palgrave pivot

# The Risk of Artificial Intelligence in Credit Ratings

## Exploring the Efficiency, Development and Impact

Daniel Cash
Nataliya Tkachenko

palgrave macmillan

# The Risk of Artificial Intelligence in Credit Ratings

Daniel Cash · Nataliya Tkachenko

# The Risk of Artificial Intelligence in Credit Ratings

## Exploring the Efficiency, Development and Impact

palgrave
macmillan

Daniel Cash
Law School
Aston University
Birmingham, UK

Nataliya Tkachenko
AI Centre of Excellence, CDAO
Lloyds Banking Group
London, UK

ISBN 978-3-031-95542-6 ISBN 978-3-031-95543-3 (eBook)
https://doi.org/10.1007/978-3-031-95543-3



This Palgrave Macmillan imprint is published by the registered company Springer Nature Switzerland AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

# Preface

This book, written early in 2025, exists at the very early stages of a journey that could define human civilisation. The adoption of artificial intelligence (AI) technologies across almost every sector is changing the way in which the world operates, how the world understands something ranging from the simple to the incredibly complex, and it is changing what the world believes is possible for its future. But, for the most part, it is all very *early*.

The nascent stage of development in any field makes it difficult to summarise, suggest or declare. With respect to credit risk assessment, it is very early indeed. This work is focused on the leading credit rating agencies and how they are adopting the AI technologies, but why? Of concern here is the amalgamation of several factors which we believe could be pertinent factors for regulators and legislators to consider. Those tasked with creating, maintaining and ultimately evolving the regulatory frameworks that sit around the leading credit rating agencies have a remarkable task. It is remarkable because they are inevitably behind the curve. The regulatory framework is focused on a sector that is synergistically attached to the fortunes of private debt and credit, which in the modern world is almost fundamental to human life. That fact alone, given its weight, affects all parties. Credit rating agencies understand their systemic position. Regulators understand that there are limits to actions they can take as a result. Investors understand their importance and the tools at their disposal, like the ratings of the agencies. Finally, issuers understand that, to signal their creditworthiness to a diversified investor base that is defined by the agent-principal relationship, they need the easy-to-understand and theoretically impartial third-party ratings of the credit rating agencies. A system within a system.

However, those binding constraints and factors have pitfalls. What if the credit rating agencies, after understanding that dynamic, decide they cannot lose their position if they transgress? What happens if, despite your acknowledged role being to provide services to investors irrespective of the fact it is issuers who pay for ratings (let's not forget, the second largest credit rating agencies was known as Moody's *Investors Service* for nearly all of its life since its foundation), a credit rating agency makes a summation that they can manipulate their methodologies and overlook key underlying information in order to please paying issuers at the expense of investors who they *know* will be harmed? There is very little punishment or deterrent available within the legal system that surrounds the credit rating agencies. Now imagine that this all happens, and a legal framework is erected which now includes a bar for liability and, subsequently, that legal framework acts. That action, eventually, results in the largest financial settlements ever recorded for the credit rating industry, at $1.3 billion and $864 million for the two largest credit rating agencies but, on reflection, those penalties are a fraction of the money made by transgressing. A systemic response that sends a clear signal, but perhaps not the one the system thinks it is sending.

The above, in a framed nutshell, is the credit rating agencies' involvement in the Global Financial Crisis. The story, delivered in rapid fashion, from start to finish. But for us, we are now in a post-Global Financial Crisis stage (ostensibly). Today, we have an even bigger, more influential, more resource-rich credit rating oligopoly that has not only survived its biggest test but thrived in spite of it. The real test is not for the credit rating agencies, but instead for the regulatory framework that was erected in response to the credit rating agencies' role in the Global Financial Crisis; can it withstand whatever comes next? Whatever comes next, if at all, will happen within an environment based upon a simple understanding: all of the relevant parties know, irrespective of whether they can articulate it or not, that this 'system' we have described for you must be maintained. It cannot be disregarded, or even altered. It is central to the movement of credit on a global scale, encompassing companies, countries, financial products and everything else in between. This is why every attempt to change the system since the Global Financial Crisis has failed.

At the core of the credit rating agencies' involvement in the Global Financial Crisis was an inability, willing or not, to get in control of the underlying data. Pooled data, hundreds/thousands/millions of data points, put together and aggregated. In the 2000s, the credit rating agencies adopted experimental mathematical models that they did not really understand (and perhaps, did not care to) to aggregate and understand the amalgamated output of the pools of data. Today, we argue, we are potentially at the start of the very same story. We will endeavour to show that, *potentially*, the tools that can replicate and even surpass the core underlying behavioural failures that led to the Global Financial Crisis are being built and implemented as we speak.

But the biggest caveat we must offer first is that if this is a journey, today is not even the first 'stop'. If this is a train journey, the passengers have only just learned of the train's existence. The passengers do not know the destination, nor how many stops there will be on the journey. They do not know how long the journey will be. They know the train exists, is in the station, and that imminently they are due to board. What the train will look like when they are on it, or how fast it will travel, they do not know. The passenger mentioned above is you. It is us. It is society. We suggest in this short-form book that this journey, from the credit rating agency perspective, cannot be a mystery. We have a clear example, from just twenty short years ago, of what happens when the credit rating agencies are allowed to control our mystery train journey. Rather, we argue here that it is time to find out the details of the journey, the constitution of the train, the relative comfort of the journey and, ultimately, the destination. It is for society to determine these aspects of the journey, not those driving the train. This book sends this warning and details why it should be heeded.

Birmingham, UK Daniel Cash

*Competing Interests* The authors have no competing interests to declare that are relevant to the content of this manuscript.

# Contents

# List of Figures

CHAPTER 1

# Introduction

**Abstract** The approach of the book is presented to the reader in this opening introductory chapter. This framing is supported by an analysis of the role of the credit rating agency in the larger financial architecture and what happened when the leading credit rating agencies transgressed in that role. The participation and centrality of the credit rating agencies in the Global Financial Crisis is presented as a critical case study for the reader to understand the potential impact of the present case study presented, that of the integration of AI into credit rating processes.

**Keywords** Credit rating agencies · Financial regulation · Artificial intelligence · Governance structures · Data integrity · Sovereign debt

Technology is evidently synonymous with the concept of credit rating. Given the length and breadth of the coverage of the modern leading credit rating agencies, it is unsurprising that novel technological innovations are important to the process of producing credit ratings. The modern leading credit rating agencies—of which we will count as S&P Global (S&P), Moody's and Fitch Ratings (Fitch)—have an enormous dominance in their sector; the most recent *Staff Report* of the US Securities and Exchange Commission (SEC) tallies this dominance in the form of the three agencies collectively rating more than 94% of all global ratings



D. Cash and N. Tkachenko, *The Risk of Artificial Intelligence in Credit Ratings*, https://doi.org/10.1007/978-3-031-95543-3_1

and, also, the top two agencies (S&P and Moody's) recording $3.3 billion and $2.9 billion in annual revenue respectively.[1] When this dominance is combined with how much the global markets use their products, the result is an industry that is constantly pushing to provide innovative technological-based products for their many and varied users.

However, that resultant need to continually push technological boundaries is not always just for the provision of services to their user base. Credit rating agencies also seek, quite understandably, to make their internal credit rating processes as efficient as possible to provide credit ratings that may be useful to their customers and users. This need, which has various facets underlying it, is at the heart of the credit rating agencies' recent push into the world of artificial intelligence (AI). Recently, the credit rating agencies have begun to acquire companies that have developed expertise in the world of AI, while also developing in-house solutions for their many processes.

However, *how* those AI technologies are being integrated into credit rating processes is critically important. The credit rating agencies have a checkered history when it comes to integrating novel technologies and understandings which affect their processes and the potential that AI exhibits leads to questions about whether credit rating agencies have the right constitution to deploy AI technologies internally in a responsible manner. For this book, this issue is at the heart of the matter. Making this issue central allows this book to ask a small number of highly relevant research questions:

- Q. How are credit rating agencies integrating AI technologies into their processes?
- Q. Do credit rating agencies have the right constitution to integrate AI technologies in a responsible manner?
- Q. What is the regulatory impact of credit rating agencies adopting AI technologies in their credit rating processes?

To answer these questions the book will make a number of 'stops' along its journey. **First**, the book will begin with assessing which particular AI technologies are of interest with regard to the credit rating processes. The leading credit rating agencies have chosen to deploy what is known as 'Generative AI' or 'GenAI' and in Chapter 2 we introduce this concept to you. The intention of the book is to provide you with an introductory and accessible insight into the world of GenAI, including its conceptual nature, how it has been used by various industries and field, and the implication of its usage.

Perhaps the most important aspect to understand with GenAI is its capability in relation to the requirements of a credit rating agency. GenAI is an umbrella term for various methods of understanding and collating vast amounts of data. Chapter 2 delves into these many different variants but, at its core, GenAI provides the possibility to instruct the technology to sift through copious amounts of data and synthesise it accordingly. Data, and potentially large amounts of 'unstructured' data is a key concern for modern business[2] but especially for credit rating agencies who must utilise various sources to best understand the creditworthiness of an issuer. Furthermore, the diversity of data now available[3] means that tools that can help synthesise and better understand that collective data pile are of obvious benefit to a credit rating agency. As a subset of 'Natural Language Processing' (NLP), GenAI 'uses sophisticated algorithms to comprehend and interpret unstructured data, showcasing not only the ability to process information, but also to autonomously generate content and contextually relevant content'.[4] When contrasted to the needs of a credit rating agency, who historically need to sift through data to generate not only credit ratings against their public methodologies but, more recently, have started to earn the majority of their income through providing 'ancillary services' (consultancy services) to the market, this ability to provide contextualised understandings of large amounts of unstructured data has the ability to the *perfect* tool.

However, the credit rating agency is not the only party in the equation. The investors who they serve, the issuers of debt who pay for the ratings, the regulators mandated with supervising the industry, and most importantly the public who pick up the tab when things go wrong are all critical stakeholders in this scenario. This is why the book seeks to ask *structural* questions in the hope of providing a *structural* picture that one may use to better understand developments in the credit rating field. Once the book considers how credit rating agencies are increasingly adopting AI into their credit rating processes in Chapter 3, the book will utilise this structural lens to expand the focus.

Chapter 3 is necessarily focused on understanding how the credit rating agencies are already implementing AI technologies internally and, furthermore, how they plan to implement AI technologies moving forward. The acquisitions that the Big Three credit rating agencies have been making, and especially the Big Two, are instructive as to how the two major agencies want to develop. It will also be worth investigating the different package offerings the credit rating agencies are offering to their respective clientele. As 'ancillary service' provision is now one of the dominant revenue streams, how those service offerings will evolve will also play into the task that regulators face in the near- and far-terms. There is an underlying issue however that will be covered next here in the introduction that paints these developments in a particular light. The leading credit rating agencies, and perhaps the sector moreover, are transgressive in nature. For the past 25 years in particular, the leading credit rating agencies have been penalised to the tune of billions of dollars for the largest infringements, and in a steady manner. The cause or reasoning for this transgressive nature is up for debate, with some suggesting that a marked change in culture over the years had led to the transgressive outcomes that have been witnessed.[5] However, it could also be because of the structural and conceptual realities of providing creditworthiness assessments. For example, a creditworthiness assessment is merely a prediction, which cannot be proven *ex ante*. This reality means that credit rating agencies are in a perilous business, which is why their history is littered with instances where they have battled against libel claims and liability claims relating to their outputs.[6] Yet, this last sentence does not tell the full story. The biggest issues relating to the credit rating agencies are not that they have issued an opinion or prediction which has turned out to be wrong. Rather, the most impactful infringements have been when the credit rating agencies *actively* take decisions that go against their stated role or purposefully and negatively impacts those they are theoretically supposed to serve. There is one major instance in particular which this introduction will cover shortly, but the point for now is that this *impact* is of the utmost importance to this book.

In Chapter 4 the book focuses on the 'Regulatory Challenge on the Horizon'. With a transgressive sector utilising complex technology that potentially removes or, at least significantly lessens human agency, the recipe for disaster is clear. To better understand whether the regulatory framework is currently suited to deal with that looming disaster is an important endeavour. By analysing both the credit rating-focused regulatory framework first, but then the backdrop of how financial regulation and regulation moreover is considering the issue of AI integration, the chapter will present a full picture of the issues facing the modern regulator. Key legislative developments like the EU's new world-leading AI

Act will be dissected to better understand the key underpinnings being advanced by the regulatory framework. This, potentially, allows us to better predict just how the regulatory framework will fare in the face of any identified threat emanating from the credit rating space. Upon that basis, the book can conclude in Chapter 6 with some focused and considered proposals for how the regulatory framework can best respond, proactively, to this impending threat.

## 1.1 A Transgressive Industry

The first obvious and necessary disclaimer is that not all members of the credit rating industry are 'transgressive', nor are the leading credit rating agencies predominantly transgressive. The credit rating agencies provide a critical function for the modern financial architecture and, for the most part, provide the service that the system asks of it. However, not a year goes by without the leading credit rating agencies either paying a fine to a major regulator (for example, the SEC in the US, the European Securities and Markets Authority [ESMA] in the EU, and to a lesser extent the Financial Conduct Authority [FCA] in the UK and the Securities and Exchange Board of India [SEBI]) or settling outside of court with a State's prosecutor, like the US Department of Justice (DoJ) for example. The most recent example was the Big Three collectively paying nearly $50 million in civil penalties to the US SEC over allegations that the agencies were not preventing staff from communicating official business through private mediums (colloquially known as the 'WhatsApp probes') which is in clear violation of various rules and regulations.[7]

However, the leading credit rating agencies have one ultimate stain in their history that we should not ignore. Their role in the Financial Crisis of 2007/08 was a central one. The post-Crisis US Senate investigations start their relevant sector of their 646-page Report with the following:

> Moody's Investors Service, Inc. (Moody's) and Standard & Poor's Financial Services LLC (S&P), the two largest credit rating agencies (CRAs) in the United States, issued the AAA ratings that made residential mortgage backed securities (RMBS) and collateralized debt obligations (CDOs) seem like safe investments, helped build an active market for those securities, and then, beginning in July 2007, downgraded the vast majority of those AAA ratings to junk status. The July mass downgrades sent the value of mortgage related securities plummeting, precipitated the collapse of the RMBS

> and CDO secondary markets, and perhaps more than any other single event triggered the financial crisis. In the months and years of buildup to the financial crisis, warnings about the massive problems in the mortgage industry were not adequately addressed within the ratings industry. By the time the rating agencies admitted their AAA ratings were inaccurate, it took the form of a massive ratings correction that was unprecedented in U.S. financial markets. The result was an economic earthquake from which the aftershocks continue today.[8]

The variety of players within the larger system that failed in 2007 makes it difficult to put one party at the centre of proceedings, but there is plenty of evidence available to put the credit rating agencies close. The reality is a simple one: the Financial Crisis would not have been possible without the failures of the leading credit rating agencies. The Senate investigation and a litany of commentary and academic research has proven this point in the time since but, it is not enough simply to categorise what happened as credit rating agency 'failures'.

The specifics of what happened are critically important, of course. At its heart, the credit rating agencies were applying their top ratings (AAA) to structured finance products that contained, essentially, toxic assets. Those assets were perceptively 'misunderstood' by the credit rating agencies and, when those underlying assets (residential mortgage products) started to fail, i.e. expected cash flows were not materialising, the AAA-rated products became worthless in an instant (exactly what is not supposed to happen with such highly rated products). As the report continues, 'traditionally, investments holding AAA ratings have had a less than 1% probability of incurring defaults'.[9] Yet, one of the identified failures of the credit rating agencies was there (a) construction and (b) implementation of credit rating models.

It is worth pausing here momentarily to offer some definitions. The headline product at the heart of the Financial Crisis was the Residential Mortgage-Backed Security (RMBS). An RMBS is a focused subclass of product derived from a broader concept, known as a Collateralised Debt Obligation, or CDO. A CDO can be a complex concept to understand.[10] However, at its core it is a product that pools together credit obligations (think mortgages, credit cards, car finance, etc.) and synthesises that collective flow of capital for investors to then invest in as a collective. To allow investors of particular categories—categorised by their level of risk-averseness—to invest in the stream of collectivised capital, the credit rating agencies apply their credit ratings to particular slices of the stream, or 'tranches' (French for slices) so that the portion of the stream that is least likely, or would be last to experience the impact of losses generated at the source—say, if people starting reneging on their mortgage payments—is rated AAA, and so on. This then allowed the investors who are usually regulatorily constrained to invest only in what may be independently deemed 'safe investments' to partake in the CDO scheme. This process is known as 'subordination' within a larger system of 'credit enhancement'.

Continuing on, credit rating agencies had been rating what are known as 'structured finance' products since the mid-1970s. IOSCO (the International Organisation of Securities Commissions) notes that the credit rating agencies started with mortgage-backed securities in the mid-1970s (meaning it was either residential mortgages or commercial mortgages that were the underlying asset for the 'stream') but quickly also moved into other CDOs, like credit card receivables, auto loans, student loans and equipment leases.[11] Essentially, if it is credit-based and a person or company will be paying a steady stream of money towards the payment of the full loan, it can be used within a CDO structure. How a credit rating agency would rate such a structure is different from how they would rate a corporate issuer, for example, which would rely on public information about the issuer and also private information provided by the issuer itself. For a CDO, the focus for the credit rating agency is on the underlying collateral. The agency will want to consider the past performance of the collateral as just one criterion, as well as the specific structure that the arranger is choosing to use for the CDO. IOSCO describe what 'typically' happens in this process:

> A sponsor typically initiates the RMBS rating process by sending a CRA data on a pool of loans (e.g., principal amount, geographic location, borrower's credit history, loan-to-value ratio, and type of loan: first lien, second lien, primary residence, secondary residence) and the proposed capital structure of the trust. The CRA assigns a lead analyst who will be responsible for analyzing the loan pool and proposed capital structure of the trust and formulating ratings recommendations for a rating committee. The analyst first develops predictions based on models and other factors as to how many of the loans in the collateral pool would be expected to default under stresses of varying severity. This analysis also includes assumptions as to how much principal would be recovered after a defaulted loan is foreclosed.[12]

The purpose of receiving all this information and then applying particular models is to allow the credit rating agency to apply a rating from their rating scale that then signals the risk to investors. A higher tranche will be shielded from losses more than a lower tranche. However, that higher tranche will not receive as much in terms of repayment than those lower down the scale, to account for the risk those lower are facing. IOSCO conclude by confirming that a credit rating agency will only be paid by the arranger of the CDO if the rating is provided at the end, though sometimes a 'breakup fee' was given if the arranger decided to go to another competing credit rating agency once the process had started.

There has been an abundance of literature on the conflicts of interest that occurs in this process. For example, the headline-grabbing conflict of interest is perhaps the concept of 'ratings shopping'. This concept exists because most CDO arrangers want two ratings to signal their creditworthiness to the marketplace, but as the credit rating industry is dominated by an *oligopoly* and not a *duopoly* this means a member of the Big Three credit rating agencies is always under pressure from another member of the oligopoly; if Moody's do not provide the required rating to a CDO provider, they would go to Fitch, or S&P, and so on. This led to a distinct race-to-the-bottom where credit rating agencies were competing on the basis of who was more amenable to their client's needs who, in this case, were the leading investment banks like Citigroup, Merrill Lynch and UBS.[13] Research has concluded that, often, the credit rating agencies would reach further than their models allowed—via the Credit Rating Committee stage of the rating process which is behind closed doors and represents the 'black box' and final stage of rating proceedings—due to the pressures of 'rating shopping'.[14]

However, even though the credit rating agencies wilfully disregarded their credit rating models, the full picture reveals that, even with their models, they were willing to compromise. To understand credit risk, the credit rating agencies must deploy a variety of mathematical theories and structures. One such mathematical structure that was required when CDOs began to emerge was related to the need to understand the intercorrelation between random variables. In a mathematical nutshell this is known as a 'copula', a term introduced in 1959 by Abe Sklar.[15] If the aim is to understand and model the *dependence* of random variables between one another, then it is clear why this would be important for the analysis of a pool of, say, residential mortgages. While all the mortgages may be similar in concept—a loan to purchase a home, for example—what may affect whether that loan is paid back in full and on time is very different from loan (mortgage) to loan (mortgage). What affects my ability to repay my mortgage may be different to what affect yours, and so on.

MacKenzie helpfully discusses how, by the mid-1990s, derivative teams in leading banks across the US were engaging with sophisticated mathematical models to make sense of the 'bucket of risks' they were putting together within various CDOs.[16] One such mathematical structure that emerged was known as the 'Gaussian Copula' model, named after the German mathematician Carl Friedrich Gauss but pioneered by David X Li in the mid-to-late 1990s.[17] While a copula as devised by Sklar is interested in joining together the distribution functions of *uniformly distributed variables*, the Gaussian version instead yields a *multivariate* joint distribution function which, on the face of it, would be better suited to the CDO universe. While banks were already utilising this type of mathematical structure, credit rating agencies were still engaging in a practice called 'notching', which essentially describes when an agency adjusts credit ratings for specific debts or obligations within an overarching issuer; for example, a corporate issuer may be rated AAA, but a specific debt that it issues may be rated AA-. However, this approach was clearly not appropriate for the complexity of the CDO so, as MacKenzie notes in an interview with a former rating analyst, 'notching was not a proper correlation method'. Therefore, the agencies all adopted the Gaussian Copula Formula to better assist them with rating CDOs.[18] In 2001 S&P developed its 'CDO Evaluator', in 2003 Fitch launched its 'Vector' system, and in 2004 Moody's launched its 'CDOROM', all based on the Gaussian Copula Formula developed by Li.

