvidar de tudo que te dizem*. Rio de Janeiro: Objetiva, 2019.

Maranhão, Juliano. A inteligência artificial não é a vilã das eleições. **Folha de São Paulo**, 6 de febrero de 2024. Available at: [https://www1.folha.uol.com.br/opiniao/2024/02/inteligencia-artificial-nao-e-a-vila-das-eleicoes.shtml?utm_source=whatsapp&utm_medium=social&utm_campaign=compwa]. Viewed: 07.02.204.

Martínez García, Jesús Ignacio. Instituciones inteligentes. In: Solar Cayón, José Ignacio (ed.). *Dimensiones éticas y jurídicas de la inteligencia artificial en el marco del Estado de Derecho*. Madrid: Universidad de Alcalá, 2020, p. 23–54.

Massanari, Adrienne. Reddit's alt-right: toxic masculinity, free speech, and /r/The_Donald. In: Zimdars, Melissa; McLeod, Kembrew (eds.). *Fake news: understanding media and misinformation in the digital age*. Cambridge: The MIT Press, 2020, p. 179–189.

Mendes, Gilmar Ferreira. A problemática das fake news no Estado de Direito: uma análise do julgamento da ADPF 572. In: Branco, Paulo Gustavo Gonet; Fonseca, Reynaldo Soares da; Branco, Pedro Henrique de Moura Gonet; Velloso, João Carlos Banhos; Fonseca, Gabriel Campos Soares da (eds.). *Eleições e democracia na era digital*. Brasília: Almedina, 2022, p. 69–84.

Montagut, Marta; Willem, Cilia; Carrillo, Nereida. Influencers y desorden informativo alrededor de la 'ley trans'. In: Carratalá, Adolfo; Iranzo-Cabrera, María; López-García, Guillermo (eds.). *De la desinformación a la conspiración: política y comunicación ante un escenario híbrido*. Valencia: Tirant lo Blanch, 2023, p. 255–275.

Montero Caro, María Dolores. Sobre el control jurídico y democrático de la inteligencia artificial: herramientas y reflexiones acerca de la inserción incontrolada de mecanismos tecnológicos. Castellanos Claramunt, Jorge (Dir.). *Inteligencia Artificial y democracia: garantías límites constitucionales y perspectiva ética ante la transformación digital*. Barcelona: Atelier, 2023, p. 183–202.

Montilla Martos, José Antonio. Inteligencia artificial y derechos de participación política. In: Balaguer Callejón, Francisco; Cotino Hueso, Lorenzo (eds.). (Coords.). *Derecho Público de la Inteligencia Artificial*. Zaragoza: Fund. Miguel Giménez Abad, 2023, p. 151–180.

Moore, Martin. *Democracia hackeada: como a tecnologia desestabiliza os governos mundiais*. São Paulo: Editora Hábito, 2022.

Morales Romo, Beatriz. Colectivos vulnerables y desinformación: retos desde el ámbito educativo. Figueruelo Burrieza, Ángela (dir.). Martín Guardado, Sergio (coord.). In: *Desinformación, odio y polarización*. Vol. I. Navarra: Arazandi, 2023, p. 259–279.

Moreno, Frank. *Inteligencia artificial, su lado oscuro y el fin del principio*. Murcia: Editatum, 2023.

Morozov, Evgeny. *Capitalismo Big Tech: ¿Welfare o neofeudalismo digital?* Madrid: Enclave de Libros, 2018.

Mukherjee, Mitali. AI deepfakes, bad laws – and a big fat Indian election. **Reuters Institute,** 19 March 2024. Available at:[https://reutersinstitute.politics.ox.ac.uk/news/ai-deepfakes-bad-laws-and-big-fat-indian-election]. Viewed: 30.03.2024.

Mulhall, Joe. Drums In The Distance: Journeys Into the Global Far Right. London:Icon Books, 2021.

Muñoz Vela, José Manuel. *Retos, riesgos, responsabilidad y regulación de la inteligencia artificial. Un enfoque de seguridad física, lógica, moral y jurídica*. Navarra: Thomson Reuters, Aranzadi, 2022.

