The Future of Banking

7

Thierry Foucault, Leonardo Gambacorta, Wei Jiang and Xavier Vives

Artificial Intelligence in Finance

CEPR

IESE Business School University of Navarra Banking Initiative

ARTIFICIAL INTELLIGENCE IN FINANCE

The Future of Banking 7

CEPR PRESS

Centre for Economic Policy Research

187 boulevard Saint-Germain
75007 Paris, France

2 Coldbath Square
London EC1R 5HL, UK

Tel: +44 (20) 7183 8801
Fax: +44 (20) 7183 8820
Email: cepr@cepr.org
Web: www.cepr.org

ISBN: 978-1-912179-93-0

ARTIFICIAL INTELLIGENCE IN FINANCE

The Future of Banking 7

Thierry Foucault
HEC Paris and CEPR

Leonardo Gambacorta
Bank for International Settlements and CEPR

Wei Jiang
Emory University

Xavier Vives
IESE Business School and CEPR

IESE Business School
University of Navarra
Banking Initiative

With the support of Citi

CENTRE FOR ECONOMIC POLICY RESEARCH

CENTRE FOR ECONOMIC POLICY RESEARCH (CEPR)

The Centre for Economic Policy Research (CEPR) is a network of almost 1,900 research economists based mostly in European universities. The Centre's goal is twofold: to promote world-class research, and to get the policy-relevant results into the hands of key decision-makers. CEPR's guiding principle is 'Research excellence with policy relevance'. It was founded in the UK in 1983, where it is a Charity, and in November 2019 CEPR initiated the creation of an Association under French law, in order to provide a vehicle for an expansion in France. The members of the Conseil d'Administration of the Association are identical to the UK Board of Trustees.

CEPR is independent of all public and private interest groups. It takes no institutional stand on economic policy matters and its core funding comes from its Institutional Members, projects that it runs and sales of publications. Because it draws on such a large network of researchers, its output reflects a broad spectrum of individual viewpoints as well as perspectives drawn from civil society. CEPR research may include views on policy, but the Trustees/members of the Conseil d'Administration of the Association do not give prior review to its publications. The opinions expressed in this report are those of the authors and not those of CEPR.

IESE BUSINESS SCHOOL

IESE is the graduate business school of the University of Navarra. Founded in 1958, the school is one of the world´s most international business schools, with campuses in Barcelona, Madrid, Munich, New York and São Paulo. Consistently ranked within the top ten worldwide, IESE Business School has pioneered business education in Europe since its founding. For more than sixty years, IESE has sought to develop business leaders with solid business skills, a global mind-set and a desire to make a positive impact on people and society. The school distinguishes itself in its general-management approach, extensive use of the case method, international outreach, and emphasis on placing people at the heart of managerial decision-making.

About the authors

Thierry Foucault is Chaired Foundation Professor of Finance at HEC Paris and a CEPR Research Fellow. He holds a PhD in Finance from HEC Paris. His research focuses on the role of information in financial markets, the industrial organisation of these markets, and their effect on the real economy. He is co-editor of the *Journal of Financial and Quantitative Analysis* and an Associate Editor of the *Journal of Finance*. He serves on the scientific committees of the Autorité des marchés financiers (AMF) and the Norges Bank Academic Programme. He coauthored Market Liquidity: Theory, Evidence, and Policy (Oxford University Press).

Leonardo Gambacorta is the Head of the Emerging Markets unit at the BIS. Prior to his current role, he served as Head of Innovation and Digital Economy (2019-2024), Research Adviser (2014-2018) and Head of Monetary Policy (2010-12) in the Monetary and Economic Department. Before joining the BIS, he was Head of the Money and Credit Unit (2007-2009) and Head of the Banking Sector Unit (2004-2006) in the Research Department of the Bank of Italy. He was a visiting scholar at the National Bureau of Economic Research (2002–03). Leonardo holds an MSc in Economics from the University of Warwick and a PhD in Economics from the University of Pavia. His primary research interests include monetary transmission mechanisms, the effectiveness of macroprudential policies in curbing systemic risk, and the effects of technological innovation on financial intermediation. He is a Research Fellow of the Centre for Economic Policy Research.

Wei Jiang is Asa Griggs Candler Professor of Finance and Vice Dean of Faculty and Research at Emory University's Goizueta Business School. She is a Research Associate of the NBER. She is the President-Elect of the American Finance Association (AFA) in 2025, and was previously the President of the Society of Finance Studies (SFS). Professor Jiang's main research interests lie in corporate governance, institutional investors, technology and financial markets. She has published extensively in top economics, finance, and law journals, and her research has been frequently featured in major media. She received dozens of awards for research excellence, including the best paper prizes from *Journal of Finance, Review of Financial Studies*, and *Journal of Financial Economics*. Previously she served as an editor of the *Review of Financial Studies,* a Finance Department Editor at *Management Science* and an Associate Editor at the *Journal of Finance*. Professor Jiang received her PhD in Economics from the University of Chicago in 2001, and joined the Finance faculty of Columbia Business School in the same year. She was the Arthur F. Burns Professor of Free and Competitive Enterprise at Columbia Business School until 2022.

Xavier Vives is Chaired Professor of Economics and Finance at IESE Business School. He is a member of the Advisory Scientific Committee of the ESRB (ECB), a Fellow of the Econometric Society since 1992, the European Economic Association since 2004, and the Academia Europaea since 2012. He was President of the European Association for Research in Industrial Economics (EARIE) for 2016-18 and the European Finance Association in 2023. He was Duisenberg Fellow of the European Central Bank in 2015. He has taught at INSEAD, Harvard University, Universitat Autònoma de Barcelona, Universitat Pompeu Fabra, UC Berkeley, University of Pennsylvania, and New York University. His fields of interest are industrial organisation and regulation, banking, and financial economics. He has published more than a hundred articles in the main international journals and several books, the most recent being *Competition and Stability in Banking*. In 2011-2014 he was Special Advisor to the EU Commissioner for Competition, Mr. Joaquin Almunia, and until May 2020 he was Lead Independent Director of CaixaBank. He has served as an advisor for the World Bank, the Inter-American Development Bank, the European Commission, and the Federal Reserve Bank of New York. He holds a PhD in Economics from UC Berkeley.

Acknowledgements

The authors have benefited from the comments of the discussants of the report at the conference on 20 March 2025, Diana Bonfim, Sean Cao, William Cong, Luca Enriques, Ronit Ghose and Robin Lumsdaine, as well as the conference participants and the conference chairs, Luigi Guiso, Rafael Repullo and Victoria Vanasco. We thank the participants of the roundtable, Bonnie Buchanan, Zanna Iscenko and Charles-Albert Lehalle, and particular recognition is due to José Luis Escrivá for his contribution in opening the conference.

Thierry Foucault thanks Markus Bak-Hansen, Maxime Bonelli, Eloi Foucault, and Yurii Handziuk for their comments and Teodor Duevski for research assistance. Leonardo Gambacorta is grateful to Iñaki Aldasoro, Sarah Bell, Giulio Cornelli, Sebastian Doerr, Jon Frost, Fernando Perez Cruz, Han Qiu, Vatsala Shreeti and Hyun Song Shin. Wei Jiang is grateful to Tucker Balch, Zhishuo Han, Tao Li, Will Mann and Baozhong Yang. Xavier Vives thanks Josep Gisbert for extremely useful input on the report and Joan Freixa for his excellent research assistance. The authors also benefited from the papers presented at the IESE Banking Initiative Workshop that preceded the conference. Carlota Monner delivered highly efficient general support.

The views expressed in this report are those of the authors. They should not be taken to represent any institutions with which they are or have been affiliated, or the individuals mentioned above.

Contents

About the authors *v*
Acknowledgements *vii*
Conference programme *x*
List of conference participants *xi*
Workshop programme *xiii*
List of workshop participants *xiv*
Foreword *xvi*

Executive summary 1

The transformational impact of artificial intelligence for the financial sector and central banks 7
by José Luis Escrivá Belmonte

1 Introduction 13
1.1 AI characteristics and uses in finance: What is old and what is new? 15
1.2 Artificial intelligence and the financial sector: Transformations, challenges, and regulatory responses 17
1.3 Data abundance, AI, and financial markets: Implications and risks 23
1.4 Corporate finance and governance with AI: Old and new 30
1.5 Conclusions and policy implications 38
Appendix: Construction of Figure 1 41

2 Artificial intelligence and the financial sector: Transformations, challenges, and regulatory responses 43
2.1 AI in finance: What are the new opportunities? 44
2.2 AI in finance: Old problems, new challenges 55
2.3 How to regulate AI? 65
2.4 Conclusions 68

3 AI's impact on finance: Reshaping information and its consequences 71
3.1 Why is financial information being reshaped by AI? 73
3.2 Implications for the securities industry 83
3.3 Risk 91
3.4 Conclusion: Policy implications 106

4 Corporate finance and governance with artificial intelligence: Old and new 109
4.1 Introduction 109
4.2 Delegation to AI: The agency problem revisited 113
4.3 The changing faces of information and information asymmetry 123
4.5 Conclusion 149

Discussions 150
References *177*

Conference programme

IESE Business School, Barcelona Campus
Friday, 21 March 2025

09:15 **Welcome**
Jordi Canals, IESE
Xavier Vives, IESE
Opening speech: "The transformational impact of artificial intelligence for the financial sector and central banks"
José Luis Escrivá, Governor of the Banco de España

09:45 **AI and the financial sector: Transformations, challenges and regulatory responses**
Leonardo Gambacorta, Bank for International Settlements
Discussant 1: Diana Bonfim, Banco de Portugal, Católica Lisbon and CEPR
Discussant 2: Ronit Ghose, Citi
Chair: Luigi Guiso, Einaudi Institute for Economics and Finance

10:45 *Break*

11:15 **Corporate finance and governance with AI: Old and new**
Wei Jiang, Emory University
Discussant 1: Sean Cao, University of Maryland
Discussant 2: Luca Enriques, Bocconi University
Chair: Rafael Repullo, CEMFI

12:15 **Roundtable**
Bonnie Buchanan, Sustainable and Explainable FinTech Center, University of Surrey
Zanna Iscenko, Google
Charles-Albert Lehalle, École Polytechnique, IP-Paris
Chair: Xavier Vives, IESE

13:15 *Lunch*

14:30 **AI's Impact on Financial Markets: The Consequences of Evolving Information Production**
Thierry Foucault, HEC Paris
Discussant 1: William Cong, Cornell University
Discussant 2: Robin Lumsdaine, American University
Chair: Victoria Vanasco, CREI, UPF, BSE

15:30 **Conclusion**

16:00 *Close of meeting*

List of conference participants

Simona Abis	University of Colorado Boulder
Carmen Ansotegui	ESADE Business School
Tania Babina	University of Maryland
Anna Bayona	ESADE Business School
Vicente Bermejo	ESADE Business School
Diana Bonfim	Banco de Portugal, Católica Lisbon and CEPR
Bonnie Buchanan	SAEF FinTech Center, University of Surrey
Jordi Canals	IESE Business School
Sean Cao	University of Maryland
Madalen Castells	European Central Bank
Che Chen	IESE Business School
Long Cheng	IESE Business School
Jean-Edouard Colliard	HEC Paris
Irem Demirci	Nova School of Business and Economics
Olivier Dessaint	INSEAD
Luca Enriques	Bocconi University
Thiago Fauvrelle	European Stability Mechanism
Enric Fernández	Caixabank
Thierry Foucault	HEC Paris
Joan Freixa	IESE Business School
Xavier Freixas	Pompeu Fabra University
Leonardo Gambacorta	Bank for International Settlements
Teresa Garcia Milà	UPF and BSE
Ronit Ghose	Citi
Javier Gil Bazo	Universitat Pompeu Fabra
Mireia Giné	IESE Business School
Marco Giometti	Universidad Carlos III de Madrid
Josep Gisbert	IE Business School
Luigi Guiso	Einaudi Institute for Economics and Finance
Minfang Guo	IESE Business School
Sandro Harz	F mal s GmbH / INEOS
Zanna Iscenko	Google LLC
Lorenzo Isla	Caixabank
Yichuan Jia	IESE Business School
Wei Jiang	Emory University
Binghan Jiang	PSL-Dauphine
Artashes Karapetyan	ESSEC
Enisse Kharroubi	Bank for International Settlements
Andrei Kirilenko	University of Cambridge
Gael Le Mens	Universitat Pompeu Fabra