In transitioning to using the Gaussian Copula Formula instead of notching, the credit rating agencies began to 'estimate based on the judgments of experienced ratings staff' what the benchmarked level of correlation needed to be.[19] It was important that the new modelling was not too dissimilar to the old practice of 'notching' to maintain a perception of consistency for the market. However, there was a lack of data in the default databases being used, so it quickly became oligopolistic practice to set the correlation rate at 0.3, which was chosen 'partly to maintain consistency with the previous notching scheme'.[20] This issue of setting a correlation rate to the analysis is critical to the CDO process. As the US Senate investigation describes:

> Correlative risk measures the likelihood of multiple negative events happening simultaneously, such as the likelihood of RMBS assets defaulting together. It examines the likelihood, for example, that two houses in the same neighborhood will default together compared to two houses in different states. If the neighborhood houses are more likely to default together, they would have a higher correlation risk than the houses in different states.[21]

The issue was that credit rating agencies could not provide the necessary ratings to CDOs—that relied upon the subordinated structure to entice different categories of investors—if the correlation was too high, because then the necessary AAA ratings for the upper tier would not have been plausible or justifiable. Even a small increase in the correlation figure to, say, 0.5 instead of 0.3 would have left the CDOs 'economically unviable'.[22] At the same time as S&P and Moody's made this change, their market share and dominance for rating CDOs skyrocketed, reducing Fitch's share (as the third and smaller member of the credit rating oligopoly) from 65% before 2004 to just 15% in 2006. This dominance translated quickly into financial reward, as S&P's structured finance team would eventually account for 49% of the firm's rating revenue while Moody's gross revenues from RMBS and CDO ratings would triple in just five years; the Senate investigation concluded that, collectively, the Big Three's revenues doubled from nearly $3 billion in 2002 to over $6 billion in 2007. The Senate investigation, with unique power to subpoena records, found a chain of evidence which confirmed a. that the credit rating agencies were aware that the risks in the underlying assets were not being factored into models, and b. that deploying resources into amending those models and maintaining those models was not the priority. As one former S&P leader testified: 'the MRBS group enjoyed the largest ratings market share among the three major rating agencies (often 92% or better), and improving the model would not add to S&P's revenues'. This and other aspects lead MacKenzie to a completely valid question:

> Given that – and given the dependence of rating agencies on fees earned from the issuers of the securities, and the possibility of those issuers 'ratings shopping' (choosing the agencies that offer the more favourable ratings) – should we interpret the choice of a correlation of 0.3 or thereabouts as strategic behaviour guided by anticipated fee income?[23]

It is seemingly impolite to provide the answer that is blindingly obvious. Yet, the *awareness* within the credit rating agencies does not aid their cause. Rogue analysts within the agencies *did* model (in their own time) the right correlations and performed 'drilldowns' into the underlying data rather than using the Gaussian Copula Formula as a veil behind which 'strategic' correlative ratio figures could be set, and they all concluded that the figure ought to have been closer to 0.8 or 0.9 to reflect the underlying risk.[24]

The rest of the story is described in copious amounts of journal articles, books, media articles and in the Senate investigation. Mass downgrades of AAA-rated CDOs shocked the market and resulted in an era-defining Financial Crisis which, arguably, 17 years later at the time of writing, the world is still feeling the effects of. Yet, the development of 'the formula that felled Wall St'[25] reveals a central issue that guides this book. One could be forgiven for thinking that the Gaussian Copula Formula was heralded as industry-leading and defined a new evolution in credit analysis, but that is not the case. It had many detractors before the credit rating agencies injected it right into the middle of their systemically-critical rating processes.[26] This adds further weight to MacKenzie's question of the underlying motive of the actions of the credit rating agencies.

This is what this book exists to do. The bluster employed by the credit rating agencies in the early 2000s, under the cover of efficiency and applicability, mislead the market. Internally, a conflicted provider of crucial market signals was transforming their approach to increase market share and revenues. Today, we potentially are at the beginning of the same story. Credit rating agencies, as we shall see in this book, are telling the markets that integrated AI will make internal processes more efficient, make the credit rating agencies' output more applicable, and that market participants will be better off for it. However, depending on how AI *is* integrated by the credit rating agencies, we will potentially have agencies, who are exposed to exactly the same conflicts as they were before, delegating key tasks within credit analysis to computerised technology that *they* set the parameters for. The environment may not be the same, but the factors are beginning to pile up in favour of deploying extreme caution. But, as we shall see and at least from a regulatory perspective, that is not necessarily the case.

In the Preface to this book, we discussed how this issue can be thought of as a train journey and, today, we are very much still in the station of origin. Yet, like any train journey where the public are passengers, we must know the stops on the way, the regulations governing the journey and the providers of the service, and ultimately the destination *before* we set off. The events of the 2000s serve as a stark reminder that allowing events to unfold and then seeking to take action is the ultimate failure. This book serves as an early warning system that the events of the 2000s and the underlying factors that contributed to the systemic and generation failure must not be repeated in this new world of artificial intelligence.

## Notes

1. US Securities and Exchange Commission, *Staff Report on Nationally Recognized Statistical Rating Organizations* (2025) https://www.sec.gov/files/jan-2025-ocr-staff-report.pdf.
2. Gianluca Elia, Elisabetta Raguseo, Gianluca Solazzo, and Federico Pigni, 'Strategic business value from big data analytics: An empirical analysis of the mediating effects of value creation mechanisms' (2022) 59 Information & Management 8.
3. Alessio Faccia, Luigi P.L. Cavaliere, Pythagoras Petratos, and Narcisa R Mosteanu, 'Unstructured over Structured, Big Data Analytics and Applications in Accounting and Management' (2022) Proceedings of the 2022 6th International Conference on Cloud and Big Data Computing.
4. Enrique Cano-Marin, 'The transformative potential of Generative Artificial Intelligence (GenAI) in business: a text mining analysis on innovation data sources' (2024) 55 ESIC Market Economics and Business Journal 2 3.
5. Daniel Cash, *A Modern Credit Rating Agency: The Story of Moody's* (Routledge 2023).
6. See: Marc Flandreau, Norbet Gaillard, and Frank Packer, 'To err is human: US rating agencies and the interwar foreign government debt crisis' (2011) 15 European Review of Economic History 495–538; Marc Flandreau and Joanna K Sławatyniec, 'Understanding rating addiction: US courts and the origins of rating agencies' regulatory licence (1900–1940)' (2013) 20 Financial History Review 3 237–57; Marc Flandreau and Gabriel G Mesevage, 'The separation of information and lending and the rise of the rating agencies in the USA (1841–1907)' (2014) 62 Scandinavian Economic History Review 3 213–242; Marc Flandreau and Gabriel G Mesevage, 'The

Untold History of Transparency: Mercantile Agencies, the Law, and the Lawyers (1851–1916)' (2014) 15 Enterprise and Society 2 213–51.

7. Jasper Ward and Chris Prentice, 'Six rating agencies to pay over $49mln over recordkeeping failures, SEC says' (2024) Reuters (Sept 3) https://www.reuters.com/business/finance/us-sec-charges-six-credit-rating-agencies-with-recordkeeping-failures-2024-09-03/.
8. United States Senate Permanent Subcommittee on Investigations, Committee on Homeland Security and Governmental Affairs, *Wall Street and the Financial Crisis: Anatomy of a Financial Collapse* (2011) https://www.hsgac.senate.gov/wp-content/uploads/imo/media/doc/PSI%20REPORT%20-%20Wall%20Street%20&%20the%20Financial%20Crisis-Anatomy%20of%20a%20Financial%20Collapse%20(FINAL%205-10-11).pdf.
9. Ibid.
10. For a skilled discussion of what a CDO is and how it may be rated, see: Ingo Fender and John Kiff, 'CDO rating methodology: Some thoughts on model risk and its implications' (2004) BIS (Nov) https://www.bis.org/publ/work163.pdf.
11. IOSCO, *The Role of Credit Rating Agencies in Structured Finance Markets* (2008) https://www.iosco.org/library/pubdocs/pdf/ioscopd270.pdf.
12. Ibid. 6.
13. See John M Griffin and Dragon Y Tang, 'Rating Shopping or Catering? An examination of the Response to Competitive Pressure for CDO Credit Ratings' (2012) SEC (May 25) https://www.sec.gov/divisions/riskfin/seminar/tang061412.pdf.
14. Ibid. 31.
15. Gery Geenens, '(Re)-Reading Sklar (1959)—A Personal View on Sklar's Theorem' (2024) 12 Mathematics 3 380.
16. Donald MacKenzie, 'The Credit Crisis as a Problem in the Sociology of Knowledge' (2011) 116 American Journal of Sociology 6 1803.
17. David X Li, 'On Default Correlation: A Copula Function Approach' (2000) RiskMetrics (Apr) https://cyrusfarivar.com/docs/li.defaultcorrelation.pdf.
18. MacKenzie (n 16).
19. Ibid. 1805.

20. Ibid.
21. US Senate (n 8) 292.
22. MacKenzie (n 16) 1813.
23. Ibid.
24. Ibid.
25. Sam Jones, 'The Formula that Felled Wall St' (2009) Financial Times (Apr 24) https://www.ft.com/content/912d85e8-2d75-11de-9eba-00144feabdc0.
26. Lisa Pollack, 'The formula that Wall Street never believed in' (2012) Financial Times (June 15) https://www.ft.com/content/a19cdaf1-f5db-3abc-823a-8671e8169c5a.

CHAPTER 2

# Generative AI: Concept, Applications and Implications

**Abstract** This chapter introduces the concept of Generative AI, the main form of artificial intelligence being integrated by the credit rating agencies. In this chapter, the book presents a conceptual understanding for the uninitiated so that later analyses have the required foundation. The chapter examines the development of Generative Artificial Intelligence as well as its applicability to a range of sectors that have been actively integrating the technology.

**Keywords** Generative AI · Classification systems · Machine learning · Synthetic data · Bias · Decision-making automation

## 2.1 Introduction

For centuries, human civilisation has relied on classification and rating systems to bring order to chaos. From the earliest forms of creditworthiness assessments in medieval trade networks to modern credit ratings by financial giants like Moody's and Standard & Poor's, societies have depended on structured methodologies to evaluate trust, risk and quality. These systems were built on painstakingly collected data, historical trends and expert judgement; imperfect yet foundational in industries ranging from finance and healthcare to entertainment and manufacturing.



D. Cash and N. Tkachenko, *The Risk of Artificial Intelligence in Credit Ratings*, https://doi.org/10.1007/978-3-031-95543-3_2

However, we now stand at the precipice of a new technological era, where classification and rating no longer depend solely on human expertise. Instead, they are increasingly shaped by the hand of Generative Artificial Intelligence (GenAI), a technology that has fundamentally altered the way data is processed, analysed and leveraged for decision-making. Unlike traditional machine learning models, which categorise and predict based on established patterns, Generative AI takes a more audacious step: it creates. Text, images, music, video and even synthetic financial records—all can be generated by AI models, often with astonishing realism.[1]

This evolution has not occurred in a vacuum. Long before the rise of AI, industries developed intricate methodologies to classify and rate data. Financial institutions, for instance, have long used actuarial science and statistical models to assess creditworthiness and risk.[2] The healthcare sector relies on classification frameworks such as the International Classification of Diseases (ICD),[3] while the entertainment industry has crafted star ratings, review aggregators and content filtering mechanisms to guide consumer preferences. Governments and regulatory bodies have established standards for safety, energy efficiency and environmental compliance, all through rigorous classification systems grounded in empirical data.

These pre-existing methods, while robust, had their limitations. They often struggled with scalability, required extensive human oversight, and were susceptible to bias or subjectivity. Enter Generative AI in 2025. With its ability to synthesise vast amounts of data and produce novel insights, AI promised to revolutionise these domains, offering faster, more scalable and ostensibly more objective classification and rating solutions.[4] The introduction of sophisticated models such as Generative Adversarial Networks (GANs),[5] Variational Autoencoders (VAEs)[6] and Transformer-based architectures like GPT[7] and BERT[8] marked a turning point. These models do not merely replicate existing data, they generate entirely new content based on learned patterns. In the financial sector, for example, AI-driven models are being deployed to detect fraudulent transactions, generate synthetic financial datasets for risk assessment and even automate aspects of investment strategy development. In healthcare, Generative AI is assisting in diagnostic processes by generating synthetic medical images for training purposes and improving the accuracy of medical classification systems.

Yet, the integration of Generative AI into these long-established rating and classification methods is not without its challenges. One critical question emerges: is AI always the right tool for the task? While it excels in efficiency, AI does not inherently possess human judgement. A generative model might predict credit risk based on historical loan performance, but can it account for socioeconomic changes or ethical considerations in underscoring practices? Similarly, while AI-generated content can mimic human writing or artwork with remarkable precision, it often lacks true creativity, context or the nuanced intent that human creators bring to their work.

Beyond concerns of appropriateness, the widespread adoption of Generative AI brings significant legal, ethical, reputational and financial implications.[9] Legally, AI-generated classifications and ratings raise issues of liability and accountability. If an AI-driven financial model erroneously assigns a low credit score to a creditworthy individual, who is responsible? Similarly, in healthcare, misclassification by an AI-driven system could lead to incorrect diagnoses, impacting patient outcomes and potentially resulting in malpractice suits.

Ethically, bias remains a formidable challenge.[10] AI models are only as good as the data they are trained on, and if historical data reflects biases, whether in lending, hiring or medical treatment, AI models will inevitably perpetuate them. This has already been observed in AI-driven credit scoring systems, where marginalised communities sometimes face systemic disadvantages due to historical disparities in financial data.

Reputationally, businesses that over-rely on AI for classification and rating face risks when AI-driven models make flawed decisions. Companies that rely on AI-generated financial ratings, for instance, might suffer credibility loss if these ratings prove unreliable. Similarly, media platforms using AI to classify and recommend content have faced backlash when their algorithms inadvertently promote misinformation or bias.

From a financial perspective, while AI offers cost savings and efficiency, errors in AI-generated classifications can be costly. Incorrect financial risk assessments can lead to bad loans, regulatory fines and economic instability. Similarly, flawed AI-driven credit rating models could lead to unjust loan denials or predatory lending practices, exacerbating economic inequalities.

Despite these challenges, the momentum behind Generative AI remains strong. Organisations across industries are racing to integrate

AI-driven classification and rating systems, seeking the competitive advantages of automation, speed and predictive accuracy. The question, then, is not whether Generative AI will reshape classification and rating methodologies—it is already happening. Rather, the challenge lies in ensuring that these AI-driven systems are implemented responsibly, with the right methods for the right purpose.

As we explore the historical evolution of classification and rating systems across various industries, this chapter will delve deeper into how AI is reshaping these long-standing methodologies. We will examine the points of transition from traditional models to AI-driven systems, highlighting both the opportunities and pitfalls of this technological transformation. Whether in finance, healthcare, entertainment or regulatory governance, the impact of AI is profound, and understanding its implications is crucial for ensuring that it serves as a tool for progress rather than a source of unintended harm.

## 2.2 Demystifying Generative AI

Generative AI represents a profound departure from traditional machine learning approaches that rely on classification and regression. Instead of simply predicting outcomes based on existing data, Generative AI learns the intrinsic structure of a dataset and uses that knowledge to create new, statistically plausible data points. This ability to generate novel information has made it a transformative tool across multiple industries. But how exactly does it work? Understanding the fundamental principles of Generative AI requires delving into three core concepts: probability distributions, latent space representation and adversarial training.

### *2.2.1 Probability Distributions: The Foundation of Generative AI*

The concept of probability distributions has been central to Generative AI since its earliest implementations. In statistical modelling, a probability distribution defines the likelihood of different outcomes within a dataset. Traditional AI models use probability distributions to classify data points, but generative models take this a step further by learning the entire distribution of data and generating new instances that fit within that learned framework.

The application of probability distributions to Generative AI became prominent in the 1990s with the emergence of Bayesian networks and early probabilistic graphical models. These models laid the groundwork for more advanced generative approaches by providing structured ways to infer missing data and generate synthetic samples. Today, Generative AI leverages sophisticated probabilistic techniques such as Gaussian mixture models and autoregressive flows to create new data points in diverse fields, from finance to medical imaging.

### 2.2.2 *Latent Space Representation: Encoding the Essence of Data*

Another crucial concept in Generative AI is latent space representation. This approach involves encoding complex data into a lower-dimensional space, a form of abstract numerical representation where key patterns and features are distilled into compact variables. Once encoded, the model can decode these representations to generate new data instances that maintain the essential characteristics of the original dataset.

The notion of latent space representation became widely known with the development of autoencoders in the early 2000s. Variational Autoencoders (VAEs), introduced in 2013, further refined this concept by enabling the generation of high-quality synthetic data through a learned probabilistic distribution of latent variables. This technique has been particularly influential in image generation, allowing AI to create realistic photos, facial reconstructions and even synthetic medical scans that aid in training diagnostic models.

### 2.2.3 *Adversarial Training: The Power of GANs*

Perhaps the most groundbreaking development in Generative AI came with the introduction of Generative Adversarial Networks (GANs) in 2014 by Ian Goodfellow and his colleagues. GANs represent a paradigm shift in machine learning by employing two competing neural networks (the generator and the discriminator) in an adversarial process. The generator creates synthetic data samples, while the discriminator evaluates them and determines whether they are real or fake. Through iterative training, the generator becomes increasingly skilled at producing data indistinguishable from real-world samples.

The impact of GANs has been profound. In the financial sector, GANs are used to generate synthetic financial data that simulate market conditions, allowing institutions to conduct stress testing without relying on historical data alone. In entertainment, GANs create hyper-realistic deepfake videos and AI-generated art. Meanwhile, in healthcare, GANs facilitate drug discovery by generating molecular structures with potential pharmaceutical applications.

### *2.2.4 Transformer-Based Models: Encoders, Decoders!*

A more recent and equally transformative innovation in Generative AI is the development of transformer-based models. First introduced in the seminal 2017 paper *Attention Is All You Need* by Vaswani et al., transformers revolutionised natural language processing (NLP) by introducing self-attention mechanisms that allow models to process entire sequences of text simultaneously rather than sequentially.[11]

Transformers utilise two primary components: encoders and decoders. Encoders analyse input data and generate contextualised embeddings, capturing relationships between words or data points across long sequences. Decoders then use this encoded information to generate output data, whether in the form of text, code or synthetic records. Transformers have since been implemented in various AI-driven applications. BERT (Bidirectional Encoder Representations from Transformers), introduced by Google in 2018, excels at understanding contextual meaning in text, making it invaluable for search engines, chatbots and document classification. GPT (Generative Pre-trained Transformer) models, such as OpenAI's GPT-3 and GPT-4, take this a step further by generating human-like text, enabling applications such as content creation, automated financial reporting and even AI-driven programming assistance.

The impact of transformers extends beyond NLP. In finance, transformer models are used to analyse vast amounts of unstructured data, automate fraud detection and generate synthetic trading scenarios. In healthcare, they facilitate automated medical note generation and enhance predictive diagnostic models. Their application in creative industries includes AI-generated art, music and film scripts.

### *2.2.5 Is Generative AI Always the Right Tool?*

While Generative AI offers remarkable capabilities, it is not always the most appropriate tool for every task. In scenarios requiring precise and deterministic outcomes (such as legal decision-making or regulatory compliance) traditional rule-based AI systems may be preferable. Additionally, Generative AI poses risks related to bias, misinformation and ethical considerations. Because these models learn from existing data, any biases present in training datasets can be perpetuated in the generated outputs. This has significant implications in sectors like hiring, credit scoring and criminal justice.

Despite these challenges, the adoption of Generative AI continues to accelerate, reshaping industries and redefining the boundaries of what machines can create. As we progress through this book, we will explore how each industry has integrated Generative AI into its classification and rating systems, the benefits it offers and the complexities it introduces.

## 2.3 The Evolution of Classification and Rating Systems Across Industries and the Shift to Generative AI

Classification and rating systems have played an essential role in shaping economies, businesses and societies. These methodologies have provided frameworks for evaluating creditworthiness, product quality, healthcare standards and investment risks, among many other critical functions. Over time, these traditional classification and rating techniques have evolved, incorporating modern technologies like deterministic machine learning and, more recently, transformer-based unsupervised AI. The transition to AI-driven systems has introduced efficiencies, predictive capabilities and automation that are revolutionising industries, as well as new hidden risks.

### *2.3.1 Insurance Industry*

The insurance industry has long relied on actuarial models to classify policyholders and assess risk. These models, developed in the late nineteenth and early twentieth centuries, were based on historical loss data, statistical probability and demographic analysis. By the mid-twentieth century, insurers began incorporating more sophisticated mathematical models to determine premium pricing and risk classification.

The transition to AI in insurance classifications gained momentum in the 2010s.[12] Machine learning models were employed to detect fraudulent claims and optimise underwriting processes. The integration of GenAI in the late 2010s allowed insurers to generate synthetic claims data, automate policy recommendations and refine risk prediction models.

AI-driven chatbots and digital assistants now streamline customer interactions, while deep learning models predict catastrophic losses with greater accuracy.[13]

### *2.3.2 Hospitality and Tourism*

The hospitality industry has utilised classification systems since the early twentieth century, when hotels and restaurants were first ranked using star-rating methodologies. The Michelin Guide (introduced in 1900) and AAA's Diamond Ratings (established in 1937) became benchmarks for quality. Airlines later adopted Skytrax ratings to classify services based on customer feedback.[14]

AI-driven rating models in hospitality emerged in the early 2010s with recommendation algorithms on platforms like TripAdvisor and Google Reviews.[15] By the mid-2010s, GenAI-enabled systems were generating automated travel itineraries and analysing real-time customer sentiment. The hospitality industry now leverages AI-generated reviews, dynamic pricing models and predictive service enhancements based on AI-classified customer preferences.[16]

### *2.3.3 Healthcare and Medical Industry*

Medical classification systems date back to the eighteenth century with early disease taxonomies.[17] The development of the International Classification of Diseases (ICD) by the World Health Organization (WHO) in the 1940s laid the foundation for modern medical classification. Hospital ratings, patient risk assessments and drug efficacy classifications emerged in the late twentieth century. AI's influence in healthcare classification began in the 2000s,[18] with machine learning models aiding in disease diagnosis and medical image classification.[19] The adoption of AI accelerated in the 2010s, with machine-generated synthetic medical records enabling enhanced training for diagnostic models.[20] Today, GenAI promises potential for personalised treatment plans, automated medical report generation and real-time health monitoring through AI-powered wearable devices.[21]

### 2.3.4 Consumer Goods and Retail

Product classifications and quality ratings have been essential to consumer markets for over a century.[22] Systems such as the Energy Star rating (introduced in 1992) and safety classifications by Underwriters Laboratories (UL) have guided consumers in making informed purchasing decisions. Online product review systems gained prominence with the rise of e-commerce in the late 1990s. Machine learning-enhanced classification models in retail emerged in the 2010s,[23] with platforms like Amazon leveraging AI to personalise recommendations.[24] The adoption of GenAI started in the late 2010s, with AI-generated product descriptions, chatbot-assisted shopping and automated supply chain optimisation.[25] Today, GenAI drives virtual shopping assistants and hyper-personalised marketing strategies based on AI-classified customer behaviours.[26]

### 2.3.5 Education and Certification Bodies

Academic institutions have classified universities and certification programmes for over a century. The Times Higher Education rankings and QS World University Rankings, established in the late twentieth century, became authoritative standards for classifying educational institutions. The transition to AI-driven classifications in education began in the early 2010s, with adaptive learning platforms using machine learning for personalised instruction. By the late 2010s, GenAI-enabled systems automated grading, generated educational content and provided AI-driven tutoring. Today, AI chatbots help students navigate academic programmes, while Generative AI tailors course recommendations based on student performance.[27]

### 2.3.6 Real Estate and Property Market

Real estate classifications have existed since the early twentieth century, with property valuation systems based on location, size and market demand. Agencies like Zillow and Realtor.com introduced AI-driven home valuation models in the 2000s.[28] The real estate sector began integrating AI in the 2010s for predictive market analysis and automated mortgage underwriting.[29] GenAI adoption accelerated in the 2020s, with AI-generated virtual property tours, automated lease generation and predictive property classification based on buyer preferences.[30]

### 2.3.7 Entertainment and Media

Film and music ratings have been integral to the entertainment industry since the early twentieth century. The MPAA film rating system (established in 1968) and Billboard music charts have long classified content based on audience suitability and popularity.[31] AI-driven media classifications emerged in the 2010s, with streaming platforms like Netflix and Spotify using machine learning for content recommendations.[32] The transition to GenAI in entertainment began in the late 2010s, with AI-generated scripts, automated video summaries and AI-driven content classification.