Murgia, Michela. *Instrucciones para convertirse en fascista*. Barcelona: Seix Barral, 2019.

National Conference of State Legislatures. Artificial Intelligence (AI) in Elections and Campaigns. 7 de febrero de 2024. Available at: [https://www.ncsl.org/elections-and-campaigns/artificial-intelligence-ai-in-elections-and-campaigns]. Viewed: 11.03.2024.

Navarrete, Federico. Racismo. In: Pereda, Carlos (ed.); Aragón Ribera, Álvaro; Delgado Parra, Concepción; Vega, Julieta Marcone; Leroux, Sergio Ortiz; Sermeño Quezada, Ángel (coords.). *Diccionario de injusticias*. Ciudad de México: Siglo XXI, UNAM, 2022, p. 668–674.

Neisser, Frederico; Mattiuzzo, Mariana. Estratégia do TSE de regrar uso de IA nas eleições não é isenta de riscos. **UOL,** 4 February 2024. Available at: [https://noticias.uol.com.br/opiniao/coluna/2024/04/02/inteligencia-artifical-eleicoes-2024.htm]. Viewed: 02.04.2024.

Pérez-Curiel, Concha; Rivas-De-Roca, Rubén; García-Gordillo, Mar. Narrativas populistas con alcance global: el caso de la retórica de Trump en las elecciones de Estados Unidos de 2020. In: Carratalá, Adolfo; Iranzo-Cabrera, María; López-García, Guillermo (eds.). *De la desinformación a la conspiración: política y comunicación ante un escenario híbrido*. Valencia: Tirant lo Blanch, 2023, p. 27–45.

Piras, Elisa. Inequality in the public sphere: epistemic injustice, discrimination and violence. In: Giusti, Serena; Piras, Elisa (eds.). *Democracy and fake news: information manipulation and post-truth politics*. New York: Routledge, 2021, p. 30–39.

Prado, Michele. *Tempestade ideológica. Bolsonarismo: a altright e o populismo iliberal no Brasil*. São Paulo: Todos Livros, 2023.

Prado, Paulo. La inteligencia artificial y su impacto en la política electoral. In: Vázquez-Barrio, Tamara; Salazar García, Idoia (eds.). *Inteligencia artificial, periodismo y democracia*. Valencia: Tirant lo Blanch, 2023, p. 80–226.

Rebollo Delgado, L. *Inteligencia artificial y derechos fundamentales*. Madrid: Dykinson, 2023.

REGULATION (EU) 2024/1689 OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL of 13 June 2024 laying down harmonised rules on artificial intelligence and amending Regulations (EC) No 300/2008, (EU) No 167/2013, (EU) No 168/2013, (EU) 2018/858, (EU) 2018/1139 and (EU) 2019/2144 and Directives 2014/90/EU, (EU) 2016/797 and (EU) 2020/1828 (Artificial Intelligence Act) https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=OJ:L_202401689

Revoredo, Tatiana. A lei de inteligência artificial da União Europeia. **Conjur,** 20 December 2023. Available at: [www.conjur.com.br/2023-dez-20/a-lei-de- inteligencia- artificial- da- uniao- europeia]. Viewed: 28.12.2023.

Rivero Ortega, Ricardo. *Derecho y Inteligencia Artificial: cuatro estudios*. Santiago: Olejnik, 2023.

Ryan-Mosley, Tate. Vuelta al mundo por las regulaciones de la IA en 2024. **MIT Technology Review,** 23 January 2024. Available at: [https://technologyreview.es/article/vuelta-al-mundo-por-las-regulaciones-de-la-ia-en-2024/]. Viewed: 30.01.2024.

Saccomani, Edda. Fascismo. In: Bobbio, Norberto; Matteucci, Nicola; Pasquino, Gianfranco (eds.). (coords.). *Dicionário da Política*. Vol. 1. Brasília: Universidade de Brasília, 2009.

Sánchez Muñoz, Óscar. *La regulación de las campañas electorales en la era digital. Desinformación y microsegmentación en las redes sociales con fines electorales*. Madrid, Valladolid: Centro de Estudios Políticos y Constitucionales, 2020.