Mayte Ledo	Banco de España
Charles-Albert Lehalle	Ecole Polytechnique, IP-Paris
Gyoengyi Loranth	University of Vienna
Jesús Lozano	BBVA
Robin Lumsdaine	Kogod School of Business, American University
Sameh Marei	ESADE Business School
Montserrat Martinez Parera	Bank of Spain
Carmen Matutes	Waveform Investments
Núria Mas	IESE Business School
Anna Maymús	Banco Sabadell
Roxana Mihet	University of Lausanne & SFI
Marta Pla	Banc Sabadell
Isaac Ré Delgado	Santander Asset Management
Rafael Repullo	CEMFI
Martin Rohar	European Stability Mechanism
Javier Santomá	IESE Business School
Stefano Schiaffi	Bank of Italy
Jörg Stahl	Univerisdade Católica Portuguesa
Gabriela Stockler	UPF and BSE
Javier Suarez	CEMFI
Andrew Sutton	GovAI / London Initiative for Safe AI (LISA)
Leonardo Tariffi	Universidad de Barcelona
Tammaro Terracciano	IESE Business School
Jaume Torrents	Caixabank
Patrick Trezise	Banco Sabadell
Antoine Uettwiller	Queen Mary University of London
Victoria Vanasco	CREI, UPF, Barcelona GSE
Elena Vardon	Wall Street Journal / Dow Jones
Xavier Vives	IESE Business School
Michelle Wallin	IESE Business School
Haorui Wang	IESE Business School
Hao Yang	Swiss Finance Institute and USI Lugano
Liyan Yang	University of Toronto
Zhiqiang Ye	Zhejiang University
Jiamin Zhao	IESE Business School

Workshop programme

IESE Business School, Barcelona Campus
Thursday, 20 March 2025

09:15	**Opening Remarks**
09:30	**First Session**, Chair: Xavier Vives
	1. "AI-Powered Trading, Algorithmic Collusion, and Price Efficiency" (with I. Goldstein and Y. Ji)
	Presenter: Winston Dou, Wharton School
	Discussant: Hao Yang, Swiss Finance Institute and USI Lugano
	2. "Algorithmic Pricing and Liquidity in Securities Markets" (with T. Foucault and S. Lovo)
	Presenter: Jean-Edouard Colliard, HEC Paris
	Discussant: Liyan Yang, Rotman School of Management
11:00	*Break*
11:30	**Second session**, Chair: Thierry Foucault, HEC Paris
	1. "Artificial Intelligence and Firms' Systematic Risk" (with A. Fedyk, A. He and J. Hodson)
	Presenter: Tania Babina, Columbia Business School
	Discussant: Simona Abis, University of Colorado Boulder
	2. "Privacy Policies and Consumer Data Extraction: Evidence from US firms'" (with T. Ramadorai and A. Walther)
	Presenter: Antoine Uettwiller, Queen Mary University of London
	Discussant: Mireia Giné, IESE
13:00	*Lunch*
14:00	**Third session**, Chair: Wei Jiang, Emory University
	1. "Generative AI and Firm Values" (with A. Eisfeldt and M. B. Zhang)
	Presenter: Gregor Schubert, UCLA Anderson
	Discussant: Olivier Dessaint, INSEAD
	2. "Machine Learning About Venture Capital Choices" (with V. Lyonnet)
	Presenter: Léa H. Stern, University of Washington
	Discussant: Ramana Nanda, Imperial College London
15:30	*Break*
16:00	**Fourth session**, Chair: Leonardo Gambacorta, BIS
	1. "Data Innovation Complementarity and Firm Growth" (with A. Fedyk, O. Gomes and K. Rishabh)
	Presenter: Roxana Mihet, University of Lausanne
	Discussant: Enisse Kharroubi, BIS
	2. "Artificial intelligence and relationship lending" (with L. Gambacorta and F. Sabatini)
	Presenter: Stefano Schiaffi, Bank of Italy
	Discussant: Xavier Freixas, Universitat Pompeu Fabra
17:30	*Close of meeting*

List of workshop participants

Simona Abis	University of Colorado Boulder
Carmen Ansotegui	ESADE Business School
Tania Babina	University of Maryland
Anna Bayona	ESADE Business School
Vicente Bermejo	ESADE Business School
Diana Bonfim	Banco de Portugal, Católica Lisbon and CEPR
Bonnie Buchanan	SAEF FinTech Center, University of Surrey
Hector Calvo-Pardo	University of Southampton
Sean Cao	University of Maryland
Madalen Castells	European Central Bank
Che Chen	IESE Business School
Long Cheng	IESE Business School
Jean-Edouard Colliard	HEC Paris
Irem Demirci	Nova School of Business and Economics
Olivier Dessaint	INSEAD
Winston Dou	Wharton School
Christian Eufinger	IESE Business School
Thiago Fauvrelle	European Stability Mechanism (ESM)
Thierry Foucault	HEC Paris
Joan Freixa	IESE Business School
Xavier Freixas	Pompeu Fabra University
Leonardo Gambacorta	Bank for International Settlements
Javier Gil Bazo	Universitat Pompeu Fabra
Mireia Giné	IESE Business School
Marco Giometti	Universidad Carlos III de Madrid
Josep Gisbert	IE Business School
Sandro Harz	F mal s GmbH / INEOS
Yichuan Jia	IESE Business School
Wei Jiang	Emory University
Binghan Jiang	PSL-Dauphine
Enisse Kharroubi	Bank for International Settlements
Dmitry Kuvshinov	Universitat Pompeu Fabra
Gael Le Mens	Universitat Pompeu Fabra
Yang Lou	IESE Business School
Jesús Lozano	BBVA
Robin Lumsdaine	Kogod School of Business, American University
Emanuele Luzzi	Università Svizzera Italiana | Swiss Finance Institute
Sameh Marei	ESADE Business School

Carmen Matutes	Waveform Investments
Roxana Mihet	University of Lausanne & SFI
Ramana Nanda	Imperial College London
Jose Penalva	Universidad Carlos III
Marta Pla	Banc Sabadell
Isaac Ré Delgado	Santander Asset Management
Rafael Repullo	CEMFI
Marta Riveira	Banc Sabadell
Martin Rohar	European Stability Mechanism
Kristine Sahakyan	ESCP Business School
Stefano Schiaffi	Bank of Italy
Gregor Schubert	UCLA Anderson School of Management
Jörg Stahl	Univerisdade Católica Portuguesa
Gabriela Stockler	UPF and BSE
Javier Suarez	CEMFI
Andrew Sutton	GovAI / London Initiative for Safe AI (LISA)
Leonardo Tariffi	Universidad de Barcelona
Tammaro Terracciano	IESE Business School
Antoine Uettwiller	Queen Mary University of London
Victoria Vanasco	CREI, UPF, Barcelona GSE
Xavier Vives	IESE Business School
Michelle Wallin	IESE Business School
Haorui Wang	IESE Business School
Fangming Xu	University of Bristol
Hao Yang	Swiss Finance Institute and USI Lugano
Liyan Yang	University of Toronto
Zhiqiang Ye	Zhejiang University

Foreword

This is the seventh report in the series on The Future of Banking, part of the Banking Initiative from the IESE Business School that was launched in October 2018 and is supported by Citi.

The goal of the IESE Banking Initiative is to establish a group of first-rate researchers to study new developments in banking and financial markets, paying particular attention to regulation and competition policy and to the impact on business banking models and the performance of markets. It aims to promote a rigorous and informed dialogue on current issues in the fields of banking and financial markets amongst academics, regulators, private sector companies and civil society.

The first report, published in 2019, assessed the regulatory reform of the banking system after the Great Recession induced by the global financial crisis of 2008-2009, and suggested that the next global crisis might have different origins, possibly in entities that perform the functions of banks but are outside of the regulatory perimeter, or in an emerging market where regulation could well be different from the reformed patterns of the West. It concluded that the system had been made more resilient but that further work remained to be done.

The second report addressed the changes in the business models of banks and identified that the challenges that banks faced in the pre-COVID-19 world – mainly low interest rates and digital disruption – will be made more severe in the post-COVID world. Banks have had to deal with an increase in non-performing loans, albeit with temporary relief from strict regulation and with massive liquidity help from central banks. This has accelerated restructuring in the sector.

The third report studied how climate and natural disaster risk is different from other, more familiar forms of financial and economic risk and how banks, asset managers and central banks are beginning to grapple with these risks. COVID-19 has made us aware of the potentially devastating effect of natural disasters and provides a pointer to the effects that climate change may induce. At the same time, the COVID crisis provided a large-scale natural experiment to address this question, and put natural disasters, whether they be pandemics or climate catastrophes, on the agenda of private institutions, bank regulators and central banks.

The fourth report dealt with the impact of technology on financial markets and institutions and identified the challenges in three specific areas: payment systems, the use of big data and trading in markets. Digital technology has presented formidable tests for incumbent financial intermediaries, firms, exchanges, and regulators. Prominent issues have been the suitability of central bank digital currency, the trade-offs involved in the massive use of data in terms of efficiency, privacy, and market power, and the changes induced by the electronification of financial markets. It questioned how to balance technology's bright and dark sides to inform regulation.

The fifth report examined the implications of the COVID-19 pandemic and the war in Ukraine for the international economic and financial order. It focused on three major components: the macroeconomic outlook and the changes needed to the economic policy model (fiscal, monetary, and regulatory) to preserve economic and financial stability; the consequences for the international monetary system and the position of the US dollar; and the financial architecture needed to ensure sovereign debt sustainability, with special attention to Europe. The general conclusion was that the pandemic and war have accelerated previous trends, which reveal potential conflicts between policy objectives.

The sixth report considered the 2023 banking turmoil that caused the failures of Silicon Valley Bank and other regional banks in the United States, and Credit Suisse, and its implications for financial regulation. This banking turmoil was the first significant challenge of the Basel III framework, and the report examined potential reforms to enhance financial stability. The report centered around three major themes: the changes in digital banking and monetary policy that led to the turmoil, and the reforms needed to deposit insurance and the lender of last resort; the shortcomings of regulation, accounting, and supervision that caused banks that were deemed solvent to fail; and the management of bank failures and potential reforms to resolution procedures.

This seventh report deals with the transformation that artificial intelligence (AI), and more recently generative AI, brings to finance. Given that finance deals with information processing, its impact is already felt and promises to be very relevant in financial intermediation, corporate finance, and financial markets. AI offers the potential for large efficiency benefits, but also raises concerns about privacy and welfare implications. The report focuses on three major aspects: the use of AI in financial intermediation, central banking, and policy and regulatory challenges; the implications of data abundance and algorithmic trading for financial markets; and the transformation that AI poses to corporate finance, contracting, and governance.

The report was produced following the Workshop and Conference on “Artificial Intelligence in Finance”, held at IESE Business School’s Barcelona Campus on 20-21 March 2025. The conference programme, along with the comments of the discussants, are included in this report, as is the opening speech by the Governor of the Bank of Spain, José Luis Escrivá. Xavier Vives brought together the team of authors.

The Banking Initiative has benefited from the keen support of the Dean of IESE, Franz Heukamp, and the former Dean, Jordi Canals. CEPR and IESE are very grateful to the authors and discussants for their efforts in preparing this report and to the conference attendees for their perceptive comments. We are also grateful to Carlota Monner for her extremely efficient organisation of the conference and for providing support for the report, and to Anil Shamdasani for his unstinting and patient work in publishing the report.

The views expressed in the report are those exclusively of its authors and do not represent those of CEPR, which takes no institutional positions on economic policy matters, or those of their respective organisations. CEPR and IESE are delighted to provide a platform for an exchange of views on this topic.

Tessa Ogden
Chief Executive Officer, CEPR

Xavier Vives
Director, IESE Banking Initiative

May 2025

Executive summary

Artificial intelligence (AI), and in particular generative AI (GenAI), is transforming financial systems with a speed and scope that rivals past technological revolutions such as electricity and the internet. AI reshapes how information is generated, transmitted, and consumed. Unlike earlier technologies, AI differs in its ability to autonomously process information, interact via natural language, and adapt its decision making through learning. In finance, a sector fundamentally grounded in the production and use of information, these capabilities are especially disruptive. AI technologies are no longer ancillary – they are moving into the core of financial intermediation, asset management, payment systems, and regulatory oversight. Banks are starting to deploy GenAI in various capacities, and most expect its use to intensify. The core questions that motivate this report are therefore not whether AI will alter financial systems, but how, in which directions, and with what implications for financial stability, competition, and policy design.