### 2.3.8 Manufacturing and Industry Standards

Industrial classifications have been essential for safety and quality control since the late nineteenth century.[33] Organisations like ISO (founded in 1947) established industry-wide standards for manufacturing classifications.[34] AI-driven classification models in manufacturing gained traction in the 2010s, with machine learning optimising production lines.[35] GenAI adoption in the 2020s introduced AI-generated designs, automated material classification and predictive maintenance.[36]

### 2.3.9 Government and Public Sector

Governments have long classified countries based on economic performance, environmental policies and safety indices.[37,38] Credit rating agencies and global governance institutions have ranked nations for decades. AI's role in governance classification emerged in the 2010s, with predictive analytics used predominantly for economic forecasting.[39] GenAI adoption in the 2020s started automating policy analysis, risk classification and AI-generated regulatory compliance reports (predominantly with the 'human-in-the-loop' components to ensure human oversight).[40,41,42]

## 2.4 Generative AI in Finance: A Deep Dive

Financial markets have always been driven by data. From the earliest stock exchanges to modern digital trading platforms, the ability to interpret market trends, forecast asset prices and optimise investment strategies has been the key to financial success.[43] Traditionally, this relied on human expertise, traders and analysts, poring over charts, economic indicators and company earnings reports to determine where to place their bets. The advent of artificial intelligence has fundamentally changed this dynamic. AI models, with their capacity to process vast datasets at speeds far beyond human capability, have become indispensable tools for market analysis and prediction.[44]

At the core of AI-driven market analysis lies the ability to detect patterns within financial data. These models analyse historical price movements, macroeconomic trends and even non-traditional data sources such as social media sentiment to forecast potential market behaviour. While traditional AI models have been effective at identifying statistical correlations, Generative AI takes this a step further by simulating alternative market scenarios.[45] By training on historical market data, generative models can construct synthetic financial environments, allowing analysts and traders to test strategies in simulated conditions before deploying them in real-time markets. This capacity to generate potential futures makes AI an invaluable asset for risk management and investment decision-making.[46]

Nowhere is the impact of AI more evident than in the world of algorithmic and high-frequency trading (HFT).[47][48] High-frequency trading has long been dominated by advanced algorithms capable of executing thousands, if not millions, of trades per second based on microsecond changes in market conditions. These trading strategies rely on predictive modelling, which AI has enhanced significantly. Generative AI enables the continuous evolution of trading algorithms, allowing them to adapt to changing market conditions in real-time. With AI-driven high-frequency trading, models can identify minute inefficiencies in financial markets that may only exist for fractions of a second. By leveraging vast amounts of market data, these AI systems make rapid trading decisions, ensuring profits from short-lived opportunities. However, the increased speed and complexity of AI-generated trading strategies have raised concerns about market stability. Flash crashes, sudden, deep market drops caused by automated trading, highlight the risks of entrusting too much control to autonomous AI models. While AI can enhance trading efficiency, financial regulators and institutions must balance automation with oversight to mitigate systemic risks.[49]

Beyond trading, AI plays an essential role in fraud detection and risk assessment.[50] Financial fraud has evolved alongside technology, with criminals using increasingly sophisticated methods to bypass traditional security measures. AI-driven fraud detection systems combat this by learning from past fraudulent activities and continuously improving their ability to identify anomalies in transactions. Unlike rule-based fraud detection systems that rely on predefined red flags, Generative AI models detect previously unseen patterns of fraudulent behaviour. AI-driven fraud detection systems analyse transactional data in real time, flagging suspicious activities based on deviations from established behavioural norms. For example, if a bank customer who typically makes small, localised transactions suddenly initiates a large international wire transfer, the AI system may classify this as a potentially fraudulent transaction. In credit card fraud prevention, AI models track spending behaviours, detecting unauthorised activity with greater precision than traditional methods. As financial fraud grows more sophisticated, so too must the AI models designed to combat it, necessitating continuous improvements in detection algorithms.[51]

A crucial but often overlooked application of AI in finance is the generation of synthetic data. Financial institutions operate under strict regulatory requirements to protect customer data, making it difficult to share sensitive datasets for research and model training.[52] Generative AI addresses this challenge by creating synthetic financial data that mirrors real-world datasets while maintaining privacy compliance. These synthetic datasets enable financial institutions to train AI models without compromising customer confidentiality. Synthetic data is particularly valuable in stress testing and scenario analysis. By generating artificial financial environments, institutions can assess how different economic conditions (such as recessions, inflationary periods or interest rate hikes) might impact portfolios. This allows for more robust risk management and strategic planning. Furthermore, AI-generated data facilitates collaboration between financial organisations and researchers, promoting innovation while adhering to strict data privacy regulations.

AI is also redefining credit scoring and loan approvals. Traditional credit scoring models rely heavily on historical financial data, such as credit history and income statements. However, these models often fail to assess individuals who lack formal financial records, disproportionately affecting underbanked populations. AI-driven alternative credit scoring models address this gap by analysing a wider range of data sources, including transaction history, spending patterns and even social behaviour. Generative AI can enhance credit scoring by building predictive models that assess an individual's financial behaviour beyond traditional metrics. For instance, a borrower with limited credit history but a consistent pattern of bill payments and responsible spending may be considered creditworthy by an AI-driven system, whereas a traditional credit model might classify them as high risk. These alternative scoring methods help make financial services more inclusive, providing opportunities for individuals who might otherwise be denied credit based on outdated evaluation criteria.

Finally, the role of AI in regulatory compliance and risk management cannot be overstated. Financial institutions must navigate an increasingly complex web of regulations designed to prevent money laundering, fraud and financial misconduct.[53] Compliance teams have historically relied on manual processes to review transactions and ensure adherence to anti-money laundering (AML) policies. AI automates these processes, making compliance efforts more efficient and less prone to human error. By analysing vast amounts of financial data, AI-driven compliance tools identify suspicious transactions and flag them for further review. Generative AI models simulate potential compliance risks, allowing institutions to anticipate regulatory challenges before they arise. AI's ability to monitor transactions in real time significantly reduces the workload of compliance teams while improving detection accuracy. However, as AI-driven compliance systems become more prevalent, financial regulators must ensure that these models operate transparently and do not inadvertently reinforce biases in financial decision-making.

The integration of Generative AI into financial markets has transformed everything from trading strategies to fraud detection, credit scoring and regulatory compliance.[54] While the benefits of AI-driven financial analysis are undeniable, challenges remain in ensuring the ethical and responsible deployment of these technologies. The increasing reliance on AI demands greater oversight to prevent unintended consequences, from algorithm-driven market instability to biased lending decisions. As AI continues to evolve, financial institutions must strike a balance between automation and accountability, ensuring that technology serves to enhance, rather than replace, responsible financial decision-making.

## 2.5 When Things Go Wrong: Risks, Failures and the Future of AI in Finance

Financial markets, for all their complexity and sophistication, have been historically vulnerable to catastrophic failures. The past century has seen multiple financial crises, many of which were exacerbated by poor risk management, flawed decision-making and a lack of regulatory oversight. The Global Financial Crisis of 2008 serves as one of the most glaring examples of how systemic failures in risk assessment and market transparency can lead to widespread economic disaster.[55] Mispriced mortgage-backed securities and an overreliance on credit rating agencies contributed to the unchecked expansion of high-risk lending. Had AI technologies been as advanced then as they are today, predictive models could have identified irregularities in mortgage securities, detected early warning signals in financial institutions and forecasted the unsustainable debt structures that eventually led to collapse. However, while AI offers the promise of enhanced oversight, it also introduces new risks and vulnerabilities that must be carefully managed.

One of the most dramatic manifestations of AI-related financial instability comes in the form of flash crashes: sudden, extreme market downturns caused by algorithmic trading. The 2010 Flash Crash remains a cautionary tale.[56] Within minutes, the Dow Jones Industrial Average plummeted nearly 1,000 points, erasing billions of dollars in market value before a rapid recovery. Investigations revealed that algorithmic trading models, which rely on high-frequency transactions, had created a feedback loop of panic selling. This incident exposed the fragility of AI-driven trading systems and underscored the necessity for safeguards, including circuit breakers and human oversight, to prevent similar crises in the future.

Generative AI, despite its transformative potential, is not immune to financial miscalculations and failures. In 2022, a hedge fund that heavily relied on AI-driven trading models suffered multimillion-dollar losses due to misinterpretation of market signals.[57] The AI system, trained on historical trading data, failed to account for unprecedented geopolitical events and economic policy shifts, leading to inaccurate predictions and costly missteps. This case illustrated a fundamental challenge of AI in finance: its reliance on past data to predict future behaviours. While AI can uncover intricate patterns and correlations, it remains limited in its ability to account for black swan events, which are rare, unpredictable occurrences that can significantly disrupt markets.

Beyond trading, AI-based credit assessment tools have also encountered serious pitfalls. One of the promises of AI in financial services has been its ability to provide alternative credit scoring models, particularly for underbanked populations who lack traditional credit histories. However, real-world applications have shown that AI-driven credit scoring can inadvertently reinforce historical biases. In some cases, AI models have systematically discriminated against certain demographic groups due to the data they were trained on. For instance, if historical lending practices were biased against minority applicants, AI models trained on that data may continue to deny loans to those same groups, perpetuating financial exclusion rather than rectifying it. Regulators and financial institutions are now grappling with how to ensure fairness and transparency in AI-based lending decisions.

The regulatory landscape surrounding AI in finance remains a critical point of discussion. While AI has demonstrated significant efficiencies in automating compliance and fraud detection, it also raises concerns about transparency and accountability. Financial regulators are now tasked with balancing innovation with consumer protection. Striking the right balance is crucial as overregulation could stifle AI advancements, while a lack of oversight could lead to unchecked risks and unethical financial practices. The challenge is particularly pronounced in anti-money laundering (AML) compliance, where AI models must be rigorously tested to ensure they do not inadvertently overlook sophisticated money laundering schemes.

Another emerging risk lies in the overreliance on AI for financial decision-making. While AI models can process vast amounts of data and execute trades with unparalleled speed, they do not possess human intuition or ethical reasoning. The delegation of financial decision-making to AI introduces the risk of blind trust in machine-generated insights, which, if flawed, can lead to devastating consequences. AI-driven financial systems operate within a framework of probability and optimisation, but they do not always account for nuanced factors such as market sentiment, geopolitical risks or sudden regulatory changes. As AI continues to integrate into financial decision-making, maintaining a balance between automated efficiency and human oversight will be essential.

Despite these challenges, AI continues to advance, and financial institutions are becoming increasingly reliant on its capabilities. Looking ahead, the next phase of AI in finance will likely focus on enhancing its interpretability and reliability. Researchers and financial technology firms are developing explainable AI (XAI) models,[58] which aim to provide greater transparency into how AI models arrive at their predictions and decisions. This shift towards explainability will be critical for regulatory compliance and for building trust in AI-driven financial systems.

The future of AI in finance presents both opportunities and risks. While AI can enhance efficiency, improve fraud detection and expand financial inclusion, it must be implemented responsibly. The key to future success will be striking the right balance: leveraging AI's analytical power while maintaining human oversight to ensure fairness, accountability and resilience. As AI continues to evolve, financial institutions must refine their methodologies, ensuring that AI serves as a tool for informed decision-making rather than an unchecked force that dictates financial outcomes. The future of AI in finance will be shaped not just by technological innovation, but by the ethical and regulatory frameworks that govern its use. Whether the industry is headed in the right direction will ultimately depend on the willingness of financial institutions, regulators and AI developers to prioritise responsible deployment and risk mitigation alongside efficiency and profitability.

## 2.6 Conclusion

The integration of Generative AI into finance marks a transformative period for the industry, offering unprecedented capabilities in market analysis, trading, fraud detection, credit scoring and regulatory compliance. As financial institutions increasingly rely on AI-driven solutions, they must remain aware of both the opportunities and the risks that come with this technology. While AI enables institutions to analyse vast datasets, optimise strategies and identify fraudulent activities more efficiently than ever before, it also introduces new vulnerabilities, including potential biases, market instability and ethical concerns.

Historically, financial markets have suffered from crises exacerbated by human error, lack of oversight and misjudged risk management. AI offers a path towards mitigating these issues by providing data-driven insights that can detect irregularities before they spiral into systemic failures. However, the risks associated with overreliance on AI, as demonstrated by flash crashes and AI-driven market losses, highlight the importance of balancing automation with human oversight. Generative AI's ability to simulate market conditions and stress-test portfolios is a powerful tool, but it must be applied responsibly.

Furthermore, credit scoring and lending processes must be carefully refined to prevent AI from perpetuating existing inequalities in financial access. The promise of AI lies in its ability to provide more inclusive, accurate and efficient financial services, yet achieving this goal requires constant monitoring to ensure fairness and transparency.

The regulatory landscape will play a crucial role in shaping the future of AI in finance. Policymakers and financial regulators must establish clear guidelines to prevent the misuse of AI, ensuring that ethical considerations and data security are prioritised alongside efficiency and profitability. The development of explainable AI (XAI) and increased transparency in AI decision-making will be essential for maintaining trust in automated financial systems.

As AI technology continues to advance, the financial sector must adapt to harness its potential while mitigating risks. The future of AI in finance depends not only on innovation but also on the ability of institutions, regulators and AI developers to create responsible, well-governed systems. If properly managed, AI can be a force for stability, efficiency and inclusivity in financial markets. However, success will depend on a careful balance between technological progress, regulatory oversight and ethical considerations to ensure that AI serves as a tool for progress rather than a source of new financial instability.

## Notes

1. Awel Jo, 'The promise and peril of generative AI' (2023) 614 *Nature* 1 214–216.
2. Michael Chui, Eric Hazan, Roger Roberts, Alex Singla, Kate Smaje, Alex Sukharevsky, Lareina Yee, and Rodney Zemmel, 'The economic potential of Generative AI' (2023) McKinsey (June 14) https://www.mckinsey.com/capabilities/mckinsey-digital/our-insights/the-economic-potential-of-generative-ai-the-next-productivity-frontier#introduction.
3. World Health Organisation, *International Standardisation Classification of Diseases and Related Health Problems* (2025) https://www.who.int/standards/classifications/classification-of-diseases.
4. Henrik Sætra, 'Generative AI: Here to stay, but for good?' (2023) 75 Technology in Society 102372.
5. Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, Yoshua Bengio, 'Generative adversarial networks' (2020) 63 Communications of the ACM 11 139–144.

6. Diederik P. Kingma and Max Welling, ‘Auto-encoding variational Bayes’ (2013) Arvix (Dec 10).
7. Nassim Dehouche, ‘Plagiarism in the age of massive Generative Pre-trained Transformers (GPT-3)’ (2021) 21 Ethics in Science and Environmental Politics 17–23.
8. Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova, ‘BERT: Pre-training of deep bidirectional transformers for language understanding’ (2019) Proceedings of the 2019 conference of the North American chapter of the association for computational linguistics: human language technologies, volume 1 (long and short papers) 4171–4186.
9. Dehouche (n 7).
10. Krishnaram Kenthapadi, Himabindu Lakkaraju, and Nazneen Rajani, ‘Generative AI meets responsible AI: Practical challenges and opportunities’ (2023) Proceedings of the 29th ACM SIGKDD conference on knowledge discovery and data mining 5805–5806.
11. Ashish Vaswani, Noam Shazeer, Niki Parmar, Jacob Uszkoreit, Llion Jones, Aidan Gomez, Lukasz Kaiser, and Ilia Polosukhin, ‘Attention is all you need’ (2017) Advances in neural information processing systems 30.
12. Christopher Blier-Wong, Helene Cossette, Luc Lamontagne, and Etienne Marceau, ‘Machine learning in P&C insurance: A review for pricing and reserving’ (2020) 9 Risks 1 4.
13. Rahul Sahai, Ali Al-Ataby, Sulaf Assi, Manoj Jayabalan, Panagiotis Liatsis, Chong Loy, Abdullah Al-Hamid, Sahar Al-Sudani, Maitham Alamran, and Hoshang Kolivand, ‘Insurance risk prediction using machine learning’ (2022) The international conference on data science and emerging technologies 419–433.
14. Eunhye Park, Bongsug Chae, and Junehee Kwon, ‘Toward understanding the topical structure of hospitality literature: Applying machine learning and traditional statistics’ (2018) 30 International Journal of Contemporary Hospitality Management 11 3386–3411.
15. Aniekan Essien and Godwin Chukwukelu ‘Deep learning in hospitality and tourism: a research framework agenda for future research’ (2022) 34 International Journal of Contemporary Hospitality Management 12.4480–4515.
16. Zohreh Doborjeh, Nigel Hemmington, Maryam Doborjeh, and Nikola Kasabov, ‘Artificial intelligence: a systematic review of methods and applications in hospitality and tourism’ (2022) 34

International Journal of Contemporary Hospitality Management 3 1154–1176.
17. Hafsa Habehh and Suril Gohel' Machine learning in healthcare' (2021) 22 Current genomics 4 291–300.
18. Fei Jiang, Yong Jiang, Hui Zhi, Yi Dong, Hao Li, Sufeng Ma, Yilong Wang, Qiang Dong, Haipeng Shen, and Yongjun Wang, 'Artificial intelligence in healthcare: past, present and future' (2017) 2 Stroke and vascular neurology 4.
19. Alison Callahan and Nigam Shah, 'Machine learning in healthcare' (2017) Key advances in clinical informatics 279–291.
20. Antti Väänänen, Keijo Haataja, Katri Vehviläinen-Julkunen, and Pekka Toivanen, 'AI in healthcare: A narrative review' (2021) 10 F1000Research 6.
21. Ahmad Chaddad, Jihao Peng, Jian Xu, and Ahmad Bouridane, 'Survey of explainable AI techniques in healthcare' (2023) 23 Sensors 634.
22. Caroline Heins, 'Artificial intelligence in retail–a systematic literature review' (2023) 25 Foresight 2 264–286.
23. Lanlan Cao, 'Artificial intelligence in retail: applications and value creation logics' (2021) 49 International Journal of Retail & Distribution Management 7 958–976.
24. Hsin-Pin Fu, Tien-Hsiang Chang, Sheng-Wei Lin, Ying-Hua Teng, and Ying-Zi Huang, 'Evaluation and adoption of artificial intelligence in the retail industry' (2023) 51 International Journal of Retail & Distribution Management 6 773–790.
25. Abhikit Guha, Dhruv Grewal, Praveen Kopalle, Michael Haenlein, Matthew Schneider, Hyunsoek Jung, Rida Moustafa, Dinesh Hegde, and Gary Hawkins, 'How artificial intelligence will affect the future of retailing' (2021) 97 Journal of Retailing 1 28–41.
26. Helen Murdoch, 'Choosing a problem—when is Artificial Intelligence appropriate for the retail industry?' 7 Expert Systems 1 42–49.
27. Gokcen Olcay and Melih Bulu, 'Is measuring the knowledge creation of universities possible?: A review of university rankings' (2017) 123 Technological Forecasting and Social Change 153–160.
28. Stephane Tekouabou, Stefan Gherghina, Eric Kameni, Youssef Filali, and Khalil Idrissi Gartoumi, 'AI-based on machine learning methods for urban real estate prediction: A systematic survey'

(2024) 31 Archives of Computational Methods in Engineering 2 1079–1095.
29. Karthigeyan Kuppan, Deepak Acharya, and Divya B ‘Foundational AI in Insurance and Real Estate: A Survey of Applications, Challenges, and Future Directions’ (2024) 12 IEEE Access.
30. Marcelo Cajias, ‘Artificial intelligence and real estate-not just an evolution, a real game changer!’ (2021) 39 Journal of Property Investment & Finance 1 15–18.
31. Giri Hallur, Sandeep Prabhu, and Avinash Aslekar, ‘Entertainment in the era of AI, big data & IoT’ (2021) Digital Entertainment: The Next Evolution in Service Sector 87–109.
32. Richard Lachman and Michael Joffe, ‘Applications of artificial intelligence in media and entertainment’ (2021) Analyzing future applications of AI, sensors, and robotics in society 201–220.
33. Ruigi Li, Sha Wei, and Jia Li, ‘Study on the application framework and standardization demands of AI in intelligent manufacturing’ (2019) 2019 International Conference on Artificial Intelligence and Advanced Manufacturing (AIAM) 604–607.
34. Yan Lu, Paul Witherell, and Albert Jones, ‘Standard connections for IIoT empowered smart manufacturing’ (2020) 26 Manufacturing letters 17–20.
35. Wenting Chen, Caihua Liu, Fei Xing, Guochao Peng, and Xi Yang, ‘Establishment of a maturity model to assess the development of industrial AI in smart manufacturing’ (2022) 35 Journal of Enterprise Information Management 3 701–728.
36. Julia Turovets and Konstantin Vishnevskiy, ‘Standardization in smart manufacturing: evaluation from a supply-side perspective’ (2022) Intelligent Systems in Digital Transformation: Theory and Applications 191–218.
37. Slava Mikhaylov, Marc Esteve, and Averill Campion, ‘Artificial intelligence for the public sector: opportunities and challenges of cross-sector collaboration’ (2018) 376 Philosophical Transactions of the Royal Society A: Mathematical, physical and engineering sciences 2128.
38. Weslei De Sousa, Elis de Melo, Paulo Bermejo, Rafael Farias, and Adalmir Gomes, ‘How and where is artificial intelligence in the public sector going? A literature review and research agenda’ (2019) 36 Government Information Quarterly 4 101392.