Shoai, Andrés; López Molina, Adrián. Polarización e inteligencia artificial: una sistematización del conocimiento disponible. In: Vázquez-Barrio, Tamara; Salazar García, Idoia (eds.). *Inteligencia artificial, periodismo y democracia*. Valencia: Tirant lo Blanch, 2023, p. 247–264.

Simón Castellano, Pere. *La evaluación de impacto algorítmico em los derechos fundamentales*. Navarra: Aranzadi, 2023.

Sotero, José. Regulación de los algoritmos de recomendación: estudio de casos sobre los enfoques de la Unión Europea, Estados Unidos y República Popular China. In: Vázquez-Barrio, Tamara; Salazar García, Idoia (eds.). *Inteligencia artificial, periodismo y democracia*. Valencia: Tirant lo Blanch, 2023, p. 149–170.

Souza, Bruno Cezar Andrade de. *Dados pessoais: LGPD e as eleições*. Belo Horizonte: D'Plácido, 2022.

Stepanenko, Pedro. Prejuicio. In: Pereda, Carlos (ed.); Aragón Ribera, Álvaro; Delgado Parra, Concepción; Vega, Julieta Marcone; Leroux, Sergio Ortiz; Sermeño Quezada, Ángel (coords.). *Diccionario de injusticias*. Ciudad de México: Siglo XXI, UNAM, 2022, p. 652–656.

Tasioulas, John. First step towards an ethics of robots and artificial intelligence. *Journal of Practical Ethics*, 7(1) (2019), p. 61–95.

Taulli, Tom. Artificial Intelligence Basics: A Non-Technical Introduction. Germany: Apress, 2019.

Tong, Anna; Coster, Helen. Meet Ashley, the world's first AI-powered political campaign caller. Reuters, 16 de diciembre de 2016. Available at: [https://www.reuters.com/technology/meet-ashley-worlds-first-ai-powered-political-campaign-caller-2023-12-12/]. Viewed: 18.12.2023.

United States. Executive Order on the Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence. **The White House**, October 30, 2023. Available at: [https://www.whitehouse.gov/briefing-room/presidential-actions/2023/10/30/executive-order-on-the-safe-secure-and-trustworthy-development-and-use-of-artificial-intelligence/]. Viewed: 05.01.2024.

Urquijo, José Manuel. Xóchitl Gálvez, inteligencia artificial y el Uncanny Valley. **Expansión Política**, 22 de diciembre de 2023. Available at: [#Columna Invitada | Xóchitl Gálvez, la Inteligencia Artificial y el Uncanny Valley (expansion.mx)]. Viewed: 25.03.2024.

Vallespín, Fernando. *La sociedad de la intolerancia*. Barcelona: Galaxia Gutemberg, 2021.

Vázquez-Barrio, Tamara. Desafíos de la IA para ei periodismo y la salud democrática. In: Vázquez-Barrio, Tamara; Salazar García, Idoia (eds.). *Inteligencia artificial, periodismo y democracia*. Valencia: Tirant lo Blanch, 2023, p. 21–37.

Villamil, José Fernando. Discurso del odio en Internet y Derecho: algunos antecedentes. In: *Carvajal, Elizabeth. Discurso de odio, discriminación e información en internet*. Madrid: Reus, 2021, p. 33–45.

Zamperetti Júnior, Adriano. Impactos da legislação eleitoral e partidária brasileira no atual cenário político nacional. In: Ferraz, Roberto Victor Pereira Ribeiro (ed.). (Coord.). *O novo código eleitoral e as expectativas de sua aplicação: como a lei n° 14.211/2021 pode transformar as eleições brasileiras*. São Paulo: Tirant Lo Blanch, 2023, p. 21–42.

Index

AI and Electoral Campaigns, First Edition. Rafael Rubio Núñez, Frederico Franco Alvim and Vitor de Andrade Monteiro.
© 2026 John Wiley & Sons, Inc. Published 2026 by John Wiley & Sons, Inc.