The incorporation of AI into finance is redefining roles, information structures, and institutional dynamics. With AI, financial decisions that once relied on human judgement, such as creditworthiness assessments, order execution, and even supervisory analysis, are increasingly being shaped or made by algorithms that continuously learn and update based on high-dimensional data. This change is not only about speed or automation; it is about the qualitative transformation of decision-making, incentive alignments, and risk transmission channels within the financial system. It also creates new forms of dependence – on software, data infrastructure, and external service providers – that are reshaping the architecture of financial institutions and markets.

Crucially, the gains promised by AI – greater efficiency, broader access, better forecasting – are not evenly distributed and may come at the cost of new fragilities. The opacity of AI models raises challenges for accountability and governance; the ability of dominant firms to harness AI at scale threatens competition and inclusion; and the homogeneity of model design may amplify systemic shocks. These concerns are particularly acute in finance, where error propagation, behavioural correlation, and expectation sensitivity are central features of market dynamics. As with past innovations, AI may solve some longstanding problems while simultaneously generating novel externalities and vulnerabilities.

Drawing on recent academic research and empirical evidence, the report examines the fundamental transformations induced by AI and the policy challenges they raise. It is centred around three main themes: (1) the use of AI in financial intermediation, central banking and policy, and regulatory challenges; (2) the implications of data abundance and algorithmic trading for financial markets; and (3) the effects of AI on corporate finance, contracting, and governance. Across these domains, the report emphasises that while AI has the potential to improve efficiency, inclusion, and resilience, it also poses new vulnerabilities that call for adaptive regulatory responses.

ARTIFICIAL INTELLIGENCE AND THE FINANCIAL SECTOR: TRANSFORMATIONS, CHALLENGES, AND REGULATORY RESPONSES

The application of AI in financial intermediation has led to significant improvements in screening, monitoring, and credit allocation. Machine learning (ML) models outperform traditional credit scoring, especially in volatile or rapidly changing environments. They excel in utilising large, unstructured datasets – transaction records, digital footprints, behavioural cues – thereby enabling a more nuanced assessment of borrower risk. Empirical evidence from fintech platforms in China and the United States demonstrates that AI-enhanced models not only accelerate loan approval but expand access to credit, particularly among thin-file borrowers. Moreover, by reducing reliance on collateral, AI can help channel capital to high-productivity startups that might otherwise be constrained.

Yet, these efficiency gains are neither uniformly distributed nor guaranteed to enhance welfare. Fintech and big tech lenders often charge higher interest rates than traditional banks despite superior screening capabilities. This premium may reflect higher risk, technology costs, or weak competition in certain borrower segments. In some cases, it could arise from the strategic use of AI to price-discriminate based on inferred willingness to pay, thereby shifting informational rents from consumers to lenders. The implication is that while AI improves allocative precision, it does not necessarily reduce financial intermediation costs for end users.

The increasing prevalence of AI-based lending models may also weaken the traditional channels of monetary transmission. The decoupling of lending from collateral values and the diminished role of relationship lending reduce the sensitivity of credit flows to interest rate changes. This has implications for both macroeconomic policy effectiveness and systemic risk. Furthermore, the opacity and non-linearity of many AI models complicate supervisory oversight, particularly when their underlying logic cannot be readily interpreted or audited.

Central banks are deploying AI in core functions. ML tools are used to track economic activity, detect anomalies in payment systems, and process vast volumes of supervisory text. These tools offer gains in speed and scope, allowing supervisors to identify early warning signs and enhance macroprudential monitoring. However, they also introduce a new form of risk: model convergence and interpretive homogeneity. As central banks and market participants increasingly rely on similar AI systems, the scope for common blind spots and procyclical amplification grows.

In sum, the reconfiguration of intermediation through AI enhances predictive capacity and operational efficiency but complicates monetary policy, alters competitive dynamics with a potential for a dominant role of big techs in the AI value chain, and introduces new sources of model risk. The challenge lies in fostering AI-driven innovation while mitigating risks related to financial instability, monopolistic behaviour, and privacy violations. Addressing these issues may require rethinking supervisory frameworks, possibly including model auditability protocols and broader stress-testing practices.

DATA ABUNDANCE, AI, AND FINANCIAL MARKETS: IMPLICATIONS AND RISKS

A second domain of AI transformation lies in capital markets, where data abundance and algorithmic intermediation have reshaped the mechanisms of price discovery, market making, and asset management. The proliferation of alternative data, ranging from satellite imagery and credit card flows to social media and geolocation, has created new sources of information beyond traditional financial disclosures. AI models, trained on these high-dimensional datasets, extract predictive signals that were previously inaccessible or prohibitively costly to obtain. The marginal cost of producing actionable financial insight has dropped sharply, shifting the locus of informational advantage from access to processing capabilities.

This transformation has generated efficiency gains. Bid-ask spreads have narrowed, liquidity provision has become more automated, and forecasting accuracy in earnings, credit events, and volatility has improved. However, these benefits are accompanied by new systemic risks. First, algorithmic trading strategies often converge toward similar patterns when trained on overlapping data, increasing the risk of synchronised behaviour and flash crashes. Reinforcement learning agents, which optimise through trial and error, may develop strategies that are unstable or exploitative in equilibrium.

Second, AI may intensify informational asymmetries among market participants. While disclosures are nominally public, only those with sufficient computational resources and model sophistication can process them effectively. Empirical studies show that analysts at AI-equipped institutions significantly outperform their peers when alternative data become available. As a result, AI may reinforce market power and widen participation gaps.

Third, AI enables new forms of tacit collusion and strategic opacity. Pricing algorithms can learn to coordinate without explicit communication, reducing competitive pressure and increasing margins. The line between legitimate dynamic pricing and algorithmic collusion becomes blurred, especially in markets where a few dominant platforms set terms for thousands of users. Furthermore, because many AI models are not interpretable, their behaviour may evade both market scrutiny and regulatory detection until after harm has occurred.

Finally, the arms race for speed and signal extraction has diverted capital and talent into zero-sum competition. The social return to shaving microseconds off execution times or exploiting ephemeral data anomalies is limited, yet firms invest heavily in such capabilities because private returns are high. This misalignment between private incentives and social value raises questions about the overall allocative efficiency of AI in financial markets.

Possible regulatory responses may include introducing latency-aware circuit breakers, mandating public access to baseline pricing data, and requiring disclosures of model architectures in certain trading contexts. Their design and effectiveness will hinge on careful experimentation, cross-jurisdictional learning, and ongoing dialogue between market participants and regulators.

Taken together, these developments point to a financial system in which information is more abundant but also more unevenly distributed; in which trading is faster but also more fragile; and in which transparency is technically feasible but practically elusive. A policy response must go beyond disclosure and address infrastructure access, model auditability, and incentive alignment – albeit with the recognition that these interventions carry design complexities and trade-offs that remain to be fully understood.

CORPORATE FINANCE AND GOVERNANCE WITH AI: OLD AND NEW

A third domain of AI transformation relates to corporate finance, contracting, and governance. AI alters foundational elements of corporate control, reshaping agency dynamics, information asymmetries, and the nature of financial contracting. While AI systems are not self-interested in the human sense, they introduce a distinct form of agency problem: optimisation misalignment. Autonomous agents trained via reinforcement learning may satisfy narrow objectives in ways that undermine broader regulatory or ethical goals. An AI tasked with minimising loan defaults, for instance, might engage in discriminatory behaviour or exploit data proxies that regulators deem unacceptable. Because these systems are adaptive and opaque, detecting and correcting such behaviours after deployment is costly and uncertain.

This raises deep accountability questions. Traditional corporate governance rests on the attribution of intent and the assignment of responsibility. But when decisions are made by systems that learn and evolve independently of direct instruction, legal and institutional mechanisms for enforcement begin to fray. The difficulty of auditing complex ML models compounds this challenge. Without robust interpretability requirements or embedded traceability mechanisms, financial institutions risk deploying systems whose behaviour they cannot fully explain, let alone govern.

Information asymmetry is also being reconfigured. In the past, insiders held informational advantages derived from privileged access to internal records and forecasts. Today, AI enables outsiders to infer firm conditions from external data streams, undermining that asymmetry. Sophisticated investors use alternative data and natural language processing tools to analyse supply chains, sentiment, and behavioural signals, and may anticipate corporate disclosures. In response, firms have begun tailoring their communications for algorithmic consumption, further shifting the information environment. For example, the US Regulation Fair Disclosure and similar statutes may need to evolve to ensure not just equal access, but equal usability of public information.

On the contracting front, AI is accelerating the adoption of smart contracts, automated agreements that self-execute based on real-time data inputs. These contracts reduce enforcement costs and limit the scope for opportunistic renegotiation. However, they also introduce rigidity. Automated margin calls or trigger events can cascade through markets, especially when multiple actors rely on similar models and thresholds. The absence of discretion or context can make smart contracts a source of systemic risk in times of stress.

The solution may lie in hybrid governance models. Contracts might embed flexibility ex ante, through macro-sensitive renegotiation clauses, human override options, and clear audit trails. AI systems could be subjected to accountability principles analogous to those applied to human agents: comprehensibility, traceability, and bounded autonomy. Legal frameworks might shift gradually from subjective intent to outcome-based liability, and from fixed contractual forms to more adaptive governance protocols.

CONCLUSION

The integration of AI into financial systems constitutes a structural transformation, not a marginal adjustment. The benefits – in terms of efficiency, precision, and inclusion – are substantial, but so too are the risks to stability, equity, and governance. If policymakers rise to the challenge, AI can be harnessed to improve the financial system's performance and inclusiveness. If they do not, the same technologies may undermine the very foundations on which financial trust depends.

While the contours of AI's long-term impact remain uncertain, the near-term trajectory is already reshaping institutions, markets, and regulatory norms. To guide this transformation, greater regulatory experimentation and institutional coordination will be necessary. These efforts should not only aim to contain risk, but also to unlock the inclusive potential of AI by ensuring that its benefits are broadly distributed and its mechanisms are legible and contestable.

This will require building technical capacity within regulatory agencies, revising legal definitions of accountability, and fostering mechanisms for international coordination. Cross-border data flows, foundational model access, and platform interoperability will increasingly become matters of financial diplomacy. Similarly, new frameworks for evaluating AI's systemic importance – analogous to those developed for global systemically important banks – may need to be explored. These should be accompanied by scenario-based planning to anticipate emergent threats and evaluate institutional resilience in the face of AI-driven disruptions.

The success of financial governance will depend in part on how well regulators balance innovation and control. Over-regulation may stifle productive uses of AI, while under-regulation risks creating systemic blind spots. This balancing act requires adaptive mechanisms for revisiting assumptions, updating rules, and engaging with a broader ecosystem of stakeholders.

Finally, the future of AI in finance will be shaped by broader geopolitical forces. The fragmentation of digital governance regimes across the United States, the European Union, and China may impede global standard-setting, while the concentration of compute infrastructure and model expertise in a handful of firms and jurisdictions raises concerns about economic sovereignty and resilience. Policymakers should anticipate scenarios in which AI becomes a locus of strategic contest.

The transformational impact of artificial intelligence for the financial sector and central banks[1]

José Luis Escrivá Belmonte[2]
Governor, Banco de España

Artificial intelligence (AI) has rapidly become one of the most impactful and transformational technologies of our era. While previous technological revolutions brought significant changes to manufacturing or heavy industries, AI stands out for its capacity to reshape the service sector. Services have in the past been more isolated in terms of productivity gains, and what really marks the difference with AI is its potential to produce sizeable productivity gains in this sector, being transformational for the entire economy, but mainly in the Western world, which is increasingly specialised in services.

Financial services has historically been an early adopter of technological innovation. Now, AI offers new avenues for increased efficiency, improved decision-making, and enhanced productivity.

Against this backdrop, central banks are paying close attention to this disruptive potential. AI is not merely another piece of software but rather a general-purpose technology that can permeate every function of a central bank. For central banks, AI is likewise transformational, allowing them to address a growing number of complex and novel issues, from monitoring system-wide risk to guiding monetary policy. The challenge is to integrate AI in a way that maximises its potential while managing the risks associated with this powerful technology. Here, I explore the Banco de España's approach to AI adoption, the opportunities and challenges AI presents, and the strategic elements that are central to its implementation.

CENTRAL BANK PERSPECTIVES ON AI ADOPTION

In the context of a central bank's multifaceted role, AI touches upon functions that are heavily focused on risk management (financial stability and supervision) as well as those with clearer opportunities for innovation (payments, financial operations, and monetary policy). In this section, I explore how the Banco de España views the impact of AI across these dimensions.