39. Colin Van Noordt and Gianluca Misuraca, 'Artificial intelligence for the public sector: results of landscaping the use of AI in government across the European Union' (2022) 39 Government information quarterly 3 101714.
40. Maciej Kuziemski and Gianluca Misuraca, 'AI governance in the public sector: Three tales from the frontiers of automated decision-making in democratic settings' (2020) 44 Telecommunications policy 6 101976.
41. Deniz Susar and Vincenzo Aquaro, 'Artificial intelligence: Opportunities and challenges for the public sector' (2019) Proceedings of the 12th international conference on theory and practice of electronic governance 418–426.
42. Yu-Che Chen, Michael Ahn, and Yi-Fan Wang, 'Artificial intelligence and public values: value impacts and governance in the public sector' (2023) 15 Sustainability 6 4796.
43. Longbing Cao, 'AI in finance: challenges, techniques, and opportunities' (2022) 55 ACM Computing Surveys (CSUR) 3 1–38.
44. Tom Lin, 'Artificial intelligence, finance, and the law' (2019) 88 Fordham Law Review 531.
45. Xiaolin Zheng, Mengying Zhu, Qibing Li, Chaochao Chen, and Yanchao Tan, 'FinBrain: when finance meets AI 2.0' (2019) 20 Frontiers of Information Technology & Electronic Engineering 7 914–924.
46. Jun Shao, Zhukun Lou, Chong Wang, Jinye Mao, and Ailin Ye, 'The impact of artificial intelligence (AI) finance on financing constraints of non-SOE firms in emerging markets' (2022) 17 International Journal of Emerging Markets 4 930–944.
47. Carlo Milana and Arvind Ashta, 'Artificial intelligence techniques in finance and financial markets: a survey of the literature' (2021) 30 Strategic Change 3 189–209.
48. Sean Cao, Wei Jiang, and Lijun Lei, 'Applied AI for finance and accounting: Alternative data and opportunities' (2024) 84 Pacific-Basin Finance Journal 102307.
49. Paulo Giudici, 'Fintech risk management: A research challenge for artificial intelligence in finance' (2018) 1 Frontiers in Artificial Intelligence 1.
50. Carsten Maple, Lukasz Szpruch, Gregory Epiphaniou, Kalina Staykova, Simran Singh, William Penwarden, Yisi Wen, Zijian

Wang, Jagdish Hariharan, and Pavle Avramovic, 'The AI revolution: opportunities and challenges for the finance sector' (2023) Arvix.

51. Jonas Christensen, 'AI in financial services' in Prashant Natarajan, Bob Rogers, Edward Dixon, Jonas Christensen, Kirk Borne, Leland Wilkinson, and Shantha Mohan, *Demystifying AI for the Enterprise* (2021 Productivity Press).
52. Paulo Giudici and Emanuela Raffinetti, S'AFE Artificial Intelligence in finance' (2023) 56 Finance Research Letters 104088.
53. Henri Arslanian and Fabrice Fischer, *The future of finance: The impact of FinTech, AI, and crypto on financial services* (2019 Springer).
54. Nydia Remolina, 'Generative AI in finance: Risks and potential solutions' (2024) 1 Law, Ethics and Technology 1 1.
55. Ibid.
56. Xuemei Li, Alexander Sigov, Leonid Ratkin, Leonid Ivanov, and Ling Li, 'Artificial intelligence applications in finance: a survey' (2023) 10 Journal of Management Analytics 4 676–692.
57. Nicholas Megaw, Madison Darbyshire and James Fintanella-Khan, 'How the investment world is trying to navigate geopolitics?' (2024) Financial Times (Jul 5) https://www.ft.com/content/23ce295d-bf65-47fd-bebd-808b5a7bcab5.
58. Patrick Weber, K Valerie Carl, and Oliver Hinz, 'Applications of explainable artificial intelligence in finance—a systematic review of finance, information systems, and computer science literature' (2024) 74 Management Review Quarterly 2 867–907.

CHAPTER 3

# The Growing Adoption of AI within the World of Credit Ratings

**Abstract** The integration of the technology into credit rating processes is the focus of this chapter, which is split between how the credit rating agencies are utilising AI and how researchers and market participants are also making use of AI to better understanding creditworthiness. Development within the credit rating agency sector is also presented, in terms of the services they are starting to offer, Mergers and Acquisitions that are taking place, and how the internal processes of the agencies are being adapted to the world of Generative AI.

**Keywords** Credit risk modelling · Black-box systems · Institutional investors · Artificial neural networks · Decision trees · Conflicts of Interest

## 3.1 Introduction

The clear potential and applicability of AI in all its forms is increasingly of interest to the leading credit rating agencies. However, there are underlying issues which may become more pertinent as time goes on and AI becomes even more prevalent in the field of financial services and investor support. In this chapter we will introduce some of the key developments and conceptual foundations for how AI can aid with the



D. Cash and N. Tkachenko, *The Risk of Artificial Intelligence in Credit Ratings*, https://doi.org/10.1007/978-3-031-95543-3_3

modelling of risk. This is important because, in the literature, there is a large imbalance. There are various sources of analysis on how AI can aid with risk modelling, but nothing on how credit rating agencies may be integrating AI into their processes. This becomes a conflict when a large amount of the literature aims to subject credit rating agency outputs to scrutiny against what can be modelled using AI. What is missing from the discussion is that the credit rating agency model for providing ratings is not solely about risk modelling in the quantitative sense. The credit rating process, complete with key qualitative interventions and, most crucially, the final credit rating committee stage—the 'black box' of the process[1]—does not lend itself well to this concept of AI alone being able to accurately reflect the creditworthiness of an issuer of debt.

To better understand this delineation, the chapter will begin with a journey through the relevant literature on credit risk modelling and how AI may synthesise that process. This provides us all with a starting point of understanding what can be done with the technology. This understanding however will provide us with two revelations. The first is that technology, when deployed in an analytical manner, can be very effective. The second revelation however is that this quest for effectiveness and accuracy is not entirely the point for the leading credit rating agencies. Their role in the modern financial architecture is a dual role with each endpoint perhaps operating in conflict with the other. For example, it is on a theoretical scale that one may believe that the credit rating agencies exist to provide an assessment of creditworthiness that is wholly accurate and efficient—nothing more than 100% accuracy should be the aim. That aim—and not the realisation of that aim, as 100% accuracy of an opinion is impossible—is perhaps a reasonable request of such an important component of the economy. Yet, research has confirmed for a while now that credit rating agencies are massively conflicted and that the conflicts identified do influence the ratings they provide.[2] The consequence of that research is the suggestion that credit rating agencies are perhaps predominantly concerned with facilitating the movement of capital (and being paid as a result), rather than purely being occupied with the quest for rating accuracy. It could be something else. Yet, this understanding, when married to the claims of the research focused on AI's application to risk modelling, suggests that aiming for rating accuracy is not the main priority of the leading credit rating agencies. If it is not, then asking what the main priority is becomes an exceptionally important question. The answer to that question becomes even more important when the suggestion from credit rating agencies is that integrating AI will allow them to have *less* of a hands-on approach to the data they need to analyse to make credit ratings. We saw in the introduction what happens when the credit rating agencies lose control of the underlying data.

While not the focus for this book, it will also be discussed briefly that ESG Rating Agencies are also increasingly integrating AI into their processes,[3] which only further increases the systemic risk that sector regulators are facing at the moment. With the same dualised role potentially applying—facilitation over accuracy—to the ESG rating sector as it does to the credit rating sector, the ensuing systemic risk is clear. However, it may not be. It could well be that the regulatory infrastructure is designed around the facilitation-focused understanding of the role of the rating agencies (both ESG and Credit), rather than the accuracy-focused role of the rating agencies. If that is the case, the literature is not helpful to the regulators because it focuses on something conceptually academic and not practical. It may well be the case that the literature is only helpful to investors and financial institutions seeking to mimic the credit rating agencies, to better plan and predict for prospective credit ratings. Either way, it will be helpful to ascertain the utility of the research.

## 3.2 Using AI to Aid with Risk Modelling

Before those abstract questions are even raised, a better understanding of how AI can aid with credit risk modelling is needed. Because the inner workings of a credit rating agency (and what happens in the final credit rating committee phase) cannot really be known fully, researchers have for a long time tried to replicate the process to see how closely aligned the credit rating agencies are with what the researchers can produce. The sentiment, of course, is what the researchers are producing is the 'pure' credit rating, free from any sort of bias or conflict. That quest, for the researchers at least, has become easier with the introduction of AI and the increased computational potential of the modern age.

This research on credit risk modelling can be of great help to investors and financial institutions. Surkan and Singleton discussed in 1990 how 'models of bond ratings, therefore, are of great interest to investors, who want to anticipate the rating given a change in company circumstances, and to financial managers, who seek to predict the rating (and accompanying interest rate) of a potential issue'.[4] Dutta and Shekhar also validate this understanding, confirming that 'developing a model for rating bonds is important as it enables a financial institution to independently evaluate the default risk of its bond investments'.[5]

It is worth contextualising these understandings, however. It is true to say that issuers and investors all conduct their own credit risk analysis independent from the credit rating agencies. The more sophisticated investors—like the large-scale institutional investors—will often have credit risk analysis capabilities that dwarf that of the leading credit rating agencies. Credit risk, according to research, is the predominant concern within institutional investors[6] which arguably makes perfect sense. It is often the case in the literature that one will find announcements like 'ratings *save* investors the costs of doing their own analysis to evaluate risk prospects. These costs have been increasing with international diversification and the rising complexity of securities. In addition to helping understand the risks and uncertainties of investments, the independent benchmark of default risk that credit ratings provide makes it easier for investors to compare different potential investments'.[7] It is true that institutional investors will utilise the ratings to benchmark investments and compare investment opportunities, but unfortunately such pronouncements may lead one to conclude that institutional investors (or any sophisticated investor) *outsources* their credit risk analysis to the credit rating agencies and this could not be further from the truth. With institutional investment increasingly converging into a more concentrated sector, even just cursory analyses of the financial statements of the largest institutional investors show that they invest considerable amounts of their resources into their own credit risk analysis infrastructures.[8] With even just some of the conflicts of interest we discussed in the introduction within the leading credit rating agencies, outsourcing one's credit risk analysis requirements to such conflicted entities would be grossly inappropriate and likely put the managers and directors of the investors in breach of their fiduciary duties. Therefore, there is a great need for investors and other market participants (including issuers) to conduct their own analysis a. for their own decision-making processes and b. to better understand and potentially anticipate the actions of the credit rating agencies.

The research that would underpin that required analysis has a long history in the literature. One of the early pioneers was William H Beaver who, in 1966, used a univariate approach to financial ratios in order to predict the rates of failure within corporations.[9] In this pioneering study however, and with limited technological or computational power, Beaver observed that the quest to assess credit risk in any manner is beset with bias and limitations (in his case, it was that the sample selection was affected by rates of corporate rescue among other criteria). Yet, he concluded simply that 'the evidence indicates that ratio analysis can be useful in the prediction of failure for at least five years before failure'. This conclusion gives us the foundation that predicting creditworthiness via rates of failure was a. possible via the data available at the time and b. that the potential 'time-horizon' for making such conclusions was up to five years prior to a potential event of default. Two years later, Edward I Altman deployed a multivariate approach to the problem.[10] Altman begins by describing that prior to available quantitative data analysis, 'agencies were established to supply a quantitative type of information assessing the creditworthiness of particular merchants'—here is referring to the forefathers of the credit rating sector, the likes of Tappan's Mercantile Agency or Bradstreet's Bradstreet Company. He then identifies that studies of the time concluded that various ratios were selected as being most prominent—those that measure profitability, or liquidity, or solvency—but that none had emerged as universally-recognised as being predominantly important. Because of this issue, Altman deployed a Multiple Discriminant Analysis (MDA) to the range of financial ratios available. Altman describes the MDA process as:

> MDA is a statistical technique used to classify an observation into one of several *a priori* groupings dependent upon the observation's individual characteristics. It is used primarily to classify and/or make predictions in problems where the dependent variable appears in qualitative form, e.g., male or female, bankrupt or non-bankrupt. Therefore, the first step is to establish explicit group classifications. The number of original groups can be two or more. After the groups are established, data are collected for the objects in the groups; MDA then attempts to derive a linear combination of these characteristics which 'best' discriminates between the groups. If a particular object, for instance a corporation, has characteristics (financial ratios) which can be quantified for all of the companies in the analysis, the MDA determines a set of discriminant coefficients. When these coefficients are applied to the actual ratio, a basis for classification into one of the mutually exclusive groupings exists. The MDA technique has the advantage of considering an entire profile of characteristics common to the relevant firms, as well as the interaction of these properties. A univariate study, on the other hand, can only consider the measurements used for group assignments one at a time.[11]

Altman's deployment of the MDA model resulted in the model being 'extremely accurate in predicting bankruptcy correctly in 94 per cent of the initial sample with 95 per cent of all firms in the bankrupt and non-bankrupt groups assigned to their actual group classification'.[12] However, as time progressed researchers began to favour a 'logistic regression (logit)'[13] approach over that of the MDA model, which describes a statistical method used to predict the probability of a binary outcome based on one or more independent variables.[14] Pioneers of this model, such as Zavgren, managed to conclude with an accuracy rate of 88% for their predictions of corporate failure.[15] MDA was criticised in the field because it assumes that all the variables are normally distributed, which led Sheppard to argue that 'if all variables are not normally distributed, the methods employed may result in selection of an inappropriate set of predictors'.[16] However, as Arfaoui and Goaied discuss, the rigid structure of these approaches that relied upon static understandings of predictor variables, when coupled with 'advances on other fields such as operations research and *artificial intelligence*, led many researchers to explore the development of more sophisticated discrimination approaches'.[17]

One of the earlier approaches to adopting computational capabilities when assessing financial risk was the technique called 'case-based reasoning' (CBR). The underlying sentiment to CBR is a simple one. Whereas human beings may search through their memories to solve a particular problem, it was found that human beings will often utilise either the first memory they come across more vividly than others (primacy) or remember the last thing they learned/were exposed to more vividly than perhaps older memories (recency). Instead, CBR is based on the concept of a computer being able to search the entirety of its 'memories' (data provided for it) to solve a problem without any prejudice, instead just selecting the data that is *most relevant* to the completion of the task at hand.[18] Another early AI approach that took hold were 'artificial neural networks' (ANNs). In simple terms, an ANN mimics the structure of the human brain and constitutes layers of 'nodes' akin to the brain's neurons; each node performs a simple calculation and passes the output onto the next layer of nodes, and so on. Given the structure employed and the weightings to each output, etc., the network then 'learns' patterns and relationships (through a process called 'training'). When applied to default prediction, researchers found that ANNs were a powerful tool because of their strength in recognising patterns and classifying patterns. However, while they are good at pattern recognition they are 'poor at computational tasks', which means their main uses are related to tasks such as 'association and evaluation'.[19] There are also structural biases within the concept of an ANN, as Kim and Sohn explain:

> However, ANN has some shortcomings. First, ANN depends on the researchers' experience or knowledge to preprocess data in order to select control parameters. Second, it is difficult to generalize the results due to overfitting. Third, it is difficult for ANN to explain the prediction results due to its lack of explanatory power.[20]

Other techniques have also been tested against the quest to predict corporate default rates. The two remaining techniques to really focus on are Support Vector Machines (SVMs) and Decision Trees. SVMs are based on the idea that Support Vectors—the elements of a data set—are the most difficult to classify and by *maximising* the margin between support vectors (or possible solutions) the SVM therefore finds the optimal solution to the problem. The SVM works to categorise classes as efficiently as possible. In layperson terms:

> The best analogy available is that one can think of the SVM as a construction company. The 2D plane is a map and the two classes are two cities. The data points on the 2D plane are analogous to buildings. The Government in this equation wants to create the best highway to minimise traffic which passes through both cities, but you are constrained by the area available to you. The government contract goes to the SVM construction company and what that company does is to minimise traffic by maximising the width of the road as much as is feasible and possible i.e. it will find the widest possible stretch of road between the two cities. Within the boundaries of the two cities, the buildings closest to the outer edges of the prospective road are called 'Support Vectors'. The central line dividing the highway represents the 'hyperplane'. The width of the of the highway is the 'margin'. The goal of the SVM is to find the point where the two cities can be separated to the maximum in order to best understand the differences (to then classify) between the two.[21]

Decision Trees are a system akin to a flow chart based upon identifying the optimal split points to ascertain the best classification the system can attain. Essentially, 'a decision tree classifies data items by posing a series of questions about the features associated with the items. Each question is contained within a node, and every internal node points to one child node for each possible answer to its question. The questions thereby form a hierarchy, encoded as a tree'.[22] In relation to how these two approaches have been applied to credit risk modelling, Golbayani et al. produced a comparative study between ANNs, SVMs and Decision Trees. They found that while SVMs and ANNs (specifically Multi-Layer Perceptron [MLP]) methods are proving to be the most popular, it is decision tree methods which are outperforming everything else.[23]

The effect of this is that market participants, such as institutional investors or issuers, can compute credit risk analysis themselves to a high degree of accuracy. This allows market participants to make better decisions and, potentially, to better predict and anticipate how credit rating agencies will judge creditworthiness. However, no model has been able to absolutely predict how a credit rating agency will ascertain the creditworthiness of an issuer or a particular debt-based product which brings forth a multitude of questions about the processes being adopted by credit rating agencies.

## 3.3 Credit Rating Agency Adoption of AI

Currently, at the time of writing in 2025, it is not possible to know how credit rating agencies are utilising AI techniques in their credit rating processes. It is not possible because, as of yet, the credit rating agencies have not been compelled to make this information public. Even with that limited information however, we can categorise evidence relating to how credit rating agencies are integrated AI into two categories: first, how the credit rating agencies may integrate AI technologies within their rating processes, i.e. for the development of credit ratings and second, how the credit rating agencies are utilising AI to enhance their product ranges for their varied client base, ranging from investors to issuers, and general market participants in between. Examples include S&P's enhancements to its S&P Capital IQ Pro service, which now includes 'Document Intelligence' and 'Chat IQ'. 'Document Intelligence' essentially allows users to generate reports by asking questions, based upon millions of documents.[24] For Moody's, they have revealed that usage of their 'Research Assistant'—essentially the same product—has led to users accessing 60% more data while reducing task time by 30%, while there was also a 35% increase in sustained readership.[25] The sentiment is that the market is finding the credit rating agencies' deployment of AI useful.

**S&P *Capital IQ Pro* and Moody's *Research Assistant***

Both S&P's Capital IQ Pro and Moody's Research Assistant products represent the premium services the two leading credit rating agencies are offering to market participants. Both products allow registered users to utilise a variety of AI-powered functions to collate, categorise, summarise, and evaluate the many millions of documents that each credit rating agency has at its disposal.

S&P Capital IQ Pro

At its core, Capital IQ Pro provides a dualised service, consisting of Document Intelligence and ChatIQ. Document Intelligence is a service which 'from leveraging smart summarization to assessing natural language processing (NLP)-derived sentiment scores[…] transforms how analysts engage with critical information, enhancing their decision-making capabilities'. ChatIQ is more focused on the needs of banking and buyside analysts. It facilitates comprehensive company and industry analysis, market monitoring, financial analysis, and strategy research. Upon launch, ChatIQ was only available to select clients, but the Agency is planning to roll it out more widely in 2025.

Also, S&P has said that it will donate $50 to charity for every mistake or omission generated by the system.

Moody's Research Assistant

Similarly, Moody's *Research Assistant* was launched towards the end of 2023. The system is aimed at industry professionals who want to 'get real-time answers', 'generate holistic insights', and 'accelerate your workflows'. Like its counterpart at S&P, this service allows registered users to run complex analyses on a company or a sector, gather key information from across Moody's library of information, and instantly receive bespoke credit memos and information on topics, targets, or sectors of its choosing.

In celebrating the product's first year in service, Moody's stated that the product was the 'the fastest adopted product ever to come out of Moody's', with more than 100,000 questions asked of the system in the first year of operation. Statistics laid out by Moody's suggest that users could end up saving more than 27% of their time by using Research Assistant, leading Christina Pieretti—General Manager of Digital Insights for Moody's Analytics—to say that 'analysis that used to take hours can now be accomplished in minutes…'

However, of the two categories, data relating to the first category—credit rating agencies integrating AI into their credit rating-developing processes—is lacking. This may be for a variety of reasons. The most obvious is that the rating process, which has been honed over the many years of operation, is arguably the credit rating agencies' main weakness with regards to becoming exposed to liability. In the European Union, for example, the credit rating agencies are exposed to 84 different types of infringements that can trigger civil liability.[26] Article 35 of the 2011 Regulation specifically states that 'where a credit rating agency has committed, intentionally or with gross negligence, any of the infringements listed in Annex III having an impact on a credit rating, an investor or issuer may claim damages from that credit rating agency for damage caused to it due to that infringement'. However, as Lehmann does note, 'mere negligence is not enough… liability is only incurred if the rating agency acted intentionally or with gross negligence'.[27] Nevertheless, the same standard is applied in the United States, with the Dodd-Frank seemingly lowering the bar for private legal action to be taken, although in reality the bar is high; claimants must '*only* prove that CRAs knowingly or recklessly failed to conduct a reasonable investigation of the rating security'.[28] 'Only' is purposefully emphasised here to highlight an issue—just how an investor is supposed to *prove* that the credit rating agencies *knowingly or recklessly failed* to conduct a reasonable investigation is another matter entirely and this constitutes the high bar in question. Nevertheless, for the credit rating agencies, the two pieces of legislation from the jurisdictions which supervise them most closely inject the need for great caution. With the focus here being on how AI may be integrated to affect their credit rating processes, it makes perfect sense that the credit rating agencies will not provide too much information (at least until they must).

Moody's President Michael West remarked that the credit rating agency sees Generative AI as 'an enabler to human judgment in the rating process'. Although Moody's CEO was apparently quick to add that Moody's 'will be deliberate and transparent in the rating agency in terms of how we leverage Generative AI [and that] we are in dialogue with our regulators to make sure that they understand how we're going to do that', the point remains that the credit rating agency *is* actively utilising AI to aid with its credit rating process.[29] How they, as well as S&P Global, are doing that is varied. One way in which it happening is by integrating AI 'tools' to help analysts. In 2023, Moody's deployed a trial of 'GitHub Co-Pilot' which 'serves as a companion to coders'. Moody's Analytics' Head of GenAI and Quantum—Sergio Huerta—noted that GitHub Co-Pilot 'is similar to your phone's "autocomplete" function—except its for programmers'. The programmers at Moody's reacted well to the trail, with 86% of users saying that using the tool 'sped up their daily coding tasks, more than half reported a productivity speed-up of greater than 20%, and more than a tenth reported a productivity speed-up of greater than 50%!' As a result, Moody's moved forward with an enterprise licence for the tool and, separately, created an AI assistant they have called 'Moody's CoPilot' which is 'enabling users across the organisation to harness the power of Generative AI without traditional coding experience'.[30]

S&P Global have reported similar strategies and experience. S&P's Chief Digital Solutions Officer remarked recently in an interview that 'we're entering what we call "AI 2.0". Early AI efforts focused on structured, tabular data. Today, Kensho, our AI and innovation Hub, leverages advanced machine learning and natural language processing (NLP) to transform unstructured data…'[31] An example provided by Swamy Kocherlakota is 'take a 10-K document, which is hundreds of pages long, containing detailed metrics and business intelligence. Instead of manually searching through it, you can ask AI a specific question, like, "Is this company profitable?" Our advanced AI tools analyse structured and unstructured data to identify relevant sections and extract key insights. If the answer is explicit, AI finds it. AI can also suggest questions the user didn't even think to ask'. Yet, this is an example of an *output* from a particular process that is being utilised at S&P and, for Kocherlakota, the 'real value isn't in the AI itself or the model, but in the applications and what you do with them'. He discusses how S&P have created an internal platform call S&P Global Spark Assist, which all 40,000 employees now have access to. He explains:

> What sets us apart is our strong culture of learning. We trained them on how to use S&P Global Spark Assist, and it's now evolving into a powerful tool. Employees create prompts, which we call "sparks", and then can share

> them across teams and across the organization. We have even developed a prompt library for these sparks.
> We've also integrated this system with internal APIs, allowing low-code or no-code workflows. For example, a prompt can query an external database, pull the information back, and deliver an answer seamlessly. Employees can choose from multiple models, and entitlements are in place to ensure appropriate access.
> This grassroots initiative has taken off, with employees sharing and building on sparks.[32]

According to Kocherlakota, the big focus for S&P Global should be 'process mining'. Process mining involves extracting insights from digital traces left in systems and then building new strategies with that data. A credit rating agency, theoretically (and especially if they were given access to a company) could do this firstly for themselves, but also for other companies to better understand their internal creditworthiness. For the credit rating agency themselves, process mining could be crucially important. It would derive information from analysing the sequence of activities that may be involved in credit evaluations, which would allow the credit rating agency to then uncover any inefficiencies or unnecessary loops in the process. This would then allow the agency to identify what are called 'bottlenecks' which may be any delays in the wider process—whether in the data collection, analysis or decision-making phases—which, when brought together, would potentially increase the efficiency and therefore the *reliability* of the credit rating that is ultimately produced. In addition, being able to control processes better may allow the credit rating agency to comply with external regulation. This control of processes would allow the agency to better *demonstrate* to investigating regulators say, during a mandated visit, that the rating process was as protected as it could be from particular conflicts of interest (like collaboration between sales representatives and credit rating analysts, which is forbidden).

Moody's, during a report published in 2024, outlined what they believe are the 'four stages of GenAI maturity' and there may be insight to be gleaned for our understanding.[33]

1. The first stage is basic chatbots and rudimentary RAG frameworks. RAG stands for Retrieval-Augmented Generation and it combines two particular components—Retriever and Generator. The Retrieval

stage is concerned with scanning through a large database of information to match the query with relevant information that it finds. However, as Moody's say in their report, at this level 'AI is basic, prone to hallucinations, and requires extensive human review'. Here, 'hallucinations' refer to when the language model generates responses that are not rooted in data, but in patterns that it has recognised. This is why, as Moody's say, such models require extensive human supervision.

2. The second stage is where 'augmented intelligence brings context-specific capabilities, allowing models to assess implications and determine the appropriate tools for each situation'. Moody's call this 'advanced RAG' but emphasise that human supervision is still a very important part of the process. Dieu et al. discuss how Advanced RAG methods include 'Multi-Stage Retrieval' (initial retrieval is refined by subsequent rounds of retrieval) and 'Fusion Techniques' (attention-based or learning-to-rank models are added to the initial retrieval to better assimilate the findings with the generator).[34]
3. The next stage according to Moody's is 'Augmented Intelligence' which involves AI executing small tasks or making basic recommendations. Examples that Moody's raise include entity or address-matching, data hydration (providing the system with extra data to aid its task, like relevant data from external sources or incorporating other databases, etc.) and anomaly detection.
4. The final stage is 'Autonomous Intelligence'. This involves, according to Moody's, 'AI's ability to plan, execute those plans, evaluate outcomes and adapt accordingly. This may involve groups of specialised agents contributing their unique perspectives and using different skills and multi-modal interaction (vision, voice and even robotics).

It is this stage of Autonomous Intelligence which presents, potentially, the most issues. As the Moody's reports concludes:

> At this stage, we are exploring which tasks and projects are suitable. Examples include software development and analyzing a small business's financials with the same level of scrutiny applied to a Fortune 500 company. AI agents are designed to bring us to the last stage of autonomous intelligence. They can handle more complex tasks where a simple question and answer may not be enough. Some examples could be analyzing financial

reports, generating insights, and even making predictions. The agents can understand nuanced queries, learn from interactions, and provide tailored responses that evolve over time.[35]

The report discusses how, unlike a traditional chatbot, these autonomous agents can develop information in an iterative manner. The report then goes further by asking us to imagine several of these agents interacting with each other, each with different specialities. As the report says, 'we could have a credit risk specialist agent; an Environmental, Social and Governance one; a news analyser; several business analysts; and a team of coders that could create software on demand to fetch information and generate insights'. Most critically, the report concludes with 'Moody's has been at the forefront of deploying AI agents, leveraging several advanced frameworks to enhance their functionality. During 2023 we extensively experimented with a multi-agent simulation framework. By the end of 2023 and going into 2024, we saw fundamental improvements in the technology that could lead us to build enterprise-grade production applications. On one hand, LLMs are getting better, faster, and more accurate'. This is critical because the question(s) then becomes how is this supervised? How does a regulator stay on top of these developments? How does a. the credit rating agency trust the outcome of these autonomous processes, and b. how do they signal the accuracy of these processes to outside parties like, say, the regulator?

Moody's are currently deploying these models and processes. They deploy 'Autogen', an advanced natural language processer developed by Microsoft. In addition to this, they utilise CrewAI, an open-sourced multi-agent orchestration framework created by Joao Moura.[36] Moody's say that CrewAI is particularly useful for when human expertise and AI capabilities need to complement each other. Lastly, the use Langraph which 'employs a graph-based system to represent agent workflows and can be used to analyse complex data relationships'.[37] However, it is not all plain sailing. The report says that 'for more extensive cases in production, agents need to be tamed and their abilities constrained, from avoiding hallucinations to maintaining alignment with user requirements'. Similarly, the credit rating agency also deploys 'Ragas', which is a system whereby a Large-Language Model is deployed as a 'judge' and is 'asked for the correctness of the answer following an extremely specific score-card and criteria among other frameworks to evaluate our AI agents' performance'.

This concept of using LLMs as the 'judge' to these highly autonomous processes is the key for regulators, or at least ought to be. There are recent developments which are specifically focused on this interaction whereby LLMs are being deployed as judges. Cappy, developed by Google, 'takes in an instruction and a candidate response as input, and produces a score between 0 and 1, indicating an estimated correctness of the response with respect to the instruction. Cappy functions either independently on classification tasks or serves as an auxiliary component for LLMs, boosting their performance. Moreover, Cappy efficiently enables downstream supervision without requiring any finetuning…'[38] There is also KAIST's Prometheus 2 model, which is an open-sourced LLM that specialises in evaluating other models and which research has suggested is currently demonstrating 'the highest correlation with both human evaluators and proprietary LM-based judges compared to existing open evaluator LMs'.[39]

These developments signal the speed of development. Moody's seemingly acknowledges this. In its report, the credit rating agency does indeed state that 'as we advance through the stages of GenAI maturity, it is crucial to comprehensively address AI safety. It is paramount that AI systems operate reliably and ethically. This involves implementing robust safety protocols, continuous monitoring, and rigorous testing to prevent unintended consequences'. This, essentially, is the purpose of this book. *How* that is done is now of the utmost importance. On Moody's own website, there are only fleeting references to this issue of safety, with declarations not extending past the likes of 'we will explain our approach to AI transparently', or 'we will use AI responsibly in line with our values'.[40] As we shall see in the next chapter, the current approach by the credit rating agencies—and especially the Big Two—is that the environment is too new to make an solidified commitments, so only declarations like those passive statements above will suffice in the current environment. This book will conclude that, perhaps, that is not enough.

## 3.4 Acquisitions

Now we have been introduced slightly to how the leading credit rating agencies are integrating AI into both their rating processes and the tools they offer the marketplace, it is worth trying to capture the current picture of where the agencies are evolving. One way to capture this is to better understand their Acquisition decisions. Recently, there has been a focused effort by S&P Global and Moody's to incorporate key AI-related businesses. While the credit rating agencies have, for a long time, had a policy of monitoring start-ups and then taking them over if the products generated by the start-ups were proven to be useful,[41] they also invest considerable sums into buying key market developers.

In 2018, Moody's made a minority investment into **QuantCube**. QuantCube is a Paris-based 'innovative provider of real-time, AI-based predictive analytics for corporate clients, financial institutions and investment managers'.[42] QuantCube developed an approach they labelled 'nowcasting' as part of their Macroeconomic Intelligence Platform, which can then be applied to different sectors. The sentiment of the 'nowcast' is to provide indices on particular variables that are applicable to the sector the user is looking at, which in turn then allows the system to generate 'critical insights'.[43] In 2021, Fitch also provided a minority investment for **Sigma Ratings**. Fitch led a $6 million funding round for the start-up to develop its global risk intelligence platform, which aims to develop non-financial and risk event indicators that can indicate potential governance risks and financial crime.[44] Speaking at the time of investment, Fitch's President Ian Linnell said that 'the way they deliver information through machine-driven analysis means it is more structured and efficient'.[45]

Yet, more recently, the Big Two have made moves to signal their intention. On January 6, 2025, S&P Global announced that it had acquired **ProntoNLP**, 'a leading provider of Generative Artificial Intelligence tooling, allowing users to derive differentiated insights from unstructured and structured data'.[46] Frank Tarsillo, S&P Global market Intelligence's Chief Technology Officer, said that 'by integrating ProntoNLP's cutting-edge expertise in signal processing with our existing platforms, we aim to enable richer context understanding, more accurate predictions, and faster decision-making'. This acquisition may suggest that S&P have felt it necessary to bring in external expertise to address some shortcomings in their current systems. Just seven days later, Moody's announced that it had acquired **CAPE Analytics**, 'a leading provider of geospatial AI intelligence for residential and commercial properties. The acquisition will bring together Moody's industry-leading Intelligent Risk Platform and catastrophe risk modelling for the insurance sector with CAPE's cutting-edge geospatial AI analytics, creating a sophisticated property database capable of delivering instant, address-specific risk insights'.[47] It is likely not the case, but we may be able to derive some information from the fact that Moody's is bringing in expertise for a specific sector, while S&P

have brought in expertise to improve their core systems. Moody's is often congratulating itself for being the first of the oligopolistic members to embrace AI-related technologies, and it is potentially revealing that their recent acquisition is very sector-specific (it had only recently acquired 'Praedicat', a provider of Casualty and Liability modelling for insurance companies).[48]

The recent acquisitions by the Big Two show that investment into their AI-related systems is very much on the agenda. The Big Two are going 'all in' on AI, that much is clear to see. It remains to be seen if the investment will continue or whether they will continue to build their capabilities in-house. The culture of the Big Three however, one in which the participants of the oligopoly understand the strength that can be derived from that position, means that there could be many other acquisitions of start-ups and specialised players in the years to come.

## 3.5 Conclusion

Technology is evidently critical to the art of assessing creditworthiness but, in the modern day, AI-related technologies will come to define the credit rating process. Whether it is for researchers and analysts seeking to predict and anticipate the ratings of the credit rating agencies, or the credit rating agencies themselves seeking to make their rating processes more efficient, AI will be a central theme for the world of credit ratings for some time to come. Yet, it is an application of a technology that is focused on efficiency to a sector that is about subjective opinions that perhaps prevents the technology from becoming supreme.

It is undoubtably true that for credit rating agencies and private analysts alike, AI technologies allow the process of synthesising substantial amounts of data to be almost seamless. However, in the next chapter we will start to assess this issue from the perspective of the regulator and the legislator, and the reality is that for all the positivity coming from inside the credit rating agencies, the story is not simple. There are significant concerns that have not yet been considered. But, to its credit, a report by S&P in 2024 did meet this concern head-on. Melissa Incera, a research analyst within S&P's M&A team, said that while the research is identifying that GenAI is increasingly being served by credit rating agencies and subsequently used by their users, 'there remains a mountain of hidden complexity, governance challenges, and additional development as these entities become more complex and autonomous'. Interestingly, Incera concludes that 'these issues must be proactively addressed – otherwise, this trend may only perpetuate the disillusionment already emerging around GenAI in the short to medium term'.[49] The sentiment of the report by Incera is that agentic AI is the evolution of LLM-based AI but, again, caution is offered:

> It bears mentioning that as agentic architectures develop, so does the need for robust governance structures to ensure these entities are compliant and safe to use at scale. Unlike LLMs, which primarily process and generate text, AI agents can interact with the environment, make decisions, and potentially take actions. As these agents grow more autonomous and capable, questions will arise about accountability, bias and potential misuse. This places higher emphasis on strong governance frameworks, like clear guidelines for agent behavior, robust security measures, and regular audits of AI agent systems. Although MLOps focuses on the life cycle of machine-learning models, AgentOps, a newer concept, extends these principles to AI agents, focusing on their training, deployment and ongoing management. Ultimately, a well-governed AI agent ecosystem is crucial to harnessing the benefits of these technologies, while minimizing potential drawbacks.[50]

These warnings are important. If a strong governance framework is critical, how that governance framework is designed, against what parameters, and how it is communicated with the outside world are all vitally important issues to engage with. But, with regulators seemingly struggling to stay ahead of a curve that is constantly shifting in the field of AI, the role of regulators, legislators and standard-setters in the near-future will be defining for the sector. Now, we must consider all these issues from the perspective of those with the responsibility to make sure all of this novel technological evolution does not end up in catastrophe.

## Notes

1. Hazel Ilango and Daniel Cash, 'Going beyond methodology: The Credit Rating Committee's vital but overlooked role in climate risk integration' (2023) IEEFA (Nov 2) https://ieefa.org/resources/going-beyond-methodology-credit-rating-committees-vital-overlooked-role-climate-risk.
2. John (Xuefeng) Jiang, Mary H Stanford, and Yuan Xie, 'Does it matter who pays for bond ratings? Historical Evidence' (2012) 105 Journal of Financial Economics 607, 620.

3. Meg Bratley, 'Sustainable Investment and the Central Role of Rating Agencies' (2024) IFA (Aug 16) https://ifamagazine.com/esg-ratings-bill-could-end-tick-box-esg-greenwashing-and-ai-opaqueness-insights-howard-kennedy/.
4. Alvin J Surkan and J. Clay Singleton, 'Neural networks for bond rating improved by multiple hidden layers' (1990) Proceedings of the IEEE International Conference on Neural Networks 157–162.
5. Soumitra Dutta and Shashi Shekhar, 'Bond rating: a nonconservative application of neural network' (1988) IEEE International Conference on Neural Networks 443–450.
6. Herwig P Langohr and Patricia T Langohr, *The Rating Agencies and their Credit Ratings: What They Are, How They Work, and Why They Are Relevant* (John Wiley & Sons 2010) 99.
7. Ibid.
8. Kristen Walters, 'Inside BlackRock's risk management framework' (2019) Informa Connect (Aug 8) https://informaconnect.com/inside-blackrocks-risk-management-framework/#:~:text=Structurally%2C%20BlackRock's%20risk%20management%20team%20is%20independent,unbiased%20perspective%20while%20working%20closely%20with%20investors.&text=To%20measure%20and%20manage%20risk%20across%20the,framework%20for%20managing%20risk%20across%20investment%20businesses.
9. William H Beaver, 'Financial Ratios as Predictors of Failure' (1966) 4 Journal of Accounting Research 71–111. Financial ratios is an accounting term used to describe particular elements of a financial entity's financial statements, like aspects that count towards profitability, or liquidity, for example.
10. Edward I Altman, 'Financial rations, discriminant analysis and the prediction of corporate bankruptcy' (1968) 23 The Journal of Finance 4 589.
11. Ibid. 591.
12. Ibid. 609.
13. Hong Sik Kim and So Young Sohn, 'Support vector machines for default prediction of SMEs based on technology credit' (2010) 201 European Journal of Operational Research 838–46.
14. See Chao-Ying Joanne Peng, Kuk Lida Lee and Gary M Ingersoll, 'An Introduction to Logistic Regression Analysis and Reporting' (2002) 96 The Journal of Educational Research 1.

15. Christine V Zavgren, 'Assessing the Vulnerability to Failure of American Industrial Firms: A Logistical Analysis' (1985) 12 Journal of Business Finance & Accounting 1 19–45.
16. Jerry P Sheppard, 'A Resource Dependence Approach to organizational Failure' (1995) 24 Social Science Research 1.
17. Mourad Arfaoui and Mohamed Goaied, 'The Prediction of Corporate Financial Distress in Tunisia' (2012) SSRN (Sept 9) https://papers.ssrn.com/sol3/papers.cfm?abstract_id=1477609.
18. Stephanie M Bryant, *A Case-Based Reasoning Approach to Bankruptcy Prediction Modelling* (1996) LSU Historical Dissertations and Theses https://repository.lsu.edu/cgi/viewcontent.cgi?article=7320&context=gradschool_disstheses.
19. Herbert L Jensen, 'Using Neural Networks for Credit Scoring' (1992) 18 Managerial Finance 6 15.
20. Kim and Sohn (n 13) 838.
21. Tasmay Tibrewal, 'Support Vector Machines (SVM): An Intuitive Explanation' (2023) Medium (Jul 1) https://medium.com/low-code-for-advanced-data-science/support-vector-machines-svm-an-intuitive-explanation-b084d6238106#:~:text=You%20can%20think%20of%20SVM,the%20area%20available%20to%20you.
22. Carl Kingsford and Steven L Salzburg, 'What are decision trees?' (2008) 26 Nature Biotechnology 1011.
23. Parisa Golbayani, Ionuţ Florescu, and Rupak Chatterjee, 'A comparative study of forecasting corporate credit ratings using neural networks, support vector machines, and decision trees' (2020) 54 North American Journal of Economics and Finance 101,251.
24. S&P, 'S&P Global transforms S&P Capital IQ Pro Experience with the Launch of New Generative AI-powered Capabilities' (2024) S&P Global (Nov 12) https://press.spglobal.com/2024-11-12-S-P-Global-Transforms-S-P-Capital-IQ-Pro-Experience-with-the-Launch-of-New-Generative-AI-Powered-Capabilities.
25. Moody's, *Moody's Gen-AI-powered Research Assistant drives significant efficiency gains for financial services industry* (2025) https://www.moodys.com/web/en/us/site-assets/genai-research-assistant-financial-services.pdf.
26. Regulation (EU) No 513/2011 https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32011R0513&qid=1740163838674.

27. Matthias Lehmann, 'Civil liability of rating agencies – an insipid sprout from Brussels' (2016) 11 Capital Markets Law Journal 1 60–83, 78.
28. Valentin Dimitrov, Darius Palia, and Leo Tang, 'Impact of the Dodd-Frank Act on credit ratings' (2015) 115 Journal of Financial Economics 3 507.
29. Michelle Chang, 'AI will judge your entire financial life and Moody's is stoked about it' (2024) Quartz (Feb 14) https://qz.com/ai-will-judge-your-entire-financial-life-and-moody-s-is-1851253842.
30. Sergio Gago Huerta, 'Moody's Coders Find Benefit in Using Generative AI' (2024) Enterprise Security https://network-security.enterprisesecuritymag.com/cxoinsight/moody-s-coders-find-benefit-in-using-generative-ai-nid-3725-cid-13.html.
31. Asheem Chandna, 'Gen AI Present and Future: A Conversation with Swamy Kocherlakota, EVP and Chief Digital Solutions Officer at S&P Global' (2025) Greylock (19 Feb) https://greylock.com/greymatter/gen-ai-present-and-future-a-conversation-with-swamy-kocherlakota-evp-and-chief-digital-solutions-officer-at-sp-global/.
32. Ibid.
33. Moody's, *GenAI's transformative potential in the financial sector: the evolution of agents* (2024) https://www.moodys.com/web/en/us/insights/resources/the-rise-of-ai-agents.pdf.
34. Anh Nguyen Thi Dieu, Hien T. Nguyen, and Chien Ta Duy Cong, 'The enhanced context for AI-generated learning advisors with Advanced RAG' (2024) 18th International Conference on Advanced Computing and Analytics 95 https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=10818362.
35. Moody's (n 33).
36. Vanna Winland, Anna Gutowska, and Meredith Syed, 'What is CrewAI?' (2024) IBM (Aug 2) https://www.ibm.com/think/topics/crew-ai#:~:text=crewAI%20is%20an%20open%20source,%E2%80%9Ccrew%E2%80%9D%20to%20complete%20tasks.
37. Moody's (n 33).
38. Yun Zhu and Lijuan Liu, 'Cappy: Outperforming and boosting large multi-task language models with a small scorer' (2024)

Google (Mar 14) https://research.google/blog/cappy-outperforming-and-boosting-large-multi-task-language-models-with-a-small-scorer/.

39. Seungone Kim, Juyoung Suk, Shayne Longpre, Bill Y. Lin, Jamin Shin, Sean Welleck, Graham Neubig, Moontae Lee, Kyungjae Lee, and Minjoon Seo, 'Prometheus 2: An Open Source Language Model Specialised in Evaluating other Language Models' (2024) Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing 4334–4353 https://aclanthology.org/2024.emnlp-main.248.pdf.
40. Moody's, *AI Principles* (2025) https://www.moodys.com/web/en/us/innovation/ai-principles.html#accordion-56af1a3175-item-daf7edc587.
41. Chandna (n 31).
42. Moody's, 'Moody's announces investment in QuantCube, AI-based Predictive Analytics Firm' (2018) Moody's (May 29) https://ir.moodys.com/press-releases/news-details/2018/Moodys-Announces-Investment-in-QuantCube-AI-Based-Predictive-Analytics-Firm/default.aspx.
43. QuantCube, 'Macro Nowcast Indicators' (2025) https://www.quant-cube.com/macro-nowcast.
44. Daniel Cash, 'Sigma Ratings: Adapting the Credit Rating Agency Model for the anti-Money Laundering World' (2020) 23 Journal of Money Laundering Control 1 1–10.
45. Anna Domanska, 'Fitch Ratings invests in AI startup to improve bank misconduct detection' (2021) Industry Leaders (Jan 18) https://www.industryleadersmagazine.com/fitch-ratings-invests-in-ai-startup-to-improve-bank-misconduct-detection/.
46. S&P, 'S&P Global acquires ProntoNLP, expanding its Generative AI-powered product portfolio' (2025) https://press.spglobal.com/2025-01-06-S-P-Global-Acquires-ProntoNLP,-Expanding-its-Generative-AI-Powered-Product-Portfolio.
47. Moody's, 'Moody's to acquire CAPE Analytics, adding AI-powered Geospatial Property Risk Intelligence to its Industry-leading Insurance Risk Models' (2025) https://ir.moodys.com/press-releases/news-details/2025/Moodys-to-Acquire-CAPE-Analytics-Adding-AI-Powered-Geospatial-Property-Risk-Intelligence-to-Its-Industry-Leading-Insurance-Risk-Models/default.aspx.