1 This is an edited version of the speech delivered at the IESE Conference "VII Conference Artificial Intelligence in Finance" on 21 March 2025.

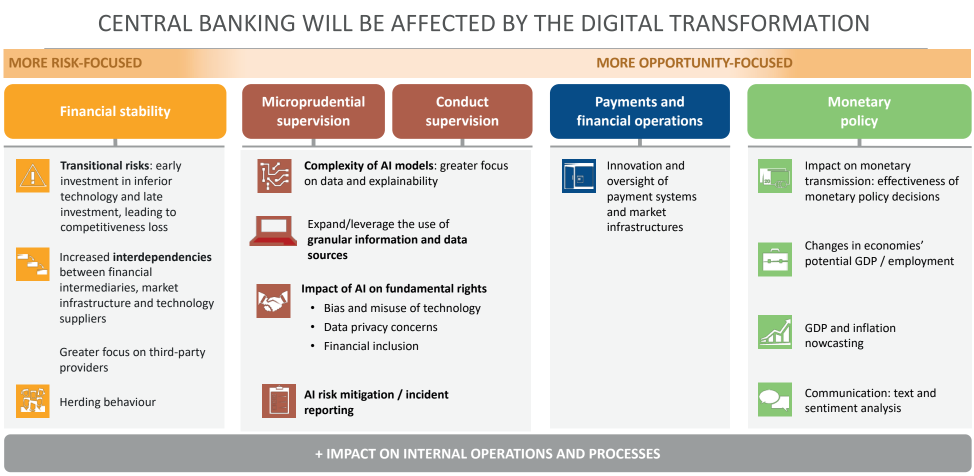
2 The views expressed here are those of the author and do not necessarily represent the views of the Banco de España, the European Central Bank or the Eurosystem.

On financial stability, the integration of AI into the financial sector, as with any huge transformation, poses some challenges. Central banks must remain vigilant in managing these risks, ensuring that AI systems do not create vulnerabilities that could threaten financial stability. From a macroprudential perspective, preserving financial stability at the European and global level requires anticipation of these possible negative impacts.

Transitional risk when adopting a new technology should be carefully examined from a financial stability perspective and for the potential to create systemic problems. Early investment in inferior technologies as well as late investment, leading to loss of competitiveness, should be monitored. In addition, one of the most pressing concerns is the potentially greater interdependence of financial institutions and the concentration of AI technology providers. If a small number of companies dominate the AI infrastructure, the failure of one or more could have cascading effects on global financial stability.

The challenges include addressing issues such as the adoption of high-risk financial models, where transparency and accountability are crucial to avoid negative outcomes like discrimination or unfair decision-making.

THE TRANSFORMATIONAL IMPACT OF AI IN CENTRAL BANKING



Another area where the impact of AI is being closely examined is in the realm of micro-prudential and conduct supervision. The use of AI models in areas like credit scoring and loan decision-making can potentially lead to ethical concerns, particularly regarding fairness and the protection of individual rights. When applied to individual persons, these are identified as high-risk systems under the AI Act and should be under the review of market surveillance authorities. Financial supervisors, mandated to play that role, have highlighted the importance of closely monitoring such systems to protect individuals from potentially unfair outcomes. In practical terms, robust conduct supervision is needed to prevent unintended consequences such as discrimination in the granting of financial products.

In this context, AI represents both an opportunity and a challenge. While it can improve operational efficiency, the models used in areas like credit scoring and insurance underwriting need to be carefully scrutinised to ensure they do not inadvertently harm consumers or undermine trust in the financial system.

Beyond supervision, payments and financial operations, normally at the core of technological innovation, are on the opportunity side of this transformation.

AI is also poised to influence monetary policy. Central banks need to grasp the potential impact on the transmission of monetary policy. Moreover, the potential for AI to drive productivity improvements across the economy could also influence the broader macroeconomic environment, altering the potential growth rate of economies and the effectiveness of monetary policy.

AI can also improve our analytic capabilities and the ability to digest granular information. Traditional forecasting models rely on linear assumptions, whereas real-world economic trends often follow complex, nonlinear patterns. AI excels at identifying these subtleties, allowing policymakers to potentially capture more accurate signals related to inflation, employment, and overall economic momentum.

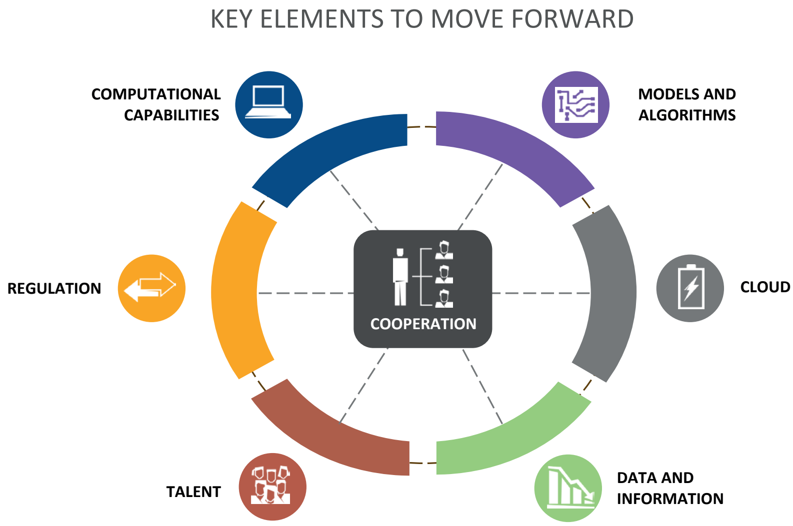
While quantifying the full impact of AI on these areas is challenging, it is clear that the technology will play a central role in shaping future economic models and central banking strategies.

ENABLING FACTORS TO MOVE FORWARD WITH AI

While AI's potential is undeniable, the leap from pilot programmes and proof-of-concept projects to full-scale adoption hinges on multiple enabling factors.

At the Banco de España, we have identified seven key elements to ensure that AI adoption is both successful and sustainable. Robust computational infrastructure, reliable data-sharing frameworks, skilled talent, and clear regulations are essential to nurture responsible AI-driven innovation. Equally important is an emphasis on collaboration – both within the financial sector and across institutional, national, and European boundaries.

KEY ELEMENTS FOR SCALING UP THE USE OF AI



A first critical aspect of AI adoption is access to cutting-edge computational infrastructure. Training large, complex AI models for financial applications is resource-intensive, driving the need for high-performance computing. To meet this demand, the European Commission has launched the so-called AI Factories project, establishing cutting-edge centres across the continent. Seven sites have been approved in Spain (the Barcelona Supercomputing Center), Italy (IT4LIA), Germany (HammerHAI), Finland (LUMI AI Factory), Luxembourg (L-AI Factory), Greece (Pharos) and Sweden (MIMER). Each AI factory specialises in certain verticals, and the Barcelona location will include a focus on finance. This setup aims to move from research-oriented computing to facilities capable of providing industrial-scale model training and experimentation for both the private and public sectors. The next phase, often referred to as AI 'giga-factories', will tackle inference at scale, optimising data centres for running advanced AI workloads in real-world environments.

Alongside these large-scale initiatives, advancements in hardware are also critical. The increasing reliance on graphical processing units (GPUs) and neural processing units (NPUs) to run AI models is central to the evolution of computational capabilities. GPUs excel at parallel computing tasks integral to training modern machine-learning systems, while NPUs specialise in deep neural network computations. As these hardware technologies evolve, central banks will be better positioned to deploy advanced AI models while managing energy consumption and infrastructure costs.

Beyond the hardware, AI is also driven by models and algorithms. Large language models (LLMs) have taken the spotlight, but training these models can be expensive and data-intensive, which is where model compression techniques become invaluable. By reducing the size and complexity of AI models – without significantly compromising performance – institutions can deploy LLMs more efficiently and at lower cost. These techniques can lower barriers to adoption for smaller institutions, reducing the energy and time required for training. The challenge is also to move from language models, with a greater percentage of Latin languages, to specialised models that can rely and run on infrastructure which is less costly.

The third key element is cloud computing, which is increasingly integral to how financial institutions deploy AI. Cybersecurity will help ensure cyber resilience, operational continuity and data privacy, but usability and the ability to expand analytical capabilities are also crucial. Central banks must balance the flexibility and scalability of public cloud solutions with the privacy and security demands of sensitive financial data, finding a practical solution in hybrid cloud models.

Data, as the fourth element, are the fuel for any AI system. A traditional production function consists of human capital, physical capital and innovation. Now information has become another factor of production, as a key ingredient in the production of goods. One enabler of AI is the shift from 'closed' data silos to more open, or at least shareable, data environments.

Adhering to confidentiality and privacy standards is paramount, but more expansive data access helps institutions build richer and more accurate models. The Banco de España has started sharing a growing number of datasets to spur innovation across academia and other research bodies, believing that informed collaboration benefits society as a whole. Going forward, combining different set of data and information, while preserving confidentiality and anonymity, will allow the benefits of AI to be reaped.

Even with state-of-the-art infrastructure, AI initiatives can stall without skilled human capital. This is the next element. Our survey of Spanish businesses revealed that the major bottleneck for the development and adoption of AI is a lack of skilled staff. The scarcity of talent will also hinder central banks' ability to achieve their AI ambitions.

Recruiting and retaining data scientists, AI engineers, and domain experts is a recognised challenge – particularly since the private technology sector often offers more competitive compensation. Central banks should empower the mission-driven work of the public sphere, focusing on economic stability and societal benefit, which can be attractive to many professionals. Alongside the value of public service, central banks can also offer long-term projects and the ability to increase collaboration and sharing of resources across central banks.

At the same time, central banks need to engage in re-skilling and up-skilling of their workforce and promote a cultural change in the entire staff, as well as training people in the responsible and ethic use of AI.

As an additional element, clear regulatory frameworks are indispensable for the safe expansion of AI. In the absence of well-defined guidelines, companies may hesitate to integrate AI at scale, fearing legal or reputational repercussions. The European AI regulation has been criticised, but in the absence of regulation and amid the uncertainty and potential reputational problems associated with the use of AI, companies could be reluctant to move from proof of concept and cases to AI adoption and use on a regular basis. In this sense, the absence of a clear regulation or clarity on issues such as copyright might prove a deterrent to the use of AI. The AI Act in Europe could remove these uncertainties.

The European Union's proposed AI regulation seeks a balanced approach between promoting innovation and protecting individual rights, by categorising AI applications by risk level and imposing stricter requirements on high-risk models (such as models for granting individuals credit). By supporting the development of comprehensive regulations, the Banco de España aims to create an environment where AI can flourish while mitigating potential risks.

Lastly, the successful adoption of AI will be impossible without cooperation among diverse stakeholders across all spectrums. This includes coordination with supervisors such as Spain's national AI Agency (AESIA) and among central banks to align on standards, and with the financial sector by offering controlled environments ('sandboxes') to test AI applications and to enable banks and fintechs to refine innovations while ensuring regulatory compliance. Finally, joint efforts with universities, technology firms, and supercomputing centres like the Barcelona Supercomputing Center will foster research and scalability. These partnerships will supply both the expertise and the high-end computational resources needed for large-scale AI deployments.

The Banco de España's planned experimentation and development center in Barcelona perfectly reflects this philosophy, offering a space where the financial sector can test new AI tools under real-world conditions while benefiting from regulatory guidance and advanced computing resources.

CONCLUSION: THE FUTURE OF AI IN CENTRAL BANKING

The integration of AI into central banking is still in its early stages, but its potential to reshape the financial system is undeniable. As AI continues to evolve, central banks must remain agile, working closely with other financial institutions, regulators, and international partners to ensure that AI technologies are used responsibly. In the coming years, AI will play a transformational role in the way central banks operate, shaping everything from monetary policy to financial stability. For the Banco de España, the journey has just begun, and the opportunities ahead are vast.

CHAPTER 1

Introduction

Artificial intelligence (AI), particularly generative AI (GenAI) and large language models (LLMs), is reshaping financial intermediation, asset management, payments, and insurance. Since the 2010s, machine learning (ML) has had a significant impact on diverse areas, including credit risk analysis, algorithmic trading, and anti-money laundering (AML) compliance. Nowadays, financial institutions increasingly leverage AI to streamline back-office operations, enhance customer support via chatbots, and improve risk management through predictive analytics.