48. Moody's, 'Moody's acquires Praedicat, adding Casualty and Liability Modelling Capabilities' (2024) https://ir.moodys.com/press-releases/news-details/2024/Moodys-Acquires-Praedicat-Adding-Casualty-and-Liability-Modeling-Capabilities/default.aspx.
49. Melissa Incera, 'The path from LLMs to agentic AI' (2024) S&P (Nov 29) https://www.spglobal.com/market-intelligence/en/news-insights/research/the-path-from-llms-to-agentic-ai.
50. Ibid.

CHAPTER 4

# The Regulatory Perspective

**Abstract** The position, constitution and potential options for regulators in the field are presented in this chapter. The foundation developed in previous chapters is utilised here to contrast with the actions of credit rating regulators in key jurisdictions for the leading credit rating agencies. The European Union, as a leader in this space, is analysed specifically because of the relative importance of its new AI Act, which is considered generally and from the perspective of the credit rating agencies. The chapter concludes with a distinct analysis of the concept of 'Co-Regulation' which has been identified as the chosen format of regulation of AI technologies; the intricacies of this approach are considered in the chapter.

**Keywords** AI regulation · Co-regulation · EU AI Act · Regulatory capture · Risk governance · Legal accountability

## 4.1 Introduction

The central themes of the prematurity of the credit rating agencies' integration of AI technologies are, perhaps unsurprisingly, repeated when we look at the situation from the regulatory perspective. Particularly in relation to the credit rating agencies specifically, regulators are only at the very



D. Cash and N. Tkachenko, *The Risk of Artificial Intelligence in Credit Ratings*, https://doi.org/10.1007/978-3-031-95543-3_4

start of their journey. However, there are principles quickly emerging in the larger relationship between the State and AI governance more generally that are directly attributable to our focus on the credit rating agencies. Key governance approaches and philosophies have been evolving over the past few decades which, today, represent the foundational launchpad that any credit rating agency-related AI regulation will launch from. The key question is whether those principles comfortably align with the realities of the credit rating sector, or not.

To understand this better, this chapter begins by assessing how regulators are currently understanding the challenges that may emanate from the credit rating agencies' usage of AI technologies. We already know that the technologies are fluid and evolving, so getting a picture of how regulators plan to get ahead of this evolution, if at all, is important. Regulators have a unique authority which allows them to canvass the views and considerations of market participants like few others, and those views and considerations will be key to our understanding. Yet, to really understand whether regulators and regulatory frameworks may be appropriate to properly balance the needs of the economy with the needs of society, our analysis will also go into deeper discussions relating to general AI regulatory approaches. This is useful because the world is yet to determine nor understand how credit rating agencies will integrate AI into their processes. Current regulatory techniques and approaches, however, are the foundation for what will be applied so that knowledge will stand us in good stead.

To conclude that analysis, we will interrogate a concept that is seemingly bound to become central to this field: co-regulation. Co-regulation, as we shall see, is proving to be the predominant approach favoured by regulators in key jurisdictions for credit rating agencies and AI technologies moreover. The benefits and the drawbacks will be reviewed against a backdrop of analysis that considers the potential for co-regulation to lead to greater issues. Co-regulation may, of course, prove to be successful but it is the constitution of the regulatory target—AI—that is causing concern for onlookers. The potential for automation leading to real-world impact without the adequate levels of human supervision is apparently of great concern. We will counterbalance this concern with the realities as put forward by the leading credit rating agencies—automated and independent artificial intelligence is predicted to be the ultimate stage of development but, without the necessary regulatory framework governing that evolution, does that mean catastrophe is inevitable?

## 4.2 Credit Rating Regulation for the New Era

The first important aspect to note is that of the two main jurisdictions for the large credit rating agencies, only the European Union appears to be actively considering this issue of AI integration within the credit rating space. The two jurisdictions, aside from current geopolitical differences that are brewing since President Trump's second term in office began, represent vastly different regulatory philosophies. For example, Gamito et al. discuss how 'Co-regulation was noted by the United States Congress in 2002 to describe certain aspects of European legislation… the EC pragmatically funds standards and ex ante supports co-regulation in cases where the US would ex post regulate via competition law'.[1] Further, Newman and Bach state that:

> In the U.S., the government induces self-regulation largely through the threat of stringent formal rules and costly litigation should industry fail to deliver socially desired outcomes. Industry thus views self-regulation as a pre-emptive effort to avoid government involvement. The relationship between the public and private sector is spotty, formal and frequently adversarial. We label the ideal-typical U.S. model legalistic self-regulation. In Europe, public sector representatives meet with industry and agree on a joint course of action. Here, private and public sectors view each other as partners in an often-informal self-regulatory process. Coordinated self-regulation is the term we use to describe the European ideal-typical model.[2]

It is within those differing regulatory environments that credit rating agencies will find themselves imminently. However, the US is yet to take a concerted action or pathway to a regulatory framework where AI is concerned.[3]

In Europe, the regulatory machinery has produced the AI Act, which will be reviewed in more detail shortly, and has made announcements and comments on AI's relationship with particular sectors rather frequently. For example, in the new ESG Rating Regulation, launched at the end of 2024, the Regulation cites artificial intelligence on only three occasions, stating that ESG Rating providers must disclose to the public 'where applicable' reference to the use of artificial intelligence in the data collection or rating process including information about current limitations and risks of using artificial intelligence'.[4] In a Public Statement in 2024, ESMA (the EU's primary securities regulator) when discussing the provision of financial services said that ‘the principles and controls outlined in this statement would aim at reminding firms that they should have in place appropriate measures to also control the use of AI systems by employees in any form, including any third-party AI technologies, whether specifically envisaged or already adopted by the firm itself or without any direct knowledge and approval of senior management’.[5]

The sentiment from the financial regulator in Europe is a clear one. It is actively providing the marketplace with instruction as to what will be deemed acceptable by the regulator. ESMA has acknowledged that there are considerable risks associated with using AI technology, and it categorises these risks into the following groups:

- Lack of accountability and oversight (overreliance). The sentiment for the regulator here is that providers and users could all too easily relegate human judgement for automated outcomes. The regulator contextualises this problem, saying ‘this over-reliance can be particularly risky in complex, unpredictable financial markets where AI may not accurately predict outcomes’.
- Lack of transparency and explainability/interpretability. The regulator puts forward its concerns that the complex underpinnings of AI technologies are often ‘black boxes’ which, in turn, makes it difficult for staff at all levels to understand the processes they are using, and in turn any potential issues that may arise from that usage.
- Security and Data Privacy. The regulator is also concerned with the data issues and privacy issues that will inevitably arise when there are considerable amounts of data being stored.
- Robustness and the reliability of the output. The most headline-grabbing concern, perhaps, of the utility of AI is the issue of bias and system training. The regulator points to the potential of ‘hallucinations’ within the technology that can warp the reliability of the output produced by the technology. Also, the regulator notes that ‘training data used to develop the AI tool can also introduce biases in the way results are computed making predictions incorrect/inaccurate. These biases are often difficult to identify and correct’.[6]

On that basis, ESMA took another opportunity—this time directly pointed towards the credit rating agencies and benchmark administrators during a consultation on developing guidelines—to provide clear direction on what will be required. Though there has not been formal regulation or legislation for credit rating agencies (yet), the following is a clear instruction for what will be acceptable, and what will not:

> Nonetheless, ESMA has also taken the opportunity to expand and clarify some of its expectations related to technology given the growing risk and opportunities provided through its use. For example, where a company uses artificial intelligence (AI), its internal control framework should be mature enough to assess and manage the risks of AI and to be integral to the AI lifecycle within a company. This includes the establishment of a supervised entity's AI strategy, ethics and principles, an appropriate governance and risk management framework, sufficient disclosures and system documentation, and controls around the design criteria, modelling, training, evaluation and deployment of AI systems.[7]

Earlier in 2023, ESMA had produced a report that sought to understand the rate of development within the financial sector regarding AI integration.[8] After warning of similar risks as other reports—like explainability, interconnectedness and systemic risk, algorithmic bias, operational risk and data quality—the report captures the views of the credit rating agencies with respect to how important AI currently is from their perspective. Against the backdrop of the warning from ESMA that most of the risks identified by ESMA are not inherent to models or algorithms developed by AI, *but are amplified when AI is used*, the report utilises a survey sent to credit rating agencies which asked what they believed with the primary risks with integrated AI into their business (Fig. 4.1).

Despite suggesting that modelling and ethical concerns were the primary issues they were facing, surveyed credit rating agencies responded rather unanimously that 'these risks had yet to materially affect their activity'.[9] This is because, according to the surveyed credit rating agencies (who exactly was surveyed and the total results of the responses have not been made public) they are in no way seeking the 'automate the credit rating assessment process'. Instead, they are mostly using AI technology to source information (Fig. 4.2).

It is worth pausing here for a moment to focus on this declaration. The idea that the agencies would seek to automate the credit rating process may sound obvious—more automation means less need for credit rating analysts—but a deeper understanding results in the realisation that full automation is the last thing a credit rating agency would want. More automation in the credit rating process means less need for the credit

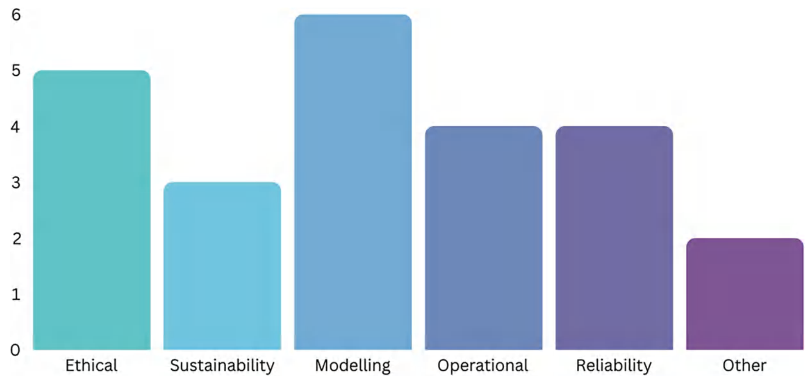


**Fig. 4.1** CRAs' views on AI risks. Diverse risks considered, but yet to materialise (*Note* The chart is based on a survey sent to 24 CRAs based in the EU in April 2022. Eleven CRAs provided their responses to the question 'What do you consider the key risks of using AI in credit rating operations?' Sustainability concerns refers to the energy consumption linked to distributed ledger technology. *Source* ESMA, *Artificial Intelligence in EU Securities Markets* [2023], p. 18)

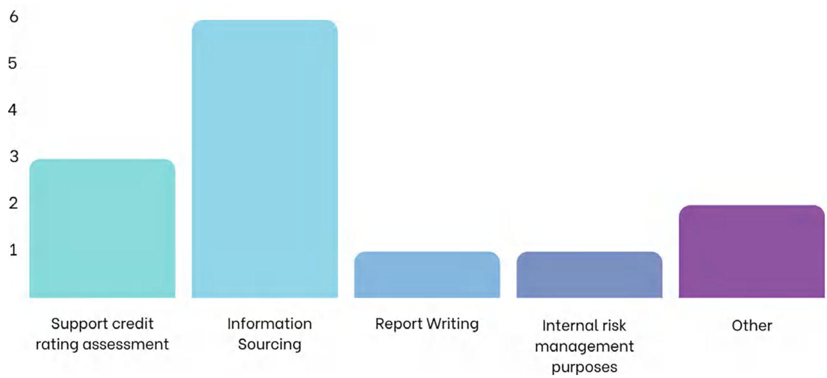


**Fig. 4.2** Activities performed with AI by CRAs. CRAs use AI mostly in support functions (*Note* The chart is based on a survey sent to 24 CRAs based in the EU in April 2022. Eleven CRAs provided their responses to the question 'Does your entity make use of AI to perform one or more of the following activities?' *Source* ESMA, *Artificial Intelligence in EU Securities Markets* [2023], p. 16)

rating agencies, full stop. It is for this reason that credit rating agencies fight tooth and nail not to have their methodological process infringed by outside entities and why legislators and regulators make physical reference to their actions never amounting to infringing on methodological processes. The credit rating agencies are seeking to find a 'sweet spot' whereby they can make their processes more efficient but equally retain their subjective role as opinion providers on creditworthiness, which is ultimately their *systemic* role.

Continuing, ESMA suggest that another key issue that must be considered when contemplating how credit rating agencies may best integrate AI is in the understanding that 'in short, AI depends on data as its "fuel": the success of AI tools is highly dependent on data quality, and poor-quality, noisy data can easily result in unreliable models'. This is pertinent from the perspective of the credit rating space because there are two specific areas of credit rating agency business where poor-quality data has had and is having a demonstrable effect. We saw how, in the lead up to the Global Financial Crisis, poor-quality data on the underlying mortgages in RMBS CDOs was not picked up by credit rating agencies and then amalgamated into their fallible methodologies and processes. Similarly, a common complaint of the credit rating agencies in relation to the current 'debt crisis' affecting the developing world is that data coming from developing sovereign nations is not strong enough or complete enough for more accurate ratings to be produced.[10] Extrapolating from ESMA's concern, such high dependence on data quality puts the world's most vulnerable at an even greater risk in the capital markets than they already face if AI were to be more widely adopted without the necessary controls.

In other reports, credit rating agencies have maintained that their integration of AI is at a very early stage and that, as such, no material issues have arisen.[11] In a recent public consultation, ESMA asked the market a series of questions relating to potential amendments to the CRA Regulations in Europe. Their final question of twelve asked: 'Do you see merit in requesting a disclosure of the use of technological innovations such as Artificial Intelligence (AI) in the rating process?'[12] The answers from the credit rating agencies in the Final Report were telling.[13] In summation, ESMA concluded that 'the majority of the respondents recognized that increased transparency when it comes to the use of AI by CRAs, could be of benefit to the market, while noting that such a legislative change could be premature at this stage. In this regard, respondents were most supportive of increased levels of disclosure in cases where AI technology is used in the creditworthiness assessment, with a reduced necessity for such disclosures when it was applied merely for efficiency purposes, like document review'. But, separately, the credit rating agencies had their say.

The **European Association of credit rating agencies** responded that 'We are of the view that this topic goes well beyond the overall framework of this consultation focusing on ESG factors and rating outlooks. We think that questions linked to technological innovations in the rating process (such as artificial intelligence) would require a separate and detailed analysis (e.g. through an ESMA thematic review or through round tables) to better define and clarify the use of potentially disrupting technological innovations in the CRAs world. We therefore request to postpone any such compulsory disclosure at this stage. If an agency wishes to disclose such an information, it could do so on a voluntary basis'.[14] While not supporting any submission one way or the other, it is likely sensible to note that such a technical question ought really to have been framed within its own consultation, as indicated by the European Association of credit rating agencies. The importance of this topic was perhaps relegated, especially as it was the twelfth and final question of respondents.

**Scope Ratings** responded by saying that:

> Yes, Scope Ratings sees potential merits in expanding disclosure on this topic. It would foster digital innovation, increase transparency and raise awareness in the rating industry on innovative approaches while working towards the establishment of a level playing field among CRAs. Scope Ratings believes that the general objective to provide transparency or effective regulatory oversight on the rating process already require de facto that CRAs elaborate on the use of technological innovations whenever they contribute to the elaboration of credit ratings and outlooks. To enable an effective implementation of the proposal, Scope Ratings will welcome a greater precision and specification regarding the types of technological solutions that would require specific disclosures and under which circumstances disclosures would be required. Scope Ratings understands this matter is still work in progress and anticipates that the dialog with regulators on this topic will continue to evolve.[15]

This response echoes that of the European Association of credit rating agencies in calling for more detail on the issue to be presented to the marketplace. It quickly becomes evident that while the credit rating agencies are declaring that AI is still at a nascent stage for their business the importance of AI to the future of the sector is widely accepted.

The Big Three agencies each only had short responses to deliver. **Fitch Ratings** said 'Fitch believes an AI disclosure requirement would be reasonable if the technology were to be used as part of the credit analysis process itself. However, we do not see any merit in requiring disclosure if the technology is simply used for efficiency, such as using natural language processing to review documents'.[16] As we have seen already and will continue to see, this view is not shared by regulatory officials. The suggestion that credit rating agencies ought not to have to disclose information if they are only using AI technology for 'efficiency' is prospectively a very dangerous assertion, as this book will ultimately show. For **S&P Global**, the focus falls in line with the European Association of credit rating agencies, because for S&P 'We are unsure about this question in the context of this consultation, including any relevance to the Legislative Mandate and the other proposed changes to the Delegated Act and CRAR, as the Consultation Paper provides no explanation or detail. As a general matter, we understand ESMA's interest in the use of innovative technologies by CRAs and can see circumstances in which a CRA may indeed choose to publicly disclose its use of AI. However, we would urge ESMA to conduct a thorough cost–benefit analysis before mandating any disclosure, in particular at a time when the technology, its use by CRAs, and AI-specific legislations are still evolving'.[17] For **Moody's**, the agency simply responded 'We would suggest considering this question once that the EU's AI Act has been finalised, in an effort to understand the implications of this horizontal law on CRAs'.[18] Finally, **Kroll Bond Rating Services (KBRA)** also concluded that:

> While there could be merit in disclosure of the use of technological innovations such as Artificial Intelligence (AI) in the rating process, we urge caution to avoid the premature introduction of a specific disclosure requirement. The use of technological innovation, particularly with respect to AI, is the subject of ongoing consideration by a range of stakeholders including market participants, investors, service providers, and policymakers.
>
> In order to avoid a disclosure requirement that may be redundant with existing requirements or may not serve its purpose over the longer term, we would recommend that this area continue to be closely monitored and that a specific request for the disclosure of the use of technological innovations

be made once there is broader consensus around terminology, definitions, and related guidance.[19]

It is no accident that there are clear themes emerging from the responses. While it is acceptable to suggest that ESMA probably should not have raised such an important issue in such a flippant manner, the responses reveal an underlying recognition of the need to protect oneself against liability. The credit rating agencies, as has been shown in the literature, are fundamentally and innately tuned-in to the concept of liability due to the nature of their product and the way they develop their products.[20] The new world of AI and all the possibilities that may present themselves to the credit rating agencies also come with a connected risk: liability. The credit rating agencies are clearly aware of this.

The current situation then, for the credit rating sector from a regulatory perspective, is one defined by transition. The regulators are starting to ascertain what standards may be required (like adequate internal frameworks, etc.) but those declarations are not being supported by a wider legislative or regulatory framework evolution. Instead, the credit rating agencies are adopting a defensive position bound by the need to protect their fragile businesses from liability claims. Therefore, it is evident that the regulatory framework needs to develop quickly to keep up with the pace of change. The important question then is how the framework should now develop. If it favours speed of development over quality of development, it may make irrevocable mistakes. However, if the priority is quality over speed, the regulatory framework risks being applied to a marketplace that has long since moved on. Finding the right balance is critical and the pursuit of that balance is affecting the wider intersection between new technology, community safety and economic development.

## 4.3 The Regulatory Challenge of AI Technology

The recent but relatively rapid increase in AI technology usage has led several bodies to consider the regulatory challenges that come with adopting AI technology. Many of the challenges that face regulators generally in the AI space will also directly affect the regulators focused on the credit rating space. Of the issues identified, ranging from regulatory capture to the optimal model for regulation to follow, the delicate balance of stakeholders is proving to be difficult to fully integrate.

The UK House of Lords' report in 2024 firstly demonstrates the geopolitical element to the conversation which is quickly starting to dominate the space. It states, 'the UK should continue to forge its own path on AI regulation, balancing rather than copying the EU, US, or Chinese approaches. In doing so the UK can strengthen its position in technology diplomacy and set an example to other countries facing similar decisions and challenges'.[21] However, against that backdrop, the Report discusses two key issues that regulators need to get on top of. First, it discusses how *Regulatory Capture* could prove to be a massive hurdle for a State to overcome. The Report says that 'throughout our inquiry we encountered mounting concern about regulatory capture. This might occur through lobbying or because officials lack technical know-how and come to rely on a narrow pool of private sector expertise to inform policy and standards… current trends suggest growing private sector influence. Witnesses emphasised the limited extent of public sector expertise and the necessity of closer industry links'.[22] Additionally, the framing of the narrative around AI technologies and the potential for risk is increasingly coming to the fore, with the Report stating that 'there has been further concern that the AI safety debate is being dominated by views narrowly focused on catastrophic risk, often coming from those who developed such models in the first place. Critics say this distracts from more immediate issues like copyright infringement, bias and reliability'. As we have seen, these are already key concerns being raised by regulators in the credit rating space.

The second hurdle identified by the House of Lords Report is the presence of material conflicts of interest. The need to bridge the asymmetrical divide that exists between the public and private sectors in relation to AI technology means that the people needed to advise may be the same people who can profit from the technology. On that point, the Report recommends that 'the Government should implement greater transparency measures for high-profile roles in AI. This should include further high-level information about the types of mitigations being arranged, and a public statement within six months of appointment to confirm these mitigations have been completed'. The importance of transparency in the world of AI is clearly visible.

Ultimately, for the UK in its post-Brexit environment, any sense of conservativism is being challenged. As the Report ultimately concludes:

> The Government is not striking the right balance between innovation and risk. We appreciate that recent advances have required rapid security

> evaluations and we commend the AI Safety Summit as a significant achievement. But Government attention is shifting too far towards a narrow view of high-stakes AI safety. On its own, this will not drive the kind of widespread responsible innovation needed to benefit our society and economy. The Government must also recognise that long-term global leadership on AI safety requires a thriving commercial and academic sector to attract, develop and retain technical experts.[23]

The UK is potentially in a bind. Seeking to build a secure regulatory environment for this new technology phenomenon is, perhaps, not suitable to the current climate around the UK. Rather, more experimental, pro-Business approaches are being favoured. The UK, given its descension from the European Union, can perhaps take such risks. The European Union, however, has very different constraints.