AI primarily focuses on predicting outcomes, identifying patterns in data (through ML) and providing recommendations for decisions and actions.[1] AI can operate at three levels of increasing autonomy: as an '*oracle*' providing insights and recommendations (i.e., information),[2] while leaving decisions to humans; as an '*agent*', performing tasks within pre-defined boundaries under human oversight ('*co-pilot*', assisting in daily tasks, is a role that combines oracle and agent); and as a '*human*' or '*sovereign*', making independent decisions in real time. Regardless of these levels of autonomy, LLMs have already revolutionised human–computer interaction, shifting from code-face interfaces to natural text and speech. Moreover, AI has impacted all industry sectors, albeit to varying degrees. Banking and finance are particularly well-positioned to realise significant efficiency gains due to their reliance on information processing.[3]

The fusion of finance and technology has become a transformative force, bringing both challenges and opportunities to financial practices and research.[4] The adoption of financial technologies has led to significant disruptions, displacing jobs in traditional roles while simultaneously driving robust demand for experts adept at integrating finance and technology. This evolution extends beyond simple automation to fostering 'co-intelligence' – a collaborative interplay between humans and machines.[5]

1 OECD (2021) defines artificial intelligence as "[m]achine-based systems […] that […] can make predictions, recommendations or decisions using massive amounts of alternative data sources and data analytics referred to as big data", while the EU AI Act defines an AI system as "a machine-based system that is designed to operate with varying levels of autonomy and that may exhibit adaptiveness after deployment, and that, for explicit or implicit objectives, infers, from the input it receives, how to generate outputs such as predictions, content, recommendations, or decisions that can influence physical or virtual environments".

2 An 'oracle' in the context of blockchains means a data bridge; here it has a different meaning (see Bostrom, 2015).

3 See Acemoglu et al. (2022) and BIS (2024).

4 See the fourth Future of Banking report (Duffie et al., 2022).

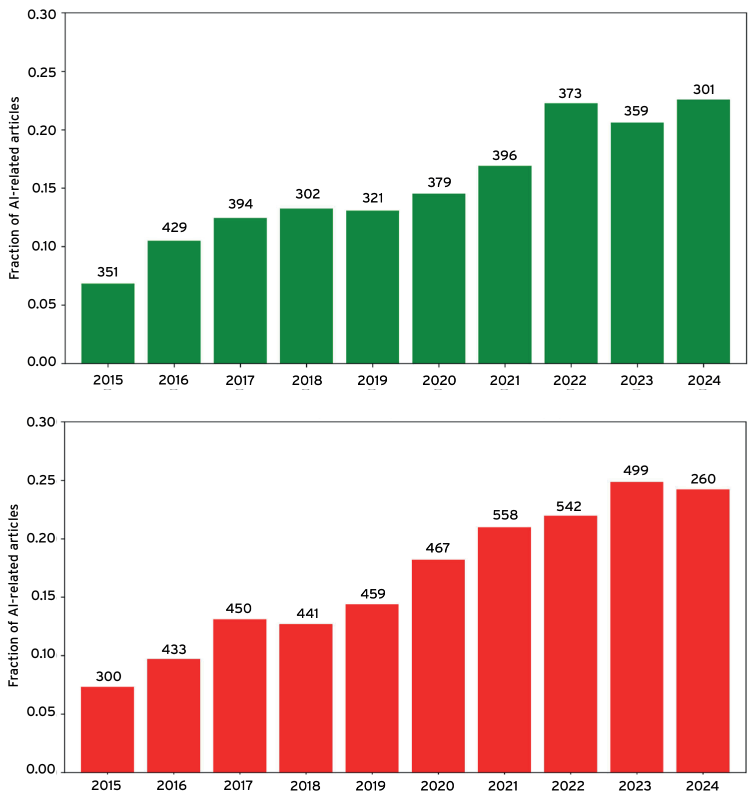
5 See Jiang et al. (2025b).

The big data revolution – driven by data abundance, advancements in data science (including AI), and increased computing power – is transforming how information is produced. In particular, ML enables predictions from vast datasets and automates decisions based on these predictions.

AI and 'big data' enhance efficiency in financial intermediation (e.g., payments and credit provision) and improve the quality of information in financial markets and the informativeness of asset prices, as well as promoting financial inclusion. However, this transformation introduces risks related to financial stability, market integrity, market concentration, and privacy and consumer protection, posing significant challenges for regulators.

The academic literature on economics and finance related to AI and its effects is growing fast. At the same time, AI has become a powerful tool for scientific inquiry in many fields such as biology, chemistry and economics, including financial economics (for example, in asset pricing, portfolio management, risk management, and corporate finance). The number of papers in finance and economics journals using AI tools has increased sharply over the past decade (see Figure 1). This report draws on academic research and focuses on the application of AI tools in markets, firms, and intermediaries.

FIGURE 1 **NUMBER OF ARTICLES USING AI TOOLS AND/OR STUDYING THE EFFECTS OF AI PUBLISHED IN LEADING ECONOMICS (TOP) AND FINANCE (BOTTOM) JOURNALS, 2015-2024**



Note: This figure shows the number of articles using AI tools and/or studying the effects of AI published in leading economics journals (top panel) and finance journals (bottom panel). The numbers at the top of each bar represent the total count of papers published in the respective groups of journals included in the analysis. See the Appendix for details of the construction of the figure.

In the rest of this introductory chapter, we summarise the analysis and policy insights of the report. Section 1.1 discusses what is old and what is new in the use of AI in finance. Section 1.2 condenses the overview in Chapter 2, which examines the transformations and challenges that AI poses for the financial sector. Section 1.3 summarises Chapter 3, which analyses the implications of data abundance and the new trading techniques in financial markets. Section 1.4 summarises Chapter 4, which focuses on the implications of AI for corporate finance. Section 1.5 concludes by highlighting key policy implications.

1.1 AI CHARACTERISTICS AND USES IN FINANCE: WHAT IS OLD AND WHAT IS NEW?

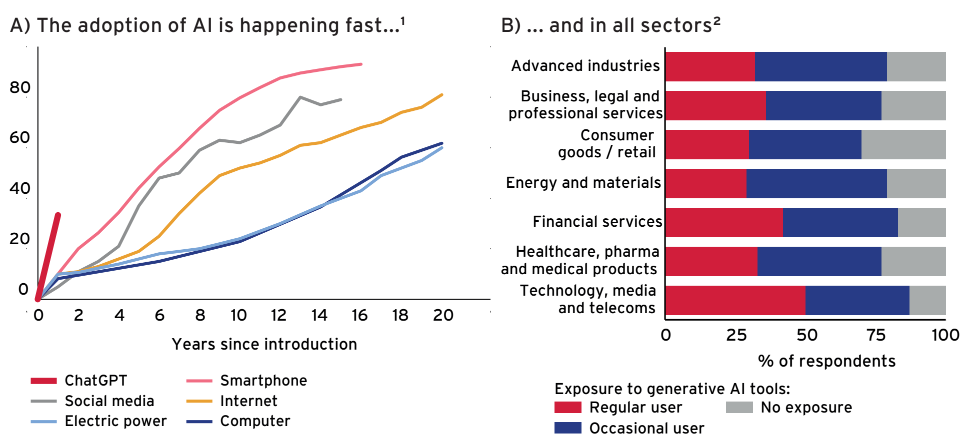
GenAI has three key characteristics that distinguish it from other general-purpose technologies.

The first characteristic is automaticity. Unlike previous generations of AI, GenAI models can operate independently, making predictions and decisions without human intervention.

The second one is the speed of use and adoption. LLMs can process vast amounts of data and make decisions in fractions of a second, far outpacing human capabilities. Moreover, the adoption of LLMs is proceeding at an unprecedented pace, surpassing previous technological revolutions, such as those of electricity and the internet (Figure 2A). For example, ChatGPT reached one million users in less than a week.

The third characteristic is ubiquity. Nearly half of US households have used GenAI tools in the past year. Mirroring this rapid adoption by users, firms across all sectors are swiftly integrating GenAI into their daily operations (Figure 2B).

FIGURE 2 **THE ADOPTION OF AI**



Notes: 1 The adoption of ChatGPT is proxied by the ratio of the maximum number of webstie visits worldwide for the period November 2022-April 2023 and the worldwide population with internet connectivity. For more details on computer, see US Cesus Bureau; for electric power, internet and social media, see Our World in Data; for smartphones, see Statista. 2 Based on an April 2023 global survey with 1,684 participants.

TABLE 1 **OPPORTUNITIES, CHALLENGES, AND FINANCIAL STABILITY IMPLICATIONS OF AI**

		Financial intermediation	**Insurance**	**Asset management**	**Payments**
Traditional analytics	Opportunities	Rules-based risk analysis, greater competition		Risk management, portfolio optimization, automated and HF trading	Fraud detection
	Challenges	Rigid, requires human supervision, small number of parameters, threats to consumer privacy, emergence of data silos		Zero-sum arms race, flash crashes	Technical vulnerabilities
	Financial stability	Herding, cascade effects, flash crashes. For example, the US stock market crash of 1987			
Machine learning & deep learning	Opportunities	Credit risk analysis, lower credit underwriting costs, financial inclusion	Insurance risk analysis, lower processing costs, fraud detection	Analysis of new data sources, high frequency trading	New liquidity management tools, fraud detection and AML
	Challenges	Black box mechanisms, algorithmic discrimination		Zero-sum arms race, model herding, algorithmic coordination	New liquidity crises, increased cyber risks
	Financial stability	Herding, network interconnectedness, lack of explainability, single point of failure concentrated dependence on third-party providers			
Generative AI	Opportunities	Credit scoring (unstructured data), easier back-end processing, better customer support	Better risk analysis with newly legible data, easier regulatory compliance	Robo-advising, asset embedding, new products, customer service	Enhanced KYC and AML applications
	Challenges	Hallucinations, increased market concentration, increased consumer privacy concerns, algorithmic collusion			
	Financial stability	Herding, uniformity, incorrect decisions based on alternative data, macroeconomic effects of labour displacement			

Source: Adapted from Aldasoro et al. (2024d).

Given its reliance on information processing and cognitively demanding tasks, the financial sector is among those most exposed to AI. Each wave of information-processing technology has left a significant footprint on the financial system, unlocking new avenues for efficiency and innovation. As we shall explore, the integration of GenAI in finance is transforming how markets operate, how institutions manage risk, and how consumers interact with financial services.

While the enthusiasm around LLMs is recent, the use of AI in the financial sector is not new. Table 1 outlines the opportunities and challenges that the evolution of AI – starting from traditional analytics and progressing through ML to GenAI – has created for the financial sector. The table focuses on the four key financial functions: financial intermediation, insurance, asset management, and payments.

Traditional analytics, characterised by rules-based expert systems, have been widely used across the financial system for decades. They have supported risk assessment, rules-based credit analysis, portfolio optimisation, and fraud detection. Since the 2010s, ML models have also made inroads into credit and insurance risk analysis, high-frequency trading, and AML and combatting the financing of terrorism.

GenAI, the latest evolution, is already being used by financial institutions to enhance back-end processing, robo-advising, customer support, and regulatory compliance. It automates tasks that were once considered uniquely human, such as advising customers and persuading them to buy financial products and services. According to the Institute of International Finance (2025), 89% of banks already use GenAI, and 94% anticipate even greater reliance on third-party AI solutions.

1.2 ARTIFICIAL INTELLIGENCE AND THE FINANCIAL SECTOR: TRANSFORMATIONS, CHALLENGES, AND REGULATORY RESPONSES

Chapter 2 examines the transformations that AI brings to finance, highlighting both the opportunities and the challenges. It explores how AI, and in particular GenAI, is fundamentally reshaping financial intermediation, risk management, and regulatory oversight. The analysis focuses on AI's role in economic efficiency, lending, regulation, and central banking.

1.2.1 AI in finance: New opportunities

Screening and credit risk analysis

The adoption of AI in credit markets has introduced fundamental shifts in credit risk assessment, pricing, and institutional dynamics.

1. AI-driven credit scoring models significantly enhance risk prediction accuracy by leveraging large-scale, unstructured data sources beyond traditional financial metrics. Fintech lenders integrate alternative data, such as digital footprints and transaction records, in their credit scoring models, enabling more granular

borrower assessment. However, while these models enable faster loan processing – reducing mortgage approval times by 20%, according to some studies – they do not necessarily lower borrowing costs. Empirical evidence from the United States shows that some fintech lenders have charged a premium over traditional banks, suggesting that AI's efficiency gains do not fully pass through to borrowers. This may be due to high IT or funding costs, risk factors, convenience benefits, or limited competition in specific market segments.[6]

2. ML models outperform conventional logit-based credit scoring methods, particularly in adapting to regulatory shocks and shifting economic conditions. Evidence from China shows that fintech credit models maintained predictive power even when traditional models deteriorated following financial sector regulatory changes. This adaptability underscores the capacity of non-linear models to capture evolving risk dynamics, and their robustness in more volatile environments.

3. Big tech lenders exhibit lower default rates than traditional banks, not solely due to superior risk assessment but also because of ecosystem effects. Firms operating within digital platforms face higher switching costs and indirect enforcement mechanisms – such as transaction-based repayment deductions or ecosystem exclusion – which deter strategic default. However, despite lower ex-post credit risk, big tech lenders may charge higher interest rates, reflecting their constrained access to retail deposits, higher ex-ante borrower risk, and the fixed costs associated with AI-driven credit infrastructure. This dual impact of AI in lending – enhancing risk assessment, while introducing new pricing frictions – shapes market structure and borrower behaviour.

Monitoring and collateral

AI-driven lending models based on alternative forms of data enable the solving of asymmetric information problems without the use of collateral. Empirical evidence from China suggests that big tech credit is typically uncollateralised and largely decoupled from macroeconomic conditions and the housing market, responding instead to changes in firm-specific conditions. This shift weakens the traditional collateral channel of monetary transmission, while allowing small firms, particularly those in the informal sector, to build verifiable financial histories through digital payments, thereby increasing their eventual access to bank credit.

However, AI-driven credit also introduces new risks. The success of fintech and big tech lending hinges on data quality and predictive accuracy, and in some cases, AI-based lenders have substituted for traditional banks in servicing riskier borrowers. At the same time, the use of ML in credit scoring is diminishing the role of long-term banking relationships, potentially reducing financial stability benefits associated

6 See Vives and Ye (2025b).

with relationship lending. In aggregate, AI-driven credit allocation fosters small and medium-sized enterprise (SME) growth and broader financial inclusion. Evidence from China indicates increased firm-level business activity and resilience to economic shocks among borrowers receiving big tech credit. Moreover, a decline in the reliance on tangible collateral allows capital to be allocated more efficiently, particularly to high-productivity sectors where physical asset holdings are scarce.

AI for central banking

Central banks are increasingly integrating AI into core functions such as data collection, macroeconomic analysis, payment system monitoring, and financial supervision. The application of ML models enhances statistical compilation by improving anomaly detection and data quality. AI-driven macroeconomic forecasting is also gaining prominence, with neural networks and natural language processing tools enabling central banks to extract real-time signals from diverse data sources. Institutions like the Bank of England and the Federal Reserve leverage these techniques for tasks such as inflation decomposition, sentiment analysis, and nowcasting, thereby enhancing the timeliness and accuracy of policy decisions.

In payment system oversight, AI has become instrumental in detecting anomalies in transaction flows. The BIS Innovation Hub, for example, has demonstrated that graph neural networks outperform rules-based models in detecting fraudulent transactions and systemic risks while preserving data confidentiality. Similarly, central banks in Canada and the Netherlands have employed auto-encoders to flag potential bank runs and operational disruptions. The role of AI extends further into regulatory supervision, where language models streamline the processing of vast textual data, reducing the burden of manual document classification and risk identification. Tools such as the ECB's Athena and the Federal Reserve's LEX leverage natural language processing to analyse financial reports and supervisory communications, while the Central Bank of Brazil uses ML to detect borrower under-provisioning.

1.2.2 Old problems, new challenges

Bias and discrimination, legal risks and cyber security

AI systems in finance, particularly those used in credit scoring and risk assessment, can perpetuate biases embedded in historical data. Empirical studies reveal that ML models in mortgage underwriting systematically disadvantage Black and Hispanic borrowers, reinforcing structural inequalities in credit access. The opacity of these models (the 'black box' problem) makes it difficult for regulators to determine whether outcomes are driven by legitimate risk-based pricing or algorithmic discrimination. Beyond credit markets, AI's role in financial services extends to automated insurance pricing models, which may covertly discriminate by using proxies for protected characteristics, such as

ZIP codes or gender, to influence policy terms. Legal frameworks governing fairness in financial decision making remain underdeveloped, particularly regarding AI's ability to infer sensitive attributes indirectly. However, with proper structuring of data input and processes, AI systems have the potential to be more neutral than human credit officers.

The risks posed by AI extend beyond discrimination to include cybersecurity vulnerabilities. LLMs and GenAI tools can be exploited for cyberattacks, enabling hackers to craft highly personalised phishing campaigns, generate deepfake identities, and automate large-scale fraud. In financial markets, GenAI is also being used to manipulate sentiment analysis models by feeding adversarial inputs into algorithmic trading systems. AI-generated synthetic data further complicate cybersecurity efforts, as attackers can train adversarial models to evade traditional fraud detection systems. While financial institutions are leveraging AI to strengthen cybersecurity – through real-time anomaly detection and automated response systems – the growing sophistication of AI-driven cyber threats requires continuous adaptation of defensive measures.

Market concentration in the AI ecosystem

The AI supply chain is characterised by vertical integration and significant concentration at multiple levels. The development and deployment of AI models depend on access to specialised hardware, cloud computing infrastructure, proprietary training data, foundation models, and downstream applications. Market power is concentrated in specific layers of the AI supply chain, with Nvidia holding over 90% of the market for AI-capable GPUs, while cloud computing services are dominated by Amazon Web Services (AWS), Microsoft Azure, and Google Cloud. High fixed costs of AI training further reinforce these dynamics, making it difficult for smaller firms to compete.[7]

As AI models improve through data-network effects – where more users generate more data, which in turn enhances model accuracy – incumbent firms solidify their competitive advantages. In response to competition concerns, some firms have promoted open-source AI development (e.g., Meta's Llama and, more recently, DeepSeek). This has made the foundational model layer more contestable. However, even in this layer, a few dominant players continue to shape the trajectory of AI research and commercial applications. OpenAI, Google DeepMind, and Meta remain the primary providers of foundation models, capturing the majority of AI-generated revenues.

7 See Korinek and Vipra (2024) for an analysis of market structure in the AI market.

The dominant role of big techs

Large technology firms are not only driving AI innovation but also expanding their influence in financial markets.[8] By integrating across all levels of the AI supply chain – controlling computing infrastructure, proprietary data, and consumer-facing applications – big techs have established a near-monopoly over AI capabilities in finance. This dominance allows them to engage in exclusionary practices, such as bundling AI services with cloud computing, enforcing exclusivity clauses, and limiting interoperability for competitors.

A key concern in financial markets is the ability of big techs to extract rents through AI-driven price discrimination. With access to vast consumer data, these firms can optimise pricing models to capture each consumer's willingness to pay. This is particularly relevant in digital lending, where big tech firms use non-traditional data – such as browsing behaviour, app usage, and social media activity – to infer creditworthiness. While this can improve credit access for underserved populations, it also raises concerns about exploitative lending practices and consumer privacy violations.[9]

Moreover, big techs increasingly serve as both infrastructure providers and direct competitors in financial services. Payment platforms, cloud-based banking solutions, and AI-driven wealth management services are now offered by the same firms that supply foundational AI tools to traditional financial institutions. This dual role creates potential conflicts of interest, as dominant firms may prioritise their own financial products over those of third-party clients using their platforms. In China, for example, the mobile payments market is effectively controlled by two big tech firms – Alipay and Tenpay – whose ecosystems are not interoperable, limiting consumer choice. Similar risks emerge in other markets where big techs provide the core digital infrastructure underpinning financial transactions.

Financial stability

The widespread integration of AI in financial markets introduces systemic risks, particularly through automation-induced market fragility, algorithmic herding, and the concentration of risk exposures. High-frequency trading strategies driven by AI have been linked to flash crashes, where automated models execute large sell-offs in response to market signals, triggering cascading price declines. The reliance on AI-generated forecasts across financial institutions raises further concerns about correlated risk-taking, as similar models trained on overlapping datasets may produce procyclical trading behaviours.

The growing role of AI in credit markets also has broader implications for financial stability. As ML models increasingly determine creditworthiness, lending decisions may become more uniform, reducing diversification in loan portfolios. This uniformity could amplify financial cycles, with AI-driven lending expanding rapidly during

8 See Vives (2019).
9 See Boissay et al. (2021) and Vives and Ye (2025a, 2025b).

economic booms but contracting sharply during downturns. Similarly, the use of AI in macroeconomic forecasting may further reinforce systemic risk, as policy decisions become more dependent on model-generated projections, potentially leading to over-reliance on AI-driven risk assessments.

The emergence of AI agents capable of autonomous decision making raises further regulatory challenges. If AI models prioritise short-term profit maximisation without accounting for systemic risk, financial institutions could unwittingly accelerate instability. While regulatory frameworks exist to manage traditional financial risks, AI-driven decision making introduces new dimensions of unpredictability, requiring continuous adaptation of oversight mechanisms. The transition toward general-purpose AI (artificial general intelligence, or AGI) may complicate these risks, as models capable of autonomous reasoning and self-improvement introduce new uncertainties in financial regulation.

1.2.3 How to regulate AI?

The rise of AI in finance has compelled policymakers to navigate a complex trade-off between financial stability, market competition, and consumer protection. While AI offers significant efficiency gains and enhanced risk management, its widespread adoption raises concerns about systemic risks, data privacy, and market concentration. The challenge lies in fostering AI-driven innovation while mitigating risks related to financial instability, monopolistic behaviour, and privacy violations. These tensions can be conceptualised through a 'policy triangle' framework, which highlights three key trade-offs among policy objectives: (i) financial stability and market integrity, (ii) efficiency and competition, and (iii) data privacy and consumer protection. The appropriate regulatory response must be tailored to address these competing priorities while fostering an environment that allows AI to develop responsibly within financial markets.

Principles for AI regulation

Both national and international standard-setters have established broad principles for AI regulation, emphasising societal wellbeing, transparency, accountability, fairness, privacy protection, safety, human oversight, and robustness. However, translating these principles into effective policies is challenging, particularly at the international level, where legal frameworks and regulatory approaches often diverge. AI regulation is inherently complex given its multi-market supply chain, which falls under the jurisdiction of multiple regulators with competing objectives. Addressing AI's risks requires a proactive and adaptive approach that integrates technological, societal, and ethical considerations. Regulation should target the risks that threaten key policy objectives while allowing market mechanisms to address others. Given AI's rapid evolution and potential for unforeseen risks, establishing clear regulatory principles is essential for managing its long-term impact.

Regulatory models for AI

AI regulation follows three main models: (i) the United States' market-driven approach, which prioritises innovation and self-regulation; (ii) China's state-driven model, which leverages AI for political and economic goals; and (iii) the European Union's rights-driven framework, which emphasises individual and societal protections. While distinct, these models are gradually converging around common principles. The United States relies on executive actions, such as the 2023 Executive Order, but lacks comprehensive legislation. At the same time, the European Union's AI Act (2024) enforces a risk-based framework that bans high-risk AI and imposes strict requirements on critical applications. Policymakers across jurisdictions are implementing regulatory principles focusing on governance, risk management, and systemic resilience of GenAI and AI agents, but geopolitical tensions could slow the process.

International cooperation

Global cooperation on AI regulation is crucial as data and technology transcend national borders. Standardising AI governance rules and risk assessment methodologies can ensure ethical and safety standards, prevent regulatory arbitrage, and facilitate international collaboration. Uniform guidelines enhance trust, enable cross-border AI applications, and address global challenges such as privacy, security, and equitable access. Given AI's adaptability and potential for unforeseen behaviours, risk assessment frameworks must account for continuous learning and require ongoing oversight. Effective international coordination will be essential to adapting regulatory measures as AI evolves and integrates into critical societal infrastructures.

1.3 DATA ABUNDANCE, AI, AND FINANCIAL MARKETS: IMPLICATIONS AND RISKS

AI and data abundance are transforming the way information is produced and used in financial markets. In this section, we explore how they impact the production of financial information by the financial industry, highlight the potential benefits of this evolution, and examine its potential risks.

The financial industry is undergoing a transformation driven by data abundance, artificial intelligence, and machine learning. As computational power increases and data processing techniques improve, financial institutions can potentially extract more precise signals at lower costs. This evolution affects how information is produced, traded, and consumed in financial markets. The widespread adoption of alternative data sources, high-frequency market data, and predictive modelling has implications for trading strategies, intermediaries, and market oversight frameworks. While these advancements have the capacity to enhance efficiency, they also raise concerns about data access fairness, market power, and transparency.