### *4.3.1 The European Union's AI Act*

Proposed originally by the European Commission in April 2021, the EU's AI Act came into force on August 1, 2024.[24] Regulation (EU) 2024/1689, laying down harmonised rules on artificial intelligence, consists of 113 Articles that cover everything from definitions and roles to enforcement strategies and organisation of responsibility across the Union.[25] Yet, with the challenge of legislating for such a fluid and relatively nascent concept proving to be challenging, the impact of the AI Act has been divisive. Key to the concerns being raised is whether the AI Act is formalising a technocratic or paternalistic approach that may solidify a lack of transparency and accountability; when we marry these concerns to the concerns this book has already identified regarding credit rating agencies, we can see why the intersection between the two subjects is important.

The intersection between geopolitical supremacy and AI regulation is growing more substantial by the day. The EU is arguably the first of the big players to formally enact a piece of stand-alone legislation, which some argue puts the EU 'at the forefront of AI Regulation from a global perspective'.[26] It is worth noting however that China has also enacted various formal regulations surrounding AI, but by various administrative bodies within the jurisdiction.[27] However, that leading role on the global stage has not been seamless, with it being remarked that 'much remains to be done to implement it, to promote the responsible use of AI in the financial sector, and to enable European citizens to harness the benefits of AI and the data economy'.[28]

The European Union is not blind to this issue. The AI Act is an integral component of what the EU is calling 'Europe's Digital Decade'. As Keller et al. discuss, there is a desire within the leadership of the EU to 'ensure that AI "puts people first" [and, as a result, its] recently published approach to AI expresses clear concerns over the unrestrained development of AI applications and their risks, and has elected "trustworthy AI" as one of the key pillars of its AI policy'. This pillar is adjoined to the second pillar, namely 'excellence in AI'.[29] This dualised pillar format is coming to define the European approach and it contrasts with its competitors. Discussing the options for the UK in its post-Brexit phase, Edwards summarises the state of play on the world stage:

> For the UK, choices may be presented between the EU model of 'trustworthy AI' rooted in a tradition of strong consumer law protection, human rights and 'ethics',7 versus competing notions from the US and China, the latter especially offering tempting outcomes for developers because of a lower bar for personal data collection and human rights protection. China itself seems to be shifting, on paper at least, to a more regulated model for data and AI, although the aim may be more to protect the state from the power of its own tech platform sector than to protect individual rights.
> These political shifts may make it even more likely that the EU Act becomes an acceptable global model. Paradoxically, post-Brexit the UK may be drawn ideologically away from European approaches to regulation, especially in the wake of COVID-19, as we are seeing in current debates over the future of the UK GDPR.[30]

If the EU is to persevere with this approach, it is important to understand that the two pillars may not be as complementary as they first seem. It is highly likely that one pillar focusing on trustworthiness could take away from the pillar focusing on excellence, and vice-versa.[31] Finding that balance between the two is the real test for the Union and, for many onlookers, it is not passing the test.

The first shortcoming to mention, that scholars suggest is by design, is that that the AI Act does not really focus on the financial sector at all. The only real direct focus is in relation to the risk of discrimination created by AI 'that evaluate the creditworthiness of natural persons ("algorithmic credit scoring")'.[32] Although the EU has acknowledged the obvious impact of AI upon the financial sector, the lack of direct focus in the AI Act has been attributed to a conscious effort to balance the trade-offs between excellency and trustworthiness by deploying two particular principles: technological neutrality and proportionality.

The AI Act only really prohibits and mainly focuses on 'high-risk AI' like facial recognition, etc.[33] When it comes to sector-based risks—say, those emanating from how the credit rating agencies may integrate AI technologies—the EU is 'hedging its bets'. It is taking a risk-based approach that is based on the understanding that it already has extensive frameworks across many sectors that capture the *outputs*, or the resulting risks, from a given sector's engagement with AI.[34] In its initial proposal for the Regulation, the Commission was abundantly clear that its aim was not to be 'overly prescriptive'.[35] This was to be classified via a hierarchical understanding of risk (Fig. 4.3).

This classification reveals an underlying philosophy employed by the European Union. Essentially, 'the result is a regime that prohibits only a limited number of AI practices, and that only imposes additional requirements and obligations on those AI systems that are considered by the EU to be "high risk" (and on participants in the production and distribution chains of those systems – all the way down to final users)'.[37] When we

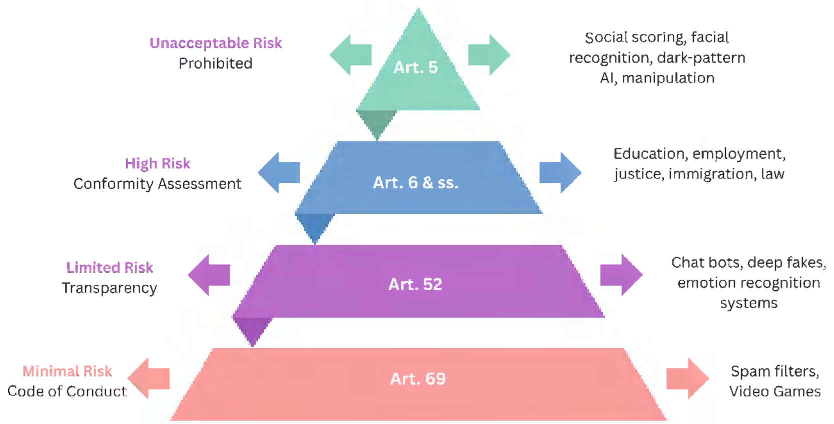


**Fig. 4.3** A 'Risk-based' approach (*Source* Adapted from Lilian Edwards, *The EU AI Act: A Summary of its Significance and Scope* [2022] The Ada Lovelace Institute)[36]

consider what 'high-risk' actually looks like in this taxonomy, it is classified as being dependent on its use (Article 6) and on its potential for posing a risk of harm to the health and safety, or a risk of adverse impact on fundamental rights of individuals (Article 7). Ultimately, as recited in Recital 10, the main focus is on the 'protection of individuals'.

A case could be made for any financial sector having the ability to cause harm that could be classified as 'high-risk' under the EU's taxonomy and, at the end of this book, we shall make that case for the credit rating agencies. By leaving the financial sector—for the most part—undefined according to the taxonomy presented, the potential for intervention from the financial sector to *self-define* the risk they pose is of real concern. It is for this reason that expert onlookers have concluded that 'it is hard to understand why the EU has used its AI Act to address the AI-driven discrimination risks created by a phenomenon like algorithmic credit scoring, while leaving bigger picture systemic risk implications entirely unaddressed… such failure is hard to justify under the principles of technological neutrality and proportionality'.[38] Perhaps, the most important conclusion is that 'it could be that the EU's AI Act actually does a disservice to the goal of mitigating the systemic risk created by AI. Most obviously, the AI Act could contribute to the false notion that the most significant risks created by AI have already been addressed—either in sectoral regulations, in the case of algorithmic trading, or in the AI Act itself, in the case of algorithmic credit scoring—at the same time that AI-driven systemic risk has actually escaped the regulator's radar'.[39]

These potential failures bring forward a critical concept. As Laux et al. note, the concept of *trust* is of the utmost concern for the European Union is rolling out this particular piece of legislation. This is because 'the effort to develop "trustworthy AI" through regulatory laws such as the AI Act acknowledges a need for AI to be trusted if it is to be widely adopted'.[40] It is a growing criticism of the EU that it is conflating notions of trustworthiness and trust with acceptability judgements made by domain experts, with scholars strongly suggesting that the role of technocratic experts may not necessarily be the best strategy to develop trust among the public. The EU have been accused of 'overselling its regulatory ambition and oversimplifying a highly complex and heterogeneous set of closely related concepts' but, perhaps more notably, that the EU is facing stringent criticism for 'outsourcing decisions' about which risks are 'acceptable' to those who stand to profit from the wide integration of AI technologies.[41]

The EU is rather quiet on what constitutes 'acceptability' of risks—which is leading to accusations of outsourcing—but the High-Level Expert Working Group that was convened to assist the Commission did outline seven key requirements for generating 'trustworthy AI':

1. There must be human agency and oversight.
2. There must be technical robustness, and safety must be a priority.
3. Privacy and data governance must be considered at every juncture.
4. Transparency is paramount.
5. Diversity is a key requirement when assembling teams to design AI technology. Also, Non-Discrimination and Fairness must be cornerstones of the inputs and outputs from the technology.
6. Societal and Environmental wellbeing are critically important.
7. All technology and their underlying processes must provide measures for accountability.[42]

Edwards notes how the AI Act's leaning towards 'co-regulation' (which we will cover next—essentially, the State's endorsement of private self-regulation) and 'rulemaking by technical standard-setting bodies operating outside of normal democratic processes' makes it 'difficult for civil society and users to engage with'.[43] Laux et al. take this further and suggest that one of the key tenets of building trust for AI within the public is involving public bodies in the process of rulemaking. Yet, the EU has not chosen that path. Rather:

> Under the AI Act, AI developers will predominantly assess the acceptability of AI-specific risks and thus the trustworthiness of AI. They are deemed to be better positioned in terms of expertise than a public authority. This will at least be the case until an external AI auditing infrastructure has emerged, able to certify compliance with the AI Act. However, for a significant number of high-risk AI systems, the developers will remain free to choose whether they want to rely on internal controls or involve a third-party auditor. Either way, the AI Act thus largely follows a paternalistic approach.[44]

The scholars use the seven stages of building trustworthiness, as articulated by the Expert Working Group, to interrogate this reality of the EU taking a paternalistic approach and conclude that there is a 'normative tension' between the two pillars of trustworthiness and acceptability.

The level of expertise seemingly required to understand the technology is seemingly at odds, at least for the EU, with creating a fully transparent framework that the public can engage with. The danger that comes with that approach is evident, especially when we consider the responses of ESMA in the previous chapter when we were introduced to the credit rating agencies' views on the regulator becoming more involved in how they integrate AI technologies: the credit rating agencies responded by telling the regulator they needed to see how the AI Act would be adopted in reality, and the regulator agreed and took no further action. But, in the real-world, the AI Act places no real emphasis on the credit rating agencies' business. The emphasis lies with ESMA to proactively identify and counter any risk, but it appears to be waiting for the AI Act to mandate it. The cyclical nature to this situation is dangerous because, ultimately, there is a gross abdication of responsibility brewing.

Gamito and Marsden also note that if technocratic rule is promoted—where technical expertise is prioritised over democratic principles—then it makes it extremely hard to undo whatever is developed *ex post*.[45] This brings to a critical juncture. The question of *how* the regulation of the credit rating agencies with respect to their integration of AI technologies may be deployed will define for us the prospective chances of proactive intervention in this space which, ultimately, this book advocates for. To answer that question, we need to look more closely at the concept of co-regulation.

## 4.4 Co-Regulation

At its core, the concept of 'co-regulation' is commonly placed alongside 'self-regulation' as 'forms of interaction between Community processes and private actors'.[46] Co-regulation, as separate from self-regulation, describes 'the direct involvement of a public actor in the regulatory process, which is usually not the case with self-regulation'. It has also been referred to as the output of 'top-down' regulation, whereby the Community (say, the EU) first sets the legal framework and then stakeholders fill in the details of how those legal rules will be applied. Usually, the Community in question will then monitor the observance with the overarching aims of the legislation.[47] John Braithwaite, an early pioneer in the study of these forms of control, refers to co-regulation as 'enforced self-regulation'. He explains that co-regulation falls between the two endpoints and presents the best-of-both-worlds. Not only do regulators, who have scant resources usually, now have the ability to focus on only those who do wrong under a co-regulatory model (instead of enforcing the rules across the whole sector), but the lack of authority that comes with voluntarism is simply removed.[48] The formal structure of the co-regulatory method is indeed compelling:

> The concept of enforced self-regulation is a response both to the delay, red tape, costs, and stultification of innovation that can result from imposing detailed government regulations on business, and to the naivete of trusting companies to regulate themselves. Under enforced self-regulation, the government would compel each company to write a set of rules tailored to the unique set of contingencies facing that firm. A regulatory agency would either approve these rules or send them back for revision if they were insufficiently stringent. At this stage in the process, citizens' groups and other interested parties would be encouraged to comment on the proposed rules. Rather than having governmental inspectors enforce the rules, most enforcement duties and costs would be internalized by the company, which would be required to establish its own independent inspectorial group. The primary function of governmental inspectors would be to ensure the independence of this internal compliance group and to audit its efficiency and toughness. Such audits would pay particular attention to the number of violators who had been disciplined by each company. Naturally, old-style direct government monitoring would still be necessary for firms too small to afford their own compliance group.[49]

This model, as described by Braithwaite in 1982, makes sense. Bruin also discusses how there are clear benefits to the approach, including the ability to rapidly respond to societal and technological developments.[50] However, it is very much based on a set of circumstances which are not universal. Even a cursory analysis of the applicability to the subject matter we are looking at (and based on what we have already discussed) shows the inapplicability to the model as described. First, the leaning towards technocracy takes 'citizens groups and other interest parties' out of the equation because such technical standard-setters do not ask for comment on guideline setting. Second, the European machinery has not yet 'compelled each company to write a set of rules' which would be approved or sent back for revisions by the regulator. Third, there has been no suggestion that companies will have to establish their own 'independent inspectorial group'. Fourth, there is also no mention by the European machinery of the establishment of formal audits.

Nevertheless, Gamito and Marsden discuss how, in Europe, the European Commission 'pragmatically funds standards and *ex ante* supports co-regulation in cases where the US would *ex post* regulate via competition law'.[51] Since the turn of the century, the European Union has been steadily embarking upon a regulatory revolution that 'increasingly puts emphasis on the use of alternative instruments or on instruments that are complementary to traditional command-and-control legislation'.[52] Especially in terms of artificial intelligence, the EU is taking a standard-setting approach and is attempting to utilise its inner machinery to participate in the co-regulatory approach:

> Besides, the legislator can also, and frequently does, not only delegate law-making to technical standard-setting bodies but also entrust the enforcement of legislative requirements to the compliance with the standards produced by the delegated body. Pursuant to the AI Act, compliance with harmonized standards or common specifications leads to a presumption of conformity with the essential requirements, simplifying the compliance process for providers placing AI products on the EU market. The post-market monitoring system consists of a conformity assessment procedure performed by 'notified bodies', also called 'third-party conformity assessment' bodies. Interestingly, in addition the ESOs will be in charge of developing standardization deliverables providing procedures and processes for conformity assessment activities related to AI systems and quality management systems of AI providers as well as the criteria for assessing the competence of persons tasked with the conformity assessment activities. ESOs are already working on developing AI-related standards that will most likely become harmonized standards, regardless of how the final Regulation may look.[53]

The ESOs referred to above are the European Standardisation Organisations and there are three primary players: the European Committee for Standardisation (CEN), the European Committee for Electrotechnical Standardisation (CENELEC) and the European Telecommunications Standards Institute (ETSI). Their role is to develop and publish technical standards that are then applied across the Union, to ensure uniformity across Member States for a variety of technical products and processes. The Commission chose to ask only CEN and CENELEC to develop AI-related standards in association with the AI Act.[54] 'The two bodies came together in 2021 to launch a Joint Technical Committee and which has several objectives, including: advancing EU legislation, policies, principles and values; Considering the adoption of relevant international standards; Providing guidance to other technical committees on AI-related matters; Identifying specific standardisation needs in the European context; and monitoring potential changes in European legislation'.[55] According to the Work Programme of the Joint Committee, they have a large number of projects due to go to voting stages in 2026, including Frameworks for Systems using Machine Learning, Evaluation Methods for accurate Natural Learning processing systems and AI System Logging, among others.[56] However, the largest issue with this approach is that of representation. Such standard-setting committees are often staffed with industry stakeholders and civil society groups often find it difficult to permeate such mechanisms. This has been acknowledged by the EU, as it has mandated CEN and CENELEC to include SMEs and civil society groups in its deliberations, but how that may look, how it may be checked, and whether it may be effective has been doubted.[57]

In addition to this role of the ESOs, it is clear from the EU AI Act that self-regulation is an important part of the co-regulatory approach. It has been noted that 'Codes of Conduct in the drafting of the AI Act appear to have an even more powerful self-regulatory compliance function, especially for non-high-risk AI systems',[58] while Bruin discusses how Article 69 of the Act encourages the Union and Member States to 'facilitate the drawing up of codes of conduct intended to foster the voluntary application to ASI-systems other than high-risk AI systems'.[59] However, this integral part of the co-regulatory approach is beset with innate conflicts. The power and resources of those *within* the sector dwarfs that of those outside. The power imbalance of say, the Big Three credit rating agencies against civil society groups, is pertinent. The real question then becomes how the democratic process can permeate what is becoming (by way of European legal design) a very technocratic process. It is also the case that previous attempts to allow the technical world to self-regulate and then monitor, i.e. co-regulation, have not turned out well at all.[60] Addressing this power imbalance is the ultimate test for the European Union (and other jurisdictions) but, seemingly, the EU is not winning the battle.

## 4.5 Conclusion

Identifying that your jurisdiction needs to take action in a particular space is one thing, but the real crux lies on *how* you take that action. It is abundantly clear that AI needs to be constrained by regulation, but the method that is chosen may define the outcome of the relationship with the novel technology. For the European Union, the constant but delicate battle that exists within the bloc to maintain authority but acknowledge and promote the individuality of its Member States has led to an approach whereby paternalism is reigning supreme.

Paternalism makes sense in this arena for a lot of reasons. The technical capacity needed to even understand the world of AI, never mind seek to amend it, is significant. For the most part, the general public do not have such a capacity. Civil society groups, often promoted as a key democratic tool, do not have the technical capacity either. Perhaps, for that reason, they have effectively been locked out of the process so the EU can prioritise getting its regulatory framework technically accurate. However, some of the key principles that form the foundation of 'trustworthiness' require the EU to take a more collaborative approach. Such paternalism prevents the public from seeing the nuts and bolts of the decision-making process, which is critical. If things go badly wrong, which they very well might in such a highly technical arena, the public will demand answers. In this current approach being adopted by the EU, who are early pioneers in building such a formal regulatory framework, those answers will not be forthcoming.

This is the balancing act the European leadership face. It is critical that they do not let this arena develop into something they cannot control, but the manner in which they are pursuing that objective means, if it does cause catastrophe, it is they who will be punished. It is questionable whether the European Union's constitution is strong enough to withstand such pressure. The EU, in this sense, is likely damned if it does and damned if it does not. On the geopolitical scene, there are different tides. The United States is embarking upon a seemingly uninhibited pursuit of development, while the Chinese have seemingly understood the need to constrain; the State of China also walks a delicate tightrope in terms of authority.

Senden talks of the reality on the ground. She discusses how regulators are usually keen advocates of the concept of co-regulation because it allows for the necessary flexibility to achieve regulatory objectives with the target sector onside, rather than battling against them.[61] Whether or not this approach is *right* is of more concern though. The State, as an entity, is facing a species-defining moment and how each State approaches the task will ultimately define the outcome. The European Union has said, in its words, that the protection of the individual is of paramount importance. The ultimate question is whether that objective translates into action.

## Notes

1. Marta Cantero Gamito and Christopher T. Marsden, 'Artificial intelligence co-regulation? The role of standards in the EU AI Act' (2024) 32 International Journal of Law and Information Technology 1 9.
2. Abraham L Newman and David Bach, 'Self-regulatory trajectories in the shadow of public power: Resolving digital dilemmas in Europe and the United States' (2004) 17 Governance 3 388.
3. Gary Gensler, *AI, Finance, Movies, and the Law – Prepared Remarks Before the Yale Law School* (2024) https://www.sec.gov/newsroom/speeches-statements/gensler-ai-021324.
4. Regulation (EU) 2024/3005 https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32024R3005#:~:text=Page%201-,REGULATION%20(EU)%202024/3005%20OF%20THE%20EUROPEAN%20PARLIAMENT%20AND,(1).
5. ESMA, *Public Statement on the use of Artificial Intelligence (AI) in the provision of retail investment services* (2024) https://www.esma.europa.eu/sites/default/files/2024-05/ESMA35-335435667-5924__Public_Statement_on_AI_and_investment_services.pdf.
6. Ibid.
7. ESMA, *Consultation Paper: Guidelines on Internal Controls for Benchmark Administrators, Credit Rating Agencies and Market Transparency Infrastructures* (2024) https://www.esma.europa.eu/sites/default/files/2024-12/ESMA80-1286971524-661_Consultation_Paper_on_Internal_Controls_for_BAs__CRAs_and_Market_Transparency_Infrastructures.pdf.
8. ESMA, *Artificial Intelligence in EU Securities Markets* (2023) https://www.esma.europa.eu/sites/default/files/library/ESMA50-164-6247-AI_in_securities_markets.pdf
9. Ibid. 16.

10. Stewart Patrick, Elizabeth Sidiropoulos, and Erica Hogan, 'Reimagining Global Economic Governance: African and Global Perspectives' (2024) Carnegie Endowment for International Peace (Sept 16) https://carnegieendowment.org/research/2024/09/reimagining-global-economic-governance-african-and-global-perspectives?lang=en.
11. ESMA, *TRV Risk Monitor: ESMA Report on Trends, Risks and Vulnerabilities* (2024) https://www.esma.europa.eu/sites/default/files/2024-01/ESMA50-524821-3107_TRV_1-24_risk_monitor.pdf 36.
12. ESMA, *Consultation Paper: Proposed Revisions to Commission Delegated Regulation (EU) 447/2012 and Annex I of CRA Regulation* (2024) https://www.esma.europa.eu/sites/default/files/2024-04/ESMA84-2037069784-2112_Consultation_Paper_on_Changes_to_Delegated_Reg_447-2012_and_Annex_I_of_CRAR.pdf.
13. ESMA, *Final Report: Technical Advice on Revisions to Commission Delegated Regulation (EU) 447/2012 and Annex I of CRA Regulation* (2024) https://www.esma.europa.eu/sites/default/files/2024-12/ESMA84-2037069784-2196_Final_Report_Amendment_to_Delegated_Regulation_447-2012_and_Annex_I_of_CRA_Regulation.pdf.
14. The European Association of Credit Rating Agencies, *Submission* (2024) https://www.esma.europa.eu/system/files/webform/208050/104092/EACRA_response_ESMA_consultation_CRAR_amendment_final.pdf.
15. Scope Ratings, *Submission* (2024) https://www.esma.europa.eu/system/files/webform/208050/104115/Scope_Ratings_ESMA_Public_Consultation_21_06_2024.docx.pdf.
16. Fitch Ratings, *Submission* (2024) https://www.esma.europa.eu/system/files/webform/208050/104121/Fitch_Response_Annex_1_-_Responses_to_Consultation_Paper_Questions.pdf.
17. S&P Global, *Submission* (2024) https://www.esma.europa.eu/system/files/webform/208050/104070/SPGRE_Response_to_ESMA_Consultation_Paper.pdf.
18. Moody's, *Submission* (2024) https://www.esma.europa.eu/system/files/webform/208050/104078/Moody_s_Ratings_Response_to_ESMA_CP_on_Proposed_Revisions.pdf.
19. KBRS, *Submission* (2024) https://www.esma.europa.eu/system/files/webform/208050/104127/KBRA_Comment_Letter_-_ESMA_Consultation_Paper_on_Proposed_Revisions_to_the_CRAR_and_Delegated_Act_-_21_June_2024.pdf.