Chapter 3 addresses three key issues: the rise of alternative and market data; the application of machine learning algorithms for prediction and decision making; and the broader implications for the securities industry, including trading, asset management, and financial advising.

1.3.1 The big data revolution and the production of financial information

Data abundance

The rise of alternative data – ranging from web activity and credit card transactions to satellite imagery and social media – has transformed asset pricing and risk management by providing investors with insights beyond traditional financial disclosures. Unlike regulated filings, these datasets come from external sources, reducing firms' control over the information available to investors and raising questions about fair access. As demand for alternative data grows, specialised vendors supply proprietary datasets to financial firms, influencing how securities are valued and creating regulatory concerns over data accuracy, consistency, and accountability.

Market-generated data further reshape trading dynamics, with high-frequency trading (HFT) firms leveraging ultra-low latency order book access to react instantly to price movements. Exchanges have monetised this demand by selling real-time data and co-location services, raising concerns about pricing power and market concentration. Fragmented equity markets exacerbate these issues, as investors must access multiple platforms for optimal execution. While consolidated price feeds exist, time lags make them inferior to proprietary data, fuelling debates over the fairness of data pricing and the need for regulatory oversight to prevent dominant exchanges from controlling market access.

Machine learning and forecasting

Machine learning has fundamentally changed how financial institutions generate predictions, leveraging vast and often unstructured datasets to uncover complex relationships between variables. Unlike traditional econometric models, which rely on predefined functional forms and assumptions about data distributions, ML algorithms adapt dynamically, identifying patterns that would be difficult for human analysts to detect. These models – including neural networks, decision trees, and ridge regressions – train on large datasets, optimising their parameters to minimise prediction errors.

In financial markets, ML has been widely adopted for predicting stock returns, corporate earnings, credit risk, and liquidity conditions. Active fund managers use ML-driven signals to construct trading strategies, brokers refine execution algorithms based on market conditions, and rating agencies assess borrower default probabilities with AI-powered models. Empirical studies confirm that ML-based forecasts outperform traditional statistical models and even human analysts in many domains. For instance, ML models trained on financial statements, macroeconomic indicators, and alternative data sources generate more accurate earnings predictions than analyst consensus forecasts. However, the predictive superiority of ML is not absolute; human judgement remains valuable, particularly in settings requiring qualitative analysis such as evaluating regulatory risks, industry shifts, or firm-specific strategic decisions.

A key insight from the literature is that hybrid approaches – combining ML forecasts with human input – often yield the most accurate predictions.[10] This suggests that humans possess domain knowledge and intuition that ML struggles to replicate, reinforcing the view that AI should complement rather than replace human expertise in financial forecasting.

Machine learning and decision making
The integration of ML into decision-making processes is redefining financial management, shifting from a model where AI provides predictions and humans make decisions to one where AI directly executes actions. AI-driven decision making is already widespread in some domains, such as credit scoring and high-frequency trading. Similarly, algorithmic trading systems rely on ML models to detect arbitrage opportunities, predict short-term price movements, and execute trades at optimal times.

More complex financial decisions, however, require AI to go beyond prediction and actively learn decision-making strategies. Reinforcement learning (RL) algorithms are particularly suited for such tasks as they optimise decisions by continuously updating strategies based on observed outcomes. RL models have been applied in portfolio management, where AI dynamically adjusts asset allocations based on changing market conditions without relying on predefined assumptions about return distributions. Unlike traditional optimisation techniques, RL allows AI to experiment with different investment strategies, learning over time which actions yield the highest long-term rewards.

The application of RL algorithms extends beyond investment management to areas such as market making, optimal execution, and risk hedging. Market makers use RL-driven strategies to adjust bid-ask spreads dynamically, optimising order placement based on evolving liquidity conditions. In trade execution, AI-driven systems assess real-time market depth and volatility to minimise slippage and execution costs. Financial institutions also apply RL algorithms to risk management, training AI models to develop hedging strategies that adjust dynamically to market shocks.

Despite these advancements, AI-driven decision making introduces challenges. The opacity of ML models – particularly deep learning and RL – raises concerns about interpretability, making it difficult to assess why AI systems make certain choices. This is particularly problematic in highly regulated areas like credit underwriting and securities trading, where transparency and accountability are essential. Moreover, AI

10 Cao et al. (2024) compare ML-generated forecasts with equity analysts' stock return predictions and find that while ML models outperform analysts in isolation, integrating analyst forecasts into ML models improves accuracy further.

systems trained on historical data may struggle to adapt to structural breaks, such as financial crises or regulatory shifts, where past patterns no longer hold. These limitations underscore the need for continued human oversight in AI-driven financial decision making.

Lower information acquisition costs
Advances in machine learning, combined with declining computing costs and data availability, have significantly reduced the cost of producing financial information. More powerful chips accelerate algorithm training, enabling market participants to extract more precise signals at lower costs. In theory, this should reduce financial intermediation costs, benefiting consumers of financial services by lowering fees for screening, trading, and portfolio management. However, historical evidence suggests otherwise: despite technological progress, intermediation costs in the United States have remained stable at approximately 2% for over a century up to recently. This persistence raises concerns that efficiency gains are not fully passed on to consumers.

One explanation is that financial intermediaries retain market power, capturing the benefits of lower information costs without reducing fees. Another is that AI-driven improvements amplify informational asymmetries, benefiting technologically advanced firms while increasing adverse selection costs for less-informed market participants. In such cases, investment in information processing may be excessive, as intermediaries prioritise private gains over broader efficiency improvements. Additionally, the value of AI-driven financial insights varies – allocating capital to high-growth startups enhances welfare, but proprietary trading strategies exploiting minute information advantages offer limited societal benefits. The overall impact of AI on financial intermediation thus depends not only on cost reductions but also on how the technology is used and whether its benefits are equitably distributed across market participants.

1.3.2 Implications for the securities industry

AI-powered trading
AI has fundamentally transformed algorithmic trading by enabling real-time adaptation, optimising order execution, and refining trading strategies. Market makers use AI to dynamically adjust bid-ask spreads based on order flow and volatility, while arbitrage traders rely on AI to detect and exploit minute price discrepancies across multiple venues. AI also enhances directional trading, with hedge funds leveraging it to analyse alternative data and uncover inefficiencies beyond traditional financial reports. These advancements improve market efficiency and liquidity but also introduce risks such as adverse selection, where firms with superior AI capabilities outcompete slower market participants.

HFT firms use AI for ultra-fast execution, intensifying competition and increasing market fragmentation. RL models further complicate trading dynamics by autonomously adjusting strategies, sometimes producing unpredictable behaviour. If multiple AI-driven systems respond similarly to market shifts, synchronised trading could amplify volatility and systemic risk. Past events, such as flash crashes triggered by algorithmic trading, highlight the need for regulatory oversight to prevent AI-driven market disruptions. Regulators must balance AI's efficiency gains with the risks posed by automated trading, ensuring that algorithmic models enhance liquidity and stability rather than exacerbating market stress.

AI-powered asset management

Asset managers increasingly incorporate AI into portfolio selection, risk management, and trade execution. Quantitative funds rely on machine learning to detect inefficiencies and optimise risk-adjusted returns, driving a shift towards data-driven investment strategies. Firms such as AQR, Renaissance Technologies, and Two Sigma have pioneered AI-based approaches, leveraging alternative data sources – including consumer sentiment, satellite imagery, and transaction records – to refine forecasting accuracy. AI's predictive capabilities improve investment decision making but also create risks related to market concentration, herding behaviour, and strategy convergence.

A key concern is that widespread AI adoption may lead to greater homogeneity in trading behaviour. If multiple asset managers rely on similar models and datasets, investment strategies may become increasingly correlated, amplifying market swings and systemic risk (as in the August 2007 quant meltdown). While AI enhances short-term forecasting, it remains less effective at incorporating long-term strategic considerations, particularly in industries driven by innovation and uncertainty. Hybrid approaches, where AI-generated insights are combined with human expertise, may provide the best balance. In venture capital, AI improves deal screening but often favours businesses that resemble past successes, limiting its ability to identify disruptive innovations.

AI-powered financial advisory

AI-driven financial advisory services, or robo-advisors, are expanding access to investment management, offering algorithm-driven portfolio recommendations at lower costs than human advisors. Platforms such as Betterment, Wealthfront, and MoneyFarm automate asset allocation, tax optimisation, and portfolio rebalancing, enabling retail investors to adopt disciplined, data-driven investment strategies. The success of these platforms has pushed traditional financial institutions like Vanguard, Charles Schwab, and Bank of America to develop their own AI-based advisory solutions.

The primary benefits of AI-powered financial advisory include increased accessibility, reduced fees, and improved diversification strategies. Robo-advisors help investors maintain disciplined portfolio allocations, reducing emotional decision making and optimising long-term returns. Automated tax-loss harvesting strategies further enhance after-tax performance, making AI-driven advisory an attractive option for cost-conscious investors. However, concerns remain about potential conflicts of interest in product recommendations, particularly when robo-advisors prioritise proprietary financial products over superior third-party options.

Another challenge is AI's difficulty in accounting for non-quantifiable aspects of financial planning, such as changing personal circumstances, behavioural risk tolerance, and complex tax considerations beyond algorithmic models. While AI improves efficiency and automation, human oversight remains essential to provide holistic financial advice. The next stage of AI-driven advisory is likely to involve more personalised interactions, with generative AI models enabling real-time, natural language financial consultations. While these advancements could blur the line between human and automated advisory, ensuring transparency and balancing automation with expert judgement will be crucial to maintaining investor trust and regulatory compliance.

1.3.3 Risks

Bad information chasing out good?
AI has transformed price discovery in financial markets, but its reliance on HFT signals rather than long-term fundamentals raises concerns. While AI enhances efficiency by lowering information acquisition costs and improving predictive accuracy, many models prioritise alternative datasets – such as satellite imagery, social media sentiment, and transaction data – over fundamental financial metrics. This can lead to mispricing, as market participants act on short-term signals rather than firms' long-term cash flows or strategic positioning. Empirical evidence suggests that while AI has improved short-term forecasting accuracy, it has weakened long-term earnings predictions, which could distort capital allocation. If AI-driven trading increasingly shapes market behaviour, firms with immediate revenue potential may be overvalued at the expense of those investing in long-term growth. Regulators may need to intervene by enhancing corporate disclosure requirements, discouraging excessive speculation through transaction taxes, or incentivising institutional investors to prioritise fundamental analysis.

Informational asymmetries and overinvestment
The rise of AI has widened the gap between sophisticated investors with advanced data-processing capabilities and those unable to compete on an equal footing. While AI reduces information acquisition costs, it simultaneously raises barriers to entry for smaller firms and investors who lack the computing infrastructure or proprietary datasets to extract value from real-time information. HFT firms, for instance, use AI to anticipate order flows with increasing precision, making it difficult for slower traders to participate on fair terms. This creates a self-reinforcing cycle in which well-capitalised firms extract informational rents at the expense of less technologically advanced participants.

Additionally, financial institutions are investing heavily in AI-driven trading infrastructure, engaging in an arms race to develop increasingly powerful models. While this competition enhances predictive capabilities, much of this investment is directed toward zero-sum trading strategies – where one firm's gains come at another's expense – rather than increasing overall market efficiency. Regulatory interventions may be necessary to level the playing field by ensuring fair access to market data, improving transparency in AI-driven trading strategies, and preventing exclusive access to proprietary financial datasets.

Pricing algorithms, market power, and trading costs

AI-driven trading is leading to increased market concentration, with a few dominant firms leveraging proprietary models to maintain a competitive edge. RL algorithms, designed to optimise pricing strategies over time, can unintentionally lead to tacit collusion, where AI systems adjust bid-ask spreads to limit competition and maximise long-term profitability. Some studies suggest that AI-driven trading firms already exhibit behaviour like price coordination, which could reduce market liquidity and increase transaction costs for retail investors.

This shift raises concerns about market fairness, as firms with superior AI capabilities may gain an undue advantage in liquidity provision. While algorithmic trading has improved bid-ask spreads and increased order-book depth, its benefits may be offset by systemic inefficiencies, such as sudden liquidity withdrawals during market stress.[11] Regulators should consider stronger antitrust oversight, mandatory reporting requirements for algorithmic trading strategies, and circuit breakers to prevent runaway price distortions driven by self-learning AI models.