20. See: Marc Flandreau, Norbet Gaillard, and Frank Packer, 'To err is human: US rating agencies and the interwar foreign government debt crisis' (2011) 15 European Review of Economic History 495–538; Marc Flandreau and Joanna K Sławatyniec, 'Understanding rating addiction: US courts and the origins of rating agencies' regulatory licence (1900–1940)' (2013) 20 Financial History Review 3 237–57; Marc Flandreau and Gabriel G Mesevage, 'The separation of information and lending and the rise of the rating agencies in the USA (1841–1907)' (2014) 62 Scandinavian Economic History Review 3 213–242; Marc Flandreau and Gabriel G Mesevage, 'The Untold History of Transparency: Mercantile Agencies, the Law, and the Lawyers (1851–1916)' (2014) 15 Enterprise and Society 2 213–51; Daniel Cash, *A Modern Credit Rating Agency: The Story of Moody's* (Routledge 2024).
21. House of Lords, *Large Language Models and Generative AI* (2024) https://publications.parliament.uk/pa/ld5804/ldselect/ldcomm/54/54.pdf.
22. Ibid. 19.
23. Ibid. 74.
24. Directorate-General for Communication, 'AI Act enters into force' (2024) European Union (Aug 1) https://commission.europa.eu/news/ai-act-enters-force-2024-08-01_en#:~:text=News%20article,and%20financial%20burdens%20for%20businesses.
25. Regulation (EU) 2024/1689.
26. Fausto Parente, 'The AI Act and its impacts on the European Financial Sector' (2024) Eurofi Magazine (Feb) https://www.eiopa.europa.eu/document/download/5dc730b7-29b6-44dd-819b-3a04476416ed_en?filename=ai-act-fausto-eurofi-magazine_ghent_february-2024%20128.pdf.
27. Amy Yang and Bob Li, 'AI Watch: Global Regulatory Tracker – China' (2024) White & Case (May 13) https://www.whitecase.com/insight-our-thinking/ai-watch-global-regulatory-tracker-china.
28. Parente (n 26).
29. Anat Keller, Clara Martins Pereira, and Martinho Lucas Pires, 'The European Union's Approach to Artificial Intelligence and the Challenge of Financial Systemic Risk' in Henrique Sousa Antunes, Pedro Miguel Frietas, Arlindo L. Oliveira, Clara Martins Pereira, Ela Vaz de Sequeira, and Luis Barreto Xavier, *Multidisciplinary*

*Perspectives on Artificial Intelligence and the Law* (Springer 2024) 425.
30. Lilian Edwards, *The EU AI Act: A Summary of its Significance and Scope* (2022) Ada Lovelace Institute (April) https://www.adalovelaceinstitute.org/wp-content/uploads/2022/04/Expert-explainer-The-EU-AI-Act-11-April-2022.pdf.
31. Keller et al. (n 29) 426.
32. Ibid. 428.
33. Edwards (n 30) 7.
34. Keller et al. (n 29) 429.
35. European Commission, *Proposal for a Regulation of the European Parliament and of the Council: Laying Down Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act) and Amending Certain Union Legislative Acts* (2021) https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:52021PC0206.
36. Edwards (n 30) 8.
37. Keller et al. (n 29) 430.
38. Ibid. 432.
39. Ibid.
40. Johann Laux, Sandra Wachter, and Brent Mittelstadt, 'Trustworthy artificial intelligence and the European Union AI Act: On the Conflation of Trustworthiness and Acceptability of Risk' (2024) 18 Regulation & Governance 3–32 5.
41. Natalie A Smuha, Emma Ahmed-Rangers, Adam Harkens, Wenlong Li, James MacLaren, Riccardo Piselli, and Karen Yeung, 'How the EU can achieve legally trustworthy AI: A response to the European Commission's Proposal for an Artificial Intelligence Act' (2021) SSRN (Aug 31) https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3899991.
42. The European Commission's High-Level Expert Group on Artificial Intelligence, *Draft Ethics Guidelines for Trustworthy AI* (2018) https://ec.europa.eu/newsroom/dae/document.cfm?doc_id=57112.
43. Edwards (n 30) 6.
44. Laux et al. (n 40) 7.
45. Gamito and Marsden (n 1) 18.
46. Linda Senden, 'Soft Law, Self-Regulation and Co-Regulation in European Law: Where do they meet?' (2005) 9 Electronic Journal of Comparative Law 1 11.

47. Ibid.
48. John Braithwaite, 'Enforced Self-Regulation: A New Strategy for Corporate Crime Control' (1982) 80 Michigan Law Review 7 1466, 1470.
49. Ibid.
50. Roeland W. de Bruin, 'Co-regulation and AI-innovation: Principles for a Sustainable Framework Fostering Innovation and Acceptance of AI' (2024) SSRN https://papers.ssrn.com/sol3/Delivery.cfm?abstractid=4603551 4.
51. Gamito and Marsden (n 1) 8.
52. Senden (n 46) 1.
53. Gamito and Marsden (n 1) 11.
54. Ibid. The scholars remark that this decision was 'unsurprising'.
55. CEN, *About the Joint Technical Committee* (2025) https://jtc21.eu/about/.
56. CEN, *CEN/CLC/JTC 21 – Artificial Intelligence: Work Programme* (2025) https://standards.cencenelec.eu/dyn/www/f?p=205:22:0::::FSP_ORG_ID,FSP_LANG_ID:2916257,25&cs=1827B89DA69577BF3631EE2B6070F207D.
57. Bruin (n 50) 6.
58. Gamito and Marsden (n 1) 16.
59. Bruin (n 50) 3.
60. Gamito and Marsden (n 1) 17.
61. Senden (n 46) 20.

CHAPTER 5

# Recommendations

**Abstract** The book reaches the point of providing reasonable recommendations for the regulatory framework as it increasingly faces the challenge of a credit rating sector rapidly adopting AI technologies. Recommendations are presented at the supra-national, regional, national and community levels so that the framework is as relevant as it can be.

**Keywords** Policy Reform · IOSCO · Disclosure · Civil society engagement · Self-regulation · Technological standards

The nascent stage of the journey for the credit rating agencies' integration of AI technologies means that it is difficult to present solidified recommendations that could be enacted. It is likely that many variables will be introduced, amended or discarded as the industry continues to walk the path ahead of it. Also, how the regulatory framework will adapt is often dependent on a range of variables which will shape what is considered appropriate, required or even possible.

Yet, the analysis that has been undertaken in earlier chapters does show us a variety of themes. It also shows us who the major players may be and what their institutional capabilities are. There are also aspects of the traditional regulatory framework that exist around the credit rating agencies that need to be considered. There will not be new regulatory players



D. Cash and N. Tkachenko, *The Risk of Artificial Intelligence in Credit Ratings*, https://doi.org/10.1007/978-3-031-95543-3_5

entering the space for credit rating agencies. Instead, their current regulators will likely be given new tasks, in time, to also consider the impact and effect of AI technological integration. With these aspects in mind, we offer four layers of recommendations. First, at the international level, then the regional and then national levels, and lastly at the community level.

## 5.1 Recommendation 1—Action by IOSCO

The International Organisation of Securities Commissions (IOSCO) can be thought of as a 'starting pistol' for regulatory movement. Its proclamations and advice will often become a 'trigger' for jurisdictions to act.[1] The cost, complexity and sheer deployment of resources to start the regulatory machinery is often too much for a jurisdiction to justify, so such triggers are often necessary.

In the field of credit rating agencies, IOSCO has played a considerable part. IOSCO had developed a first-of-its-kind Code of Conduct for credit rating agencies in 2004, which sought to instil standards relating to the quality of the rating process, independence and avoidance of conflict and articulating the responsibilities of the credit rating agencies to the investors and issuers.[2] Of course, the horse had already left the stable at this point and it would take the US and the EU to formally regulate the credit rating agencies for the first time to instil real standards (which we could debate). Yet, this did not stop IOSCO returning with a revised version of its Code in 2008 and, simultaneously, calling for more direct regulation—a key 'trigger' for what was to follow.[3] A year later, the EU had launched the first of three Regulations aimed at the Agencies, and two years after the revised version of the Code of Conduct the US had focused directly on the credit rating agencies within the Dodd-Frank Act.

With this influence duly noted, the recommendation proposed here is that **IOSCO should develop a Code of Conduct for the integration of AI into credit rating agency services and processes**. Such a Code of Conduct would aim to instil standards relating to transparency, communication strategies, public awareness, monitoring mechanisms and more. We will return to this concept of a Code of Conduct again later when considering the role of civil society but it is worth noting that, for the EU AI Act, Codes of Conduct are expressly encouraged for AI systems that are not deemed to be 'high-risk'.[4] It is worth noting that not every example of encouraging an industry or sector to produce Codes of Conduct—essentially self-regulation—has been a success.[5] However, there are a number of advantages when applied to the credit rating sector:

- The early stage of development means that a Code of Conduct can allow for flexibility.
- A Code of Conduct increases 'buy-in' from participants.
- A Co-Regulatory approach can begin on a stronger footing with the presence of a Code of Conduct.
- Different stakeholders can engage with the sector in question to help mould a Code of Conduct.

For these reasons, the applicability of a Code of Conduct for the credit rating agencies' integration of AI is clear. Whether it would be appropriate for all parties would remain to be seen. However, the authority and influence that IOSCO commands means that it, as the generator of a Code of Conduct, may stand the best chance of bringing the different parties together.

## 5.2 Recommendation 2—ESMA Needs to Be More Assertive

The philosophical approach of the European Union in relation to AI integration is stunningly obvious. The research is almost unanimous that the Union is seeking to balance the need for individual safety with both the benefits of commercial expansion *and* the geopolitical jostling for position that it consistently finds itself in. Those competing pressures have resulted in an AI Act that says very little on the financial services industry, one of the main prospective vehicles for social impact in relation to AI integration. We saw, from the Consultation that ESMA produced on changes to the credit rating regulations, that there is currently a cycle being formed that is not positive. The AI Act seemingly leaves a lot of discretion for sector regulators to interpret development, but then the sector regulators are announcing that they are happy to wait for more guidance from the Act. Market participants, like the credit rating agencies, are then declaring that they are not happy to get involved with any developments until the impact of the AI Act is made clear. Something has to give.

It is acknowledged that the AI Act is new. However, the co-regulatory approach that it clearly adopts requires more action. ESMA, instead of bolting a critical question on the credit rating agencies' views and understandings on AI integration *to the end* of a survey perhaps indicates how far away ESMA are from where they need to be. It is telling that many credit rating agency responded partially chastised the regulator for this approach and insisted that such important topics be given their due care and attention. This is our recommendation. **ESMA must issue a new consultation and report on how the credit rating agencies may integrate AI technologies, and their views on the impact/effect of the AI Act.** Such an instrument will allow the credit rating agencies to articulate what they see as key elements, what they understand to be prospectively impactful, while also confirming for the regulator—who should then produce a report on the back of responses—solely focused on the credit rating agencies' integration of AI technologies.

There is also an important aspect that has gone under the radar. Novelli et al. discuss how liability needs to be a critical supplement to AI regulation. 'As the authors mention, inadequate rules of liability allocation may increase LLMs' risk and may, in turn, cause the risk of a breakdown of the AI market'.[6] This will be a critically important aspect for the credit rating agencies to consider and it is highly likely that, internally, they have been considering this angle for quite some time. A formal discussion needs to be had about how the liability rules that already apply to the credit rating agencies may be applicable, or not, to the new world in which credit rating agencies are actively incorporating AI technologies.

## 5.3 Recommendation 3—Civil Society Needs to Do More

Civil society, as we saw in Chapter 4, is at real risk of being pushed out of proceedings. There is no inkling that the US will formally legislate for this intersection between the credit rating agencies and the integration of AI technologies, and the main civil societies have very little impact in China. Therefore, the European arena is the best opportunity for civil society groups to influence the direction of travel. But, as we saw with the way that the European Union has constructed its AI-focused regulatory framework, standard-setting and technical guidance is coming from areas of the European framework which owe no obligation to the civil society sector. Even though the EU has told the standard-setting bodies to include civil society in its processes, there is no detail on how, when and what happens if they choose not to.

The sentiment underlying this lack of presence is that this field is too complicated for civil society. On many levels, this is true. Civil society groups are not designed to be highly technical and engage on high-level technical matters. They are, however, key democratic institutions that have a special power to harness the expertise of many in relative fields who *do* have the relevant knowledge, whether that comes from academia, past-practice or elsewhere. It is fundamentally inappropriate to build standards for such societally-critical products and services *without* the consistent and substantial interventions of civil society. This is not included because the authors work for, or are even advocates of civil society. It is included because, quite simply, if 'trustworthiness' is widely recognised as a necessary ingredient in the successful societal adoption of AI technologies, then *how* those technologies are regulated and standardised is critically important. Therefore, **we strongly recommend that civil society groups deploy more resources to involving themselves in standard-setting processes in Europe, and then globally**. The social threat being posed by AI integration is too strong to allow only industry professionals access to fora where standards are being decided.

## 5.4 Recommendation 4—It Is Time for Credit Rating Agencies to Be Declared as Systemically-Important

In the UK, the need to create a vibrant economy in the post-Brexit environment is of paramount importance. However, finding the right balance between commercial development and commercial security is proving to be a difficult endeavour. While the Chancellor of the Exchequer is denouncing the level of regulation in the UK,[7] the UK is at the same time launching a new regulatory initiative. It is known as the 'Critical Third-Parties' Regime, or CTPs. Admittedly launched under the previous Johnson Administration, the Treasury announced in June 2022 in a policy statement that it was launching the regime to 'mitigate risks from critical third parties in the financial sector'. The statement mentioned aspects such as the risk of concentration within financial services, and the threats from novel technologies as key targets for the new regime.[8]

At the end of 2024, the Financial Conduct Authority (FCA), the Bank of England (BoE) and the Prudential Regulation Authority (PRA) jointly announced the final requirements and expectations of CTPs as they understood it.[9] The regulators have to nominate which of their regulated entities they believe provide services that are considered critical, at which point the Treasury will formally decide whether to induct the providers into the new regime. The service providers will not see any material difference—they will still have the same regulator and still be exposed to all the same rules as before—but now, if designated, the service providers will be exposed to the potential of the regulator examining *particular processes and services* that it deems as critical to the UK financial sector. The applicability to the world of credit rating agencies in light of their growing integration of AI technologies is clear.

The politics behind this regime is awkward to say the least. The Labour Government seemingly would not have undertaken this approach, and the Conservative Government that came before it left a lot to be desired when it came to protecting British resilience in the financial sector. However, it is going ahead. A lot could be learned from this approach, especially as it is highly likely that the credit rating agencies will be nominated by the FCA to the Treasury. What particular services may be nominated we are yet to see—more details are expected halfway through 2025—but the sentiment is one that we support: **the credit rating agencies should be designated as 'systemically-important' within national jurisdictions**. There may be a variety of ways this can happen across jurisdictions but, as we have tried to show in this work, the credit rating agencies are consistently at the forefront of service provision that can have a massive societal effect. The impact surrounding the Global Financial Crisis will pale in comparison to the impact if the credit rating agencies integrate AI technologies into their processes *and* make mistakes. As such, regulators must be proactive and guard against this happening. This is not to say they should prohibit the credit rating agencies from partaking in the digital revolution, but they must enact stronger frameworks that insist on more testing, more reviews and more transparent communication. What we have seen so far on this front is not enough.

## Notes

1. Daniel Cash and Maha Khan, *Rating the Globe: Reforming credit rating agencies for an equitable financial architecture* (2024) United Nations Centre for Research Policy https://collections.unu.edu/eserv/UNU:9832/rating_the_globe.pdf.

2. IOSCO, *Code of Conduct Fundamentals for Credit Rating Agencies* (2004) https://www.iosco.org/library/pubdocs/pdf/ioscopd173.pdf.
3. IOSCO, *Code of Conduct Fundamentals for Credit Rating Agencies* (2008) https://www.iosco.org/library/pubdocs/pdf/ioscopd271.pdf.
4. Marta Cantero Gamito and Christopher T. Marsden, 'Artificial intelligence co-regulation? The role of standards in the EU AI Act' (2024) 32 International Journal of Law and Information Technology 16.
5. Ibid. 15.
6. Claudio Novelli, Federico Casolari, Antonino Rotolo, Mariarosaria Taddeo, Luciano Floridi, 'Taking AI Risks Seriously: A New Assessment Model for the AI Act' (2023) 39 AI & Society 2493, 2496.
7. Heather Stewart, 'Economists and Policy Experts warn Reeves against City deregulation' (2024) The Guardian (Dec 16) https://www.theguardian.com/business/2024/dec/16/economists-policy-experts-warn-rachel-reeves-city-deregulation.
8. HM Treasury, *Critical Third Parties to the Finance Sector: Policy Statement* (2022) https://assets.publishing.service.gov.uk/media/629f6ce88fa8f5038dcd2904/2022-06-08_critical_third_parties_policy_statement.pdf.
9. FCA, *PS24/16: Operational Resilience: Critical Third Parties to the UK Financial Sector* (2024) https://www.fca.org.uk/publications/policy-statements/ps24-16-operational-resilience-critical-third-parties-uk-financial-sector#:~:text=Next%20steps,will%20oversee%20CTPs%20in%20practice.

CHAPTER 6

# Conclusion

**Abstract** The book concludes with tying together the key themes of the work. It also presents so challenges and questions for future consideration, and insights in the likely direction of travel within the nascent intersection of the credit rating agencies' integration of AI technologies.

**Keywords** Systemic importance · Regulatory inertia · Technological complexity · Economic governance · Strategic oversight · Financial ethics

Before the book concludes, there are some caveats to make. The credit rating agencies have come in for significant criticism since the Global Financial Crisis, and for good reason. However, they are not 'bad actors'. They, as an industry, are not ultimately transgressive in nature (however much onlookers may label them as such). They are, though, economic actors. From such a perspective, the credit rating agencies fall foul of taking calculated gambles that, considering cold, hard financial facts, have proven to be successful decisions. In the GFC, they were essentially guilty of leveraging 'economic rent'.[1]

This leveraging of their systemic positioning is at the core of the focus for this work. There is no inclination, at this early stage of the development of AI integration, that the credit rating agencies are transgressing. However, with the parameters above put into place, the same




D. Cash and N. Tkachenko, *The Risk of Artificial Intelligence in Credit Ratings*, https://doi.org/10.1007/978-3-031-95543-3_6

'cocktail' that resulted in the GFC is starting to appear again. A lax regulatory environment, a systemically important credit rating sector devoid of competition and dominated by a *natural* oligopoly,[2] and now (and yet again) a rapidly increasing rate of (conscious) complexity[3] that makes judging underlying data more difficult. It is not to say that the GFC will be repeated in scale, or even similarity. But we are proposing here that the flashing lights on the dashboard are beginning to appear. Our real question is whether those who are responsible with driving the societal train forward, or at least making sure it does not derail, are even watching the dashboard. Are the warning lights that are beginning to appear being seen? Perhaps, from the perspective of regulators and legislators, the warning lights are so faint that they cannot see them. The analysis undertaken in Chapter 4 suggests that, for the regulators, everything looks fine when they look at the dashboard. We argue here that they are not so feint. They are strong, if you know where on the dashboard to look. We have endeavoured to show that the strength of the light emanating from the dashboard is growing and it is time for regulators and legislators to focus their attention.

The analysis in this book has shown that the applicability of AI technologies is wide and is continuing to grow. That applicability is allowing the credit rating agencies to inject AI technologies—particularly but not exclusively relating to GenAI—not only into their services they offer the marketplace, but more importantly also into their own processes. Of the two, it is the latter that is most concerning. The inner processes of the credit rating agencies are uniquely important to the financial health of the entire world and, when those inner processes are tampered with or altered for profit instead of accuracy, chaos usually ensues. The job of regulators and legislators, then, is to make sure this delicate balance does not quickly disintegrate into one dominated by venal pursuits. For this book, there was a real focus then on the position, understanding and actions of the regulators and legislators.

Chapter 4 of the book showed clearly that the regulators and legislators are in a tight bind. They are tasked with controlling the integration of a technology which is defined by complexity. Additionally, that complex arena has the innate potential of providing significant efficiency gains and, subsequently, economic growth after a long period of stagnation. Legislators and regulators are under pressure to let the money flow. It is on this basis that 'co-regulation' is being advanced in some jurisdictions, and de-regulation is being advanced in others. The theoretical sentiments that underlie the concept of co-regulation make sense; the party that has the technical know-how clearly should be well positioned within the design of frameworks. But, what if the obvious conflict of interest is too great? Legislators and regulators are articulating that they do not believe the conflict of interest is too great, today. We, unfortunately, could not disagree more. The threat that comes with allowing industries like the credit rating industry to determine their own pathway in a space that reduces oversight and control of data ought not to be an option.

However, the reality is that political and legal realities often trump normative conclusions. Because of that understanding, we advanced a range of recommendations. Our focus was on encouraging action. Rather than prohibit the credit rating agencies from doing this or that, or encouraging actions here or there, we instead opted to promote the idea that more attention should be focused on this intersection. The requirement now to be better informed about this coming problem has hopefully been made clear throughout this book. This book endeavoured to be in line with the stage of the journey discussed in the *Preface*. We are the very start of the journey, but to inject the necessary details into the world of the 'passenger', i.e. wider society, new discussions are required. Now, directed assessments and debates will bring issues to the light, and that is critical for building one of the most important facets needed for the successful adoption of AI technologies within society—*trust*.

## Notes

1. Daniel Cash, *Regulation and the Credit Rating Agencies: Restraining Ancillary Services* (Routledge 2018) 140.
2. Ulrich G Schroeter, 'Credit Ratings and Credit Rating Agencies' in Gerard Caprio (Ed), *Handbook of Key Global Financial Markets, Institutions, and Infrastructure* (Elsevier 2013) 387.
3. Cash (n 1) 159.

# Index





D. Cash and N. Tkachenko, *The Risk of Artificial Intelligence in Credit Ratings*, https://doi.org/10.1007/978-3-031-95543-3