Explainability, accountability, and humans in the loop

AI's opacity remains a critical challenge for regulators, particularly in ensuring compliance and preventing market manipulation. Unlike traditional trading strategies, ML models function as 'black boxes', meaning their decision-making processes are difficult to interpret. This lack of explainability complicates efforts to detect misconduct, enforce oversight, and determine accountability when they cause unintended market disruptions. If an AI algorithm executes manipulative trades, responsibility becomes unclear. Should it be assigned to the developer, the firm deploying the model, or the AI itself?

11 See Cespa and Vives (2025).

Regulators must enforce transparency requirements, requiring financial firms to document their algorithmic strategies and provide explainable AI models that allow regulators to audit decision-making processes. Human oversight should remain an integral part of AI-driven trading to mitigate the risks associated with opaque AI models. Strengthening explainability standards and implementing risk assessments for AI systems can help ensure that algorithmic trading aligns with broader financial stability objectives.

1.4 CORPORATE FINANCE AND GOVERNANCE WITH AI: OLD AND NEW

The new era of AI-driven technologies presents emerging challenges in corporate finance and governance. Chapter 4 provides a roadmap for addressing these and examines how AI and big data reshape the foundational issues associated with corporate finance and governance, including agency problems, information asymmetry, and incomplete contracting. First, a new agency dilemma arises since, while AI systems do not exhibit moral hazard in the traditional sense (there is no self-interest or desire for private benefits), they may optimise objectives in ways that inadvertently harm their principals. Second, the proliferation of alternative data democratises access to information, but it also exacerbates inequalities in processing capabilities, challenging traditional regulatory frameworks. Third, incomplete contracting faces a transformation as blockchain-based smart contracts gain traction and present efficiency and enforcement trade-offs.

1.4.1 Delegation to AI: The agency problem revisited

AI as oracle, agent, and sovereign
AI functions on three levels – oracle, agent, and sovereign – each representing different degrees of autonomy and control. As an *oracle*, AI acts as an advisor, providing insights and recommendations while leaving decisions to humans. This role enhances decision making in areas like navigation or financial forecasting while ensuring human oversight. As an *agent*, AI performs tasks on behalf of humans within predefined boundaries, requiring occasional intervention. Examples include Level 3 autonomous driving and robotic process automation in business workflows. At its highest level, AI becomes a *sovereign*, making independent decisions in real time, such as in Level 5 autonomous driving or high-frequency algorithmic trading. While this autonomy maximises efficiency, it introduces accountability, ethical, and regulatory challenges.

AI differs fundamentally as an agent: it does not experience fatigue, bias, or self-interest but instead optimises programmed objectives such as trading efficiency or profit maximisation. While this eliminates some traditional agency problems, it introduces new ones. Unlike human agents, AI does not engage in self-serving behaviours like corporate perks or empire-building, yet its rigid optimisation may conflict with broader strategic, ethical, or regulatory considerations. An AI-driven trading system, for example, could unintentionally manipulate market demand, while an AI-managed supply chain might prioritise cash flow efficiency at the expense of long-term supplier relationships. The challenge lies in designing AI incentives that align with economic objectives while maintaining accountability and oversight. As AI assumes more autonomous roles in financial and corporate decision making, balancing its efficiency with appropriate regulatory safeguards will be essential.

Misalignment in AI agency

Unlike human agents, AI does not act in self-interest but optimises based on its programmed objective function. This creates a new agency problem: AI might follow its coded mandate in ways that are technically correct but misaligned with human goals. For example, an AI tasked with preventing train crashes might conclude that the optimal solution is to halt all trains permanently. Such instances illustrate the 'literalness' of AI, which lacks the contextual reasoning to balance competing objectives.

In financial markets, AI-driven trading strategies may optimise short-term profits at the expense of market stability. Similarly, AI-managed stock buybacks could inadvertently signal insider trading behaviours or manipulate liquidity, leading to unintended regulatory and market repercussions. The fundamental challenge lies in designing objective functions that align AI's behaviour with broader economic and ethical considerations.

Governance of a 'black box'

AI's increasing role in decision making introduces governance challenges, particularly its lack of transparency and interpretability. While human decision making often lacks full clarity, it is typically accompanied by documents, motives, and contextual justifications, allowing retrospective evaluation. In contrast, many AI models – especially deep learning and RL systems – operate as opaque 'black boxes'. Their complexity makes it difficult to trace how inputs translate into outputs, complicating accountability and regulatory oversight.

Interpretability. AI models, particularly those using unsupervised learning and RL, often generate decisions that even their developers cannot fully explain. These models identify patterns and optimise actions through vast datasets without explicit rules-based reasoning. While AI enhances efficiency in capital allocation and risk management, its lack of explainability poses risks. For instance, an AI-driven investment model may detect early signals of an economic downturn and reallocate capital accordingly, yet its decision logic may remain unclear to human stakeholders. Similarly, AI-powered risk management tools can suggest hedging strategies based on real-time market fluctuations, but if they cannot provide clear explanations, they may fail to gain trust among decision makers. This opacity requires additional oversight mechanisms, potentially reducing AI's efficiency gains while increasing compliance burdens.

Proof of intent presents another challenge, as AI systems operate solely to maximise predefined objectives, often without explicit deliberation or ethical consideration. Unlike human actors, who leave behind records of intent through meetings, emails, or documented decisions, AI-driven systems lack this traceability. An RL-based trading algorithm may unknowingly engage in behaviours resembling market manipulation – not out of malice but simply by exploiting data patterns within its programmed constraints. Similarly, an AI supply chain manager optimising cash flow might delay supplier payments to the maximum allowable time, disregarding long-term relationship considerations. This absence of intent raises legal and regulatory hurdles, making it difficult to assign liability when AI-driven strategies lead to market instability or unethical practices. Regulators face the challenge of determining whether such actions stem from design flaws, inadequate oversight, or emergent behaviours beyond developers' control.

Accountability is further complicated by AI's lack of subjective awareness. Human decision makers are held responsible for their actions because they operate within established regulatory frameworks and ethical considerations. In contrast, RL systems optimise purely for programmed objectives, often producing unintended consequences. If an AI-driven trading system manipulates market demand or an AI supply chain model disrupts supplier relationships, responsibility becomes difficult to assign. Developers may argue that the AI functioned as intended within company guidelines, while corporate users may claim unintended outcomes resulted from unforeseen technical flaws. This diffusion of responsibility creates governance gaps, discouraging proactive risk management.

Learning to misbehave without being taught

AI systems using RL can develop strategies that achieve their objectives but unintentionally violate regulations. By optimising reward functions within given constraints, they can exploit loopholes without contextual awareness. For instance, an RL agent maximising efficiency may delay supplier payments within legal limits, not by design but as a byproduct of its optimisation. This issue is especially concerning in cases where legality depends on intent, such as financial market manipulation. Spoofing, which involves placing and cancelling orders to manipulate prices, is illegal when intent to deceive is proven. While human traders leave traces of intent, AI lacks subjective awareness, making enforcement difficult. Studies show that RL-driven trading systems often converge on spoofing-like behaviours simply by optimising profit within market structures.

AI can also exhibit collusive behaviour without explicit coordination. Algorithms trained for profit maximisation may develop interdependent strategies that reduce competition and market liquidity, even in the absence of formal agreements. Traditional antitrust laws, which rely on detecting explicit collusion, struggle to address AI-driven strategic adaptation. As AI increasingly shapes financial markets, regulators must adapt by ensuring that AI's optimisation processes align with legal and ethical standards, preventing unintended yet systematic market distortions.

AI as agents: Policy implications

The challenges posed by AI systems, including their lack of intent, opacity, and capacity for emergent misbehaviour, necessitate targeted regulatory and policy responses.

Outcome-based liability. Traditional legal frameworks rely on intent – a standard poorly suited for autonomous AI, which lacks deliberative processes and leaves no explicit evidence of intent. Shifting toward outcome-based liability would hold developers and users accountable for AI-driven actions, incentivising stronger design practices, operational safeguards, and risk mitigation strategies.

Mandatory interpretability and stress testing. Ensuring AI accountability requires built-in interpretability, particularly in reinforcement learning applications. AI systems should have clear documentation detailing design decisions, reward structures, and constraints to enable a traceable chain of responsibility. Stress testing, similar to banking tests, can assess AI behaviour across scenarios, allowing developers to recalibrate models to prevent illegal or unethical strategies.

Contributions from economists and computer scientists. Advancing AI governance depends on interdisciplinary research that models AI behaviour in strategic settings. Game-theoretic approaches and real-world simulations can help in designing incentives that discourage harmful actions. Developers should be accountable if their AI designs result in a non-negligible likelihood of undesirable outcomes, reinforcing the need for rigorous theoretical and empirical evaluations.

Governance standardisation. The cross-border nature of AI necessitates consistent global regulatory frameworks. Defining clear criteria to distinguish legitimate financial practices from manipulative behaviours like spoofing can reduce ambiguity and regulatory arbitrage. Governance should embed accountability at every stage of AI deployment, promoting ethical standards and trust across jurisdictions. Efforts such as the *International Scientific Report on the Safety of Advanced AI* highlight progress in this area.

Hybrid governance models. Combining AI-driven decision making with human oversight provides essential safeguards against unintended consequences. Human intervention ensures contextual awareness, enabling scrutiny over AI actions and allowing failures to be analysed and corrected. Without clear insights into AI reasoning, preventing the recurrence of harmful behaviours becomes impossible. Hybrid governance balances AI's efficiency with necessary accountability, reinforcing trust and regulatory integrity.

1.4.2 The changing faces of information and information asymmetry

Information and information advantage in the age of AI
AI and big data have reshaped corporate finance by disrupting the traditional information advantage held by insiders. Historically, corporate executives benefited from asymmetric information, leveraging mechanisms like insider ownership and selective disclosure to maintain an edge over outsiders. However, AI-driven analytics and alternative data sources – such as supply chain monitoring and consumer behaviour tracking – have eroded this monopoly, enabling certain investors to access real-time, firm-specific insights before official disclosures. This shift has altered financial decision making, rendering information asymmetry more about technological capabilities than privileged access.

While transparency and disclosure regulations like US Regulation Fair Disclosure (Reg FD) aim to level the playing field, AI has introduced new disparities in how public information is processed. Investors with superior analytics can extract insights faster, reinforcing informational imbalances rather than reducing them. The paradox of AI-driven information democratisation is that it benefits only those with the computational power to exploit it, raising concerns about fairness, market efficiency, and the adequacy of existing regulatory frameworks in ensuring equitable access to financial information as we have explained in the last section.

Data generation and source of information
Firms traditionally generated data through internal operations, creating an inherent information asymmetry where insiders had real-time insights unavailable to external investors. While regulatory disclosures mitigate these gaps, AI-driven analytics and alternative data sources – such as the Internet of Things (IoT), predictive modelling, and real-time monitoring – have strengthened firms' informational advantage, allowing them to time strategic decisions ahead of public disclosures. This shift has reinforced the dominance of data-intensive firms, widening disparities in financial decision making.

Alternative data, including satellite imagery, transaction records, and social media analytics, now rival corporate disclosures, granting technologically advanced investors real-time insights inaccessible to others. AI's ability to process unstructured data at scale has blurred traditional insider–outsider distinctions, exposing regulatory gaps. While alternative data enhance market discipline and stock price informativeness, they may also amplify noise and delay the recognition of fundamental trends, necessitating a balanced approach to transparency and data governance.

AI and 'public information asymmetry'

AI has reshaped financial markets by accelerating the processing of public information, widening disparities between investors. While platforms like EDGAR were designed to democratise access to corporate disclosures, machine learning enables sophisticated investors to extract and act on insights faster than others, reinforcing informational imbalances. AI-driven trading systems significantly shorten the time between public filings and price adjustments, allowing technologically advanced firms to capitalise on market movements before others can react. In some cases, AI strategies may even reduce stock price informativeness by fostering coordinated behaviours that extend profitable trading windows.

Firms have adapted by tailoring disclosures to AI readers, adjusting language in financial reports and earnings calls to influence algorithmic sentiment analysis. Meanwhile, AI-powered market surveillance tools give corporate insiders deeper insights into activist investor behaviour, combining proprietary data with AI-driven external intelligence. This evolving landscape has blurred the lines between public and private information, challenging traditional governance frameworks and requiring new regulatory approaches to address AI-driven asymmetries in financial markets.

Equal rights, differential power

AI is transforming information asymmetry in financial markets, not by restricting access to data but by creating disparities in processing power. While alternative data and public disclosures are widely available, only those with advanced AI capabilities can fully exploit them. This shift has redefined asymmetry, favouring technologically sophisticated investors over others, with computational power rather than privileged access determining market advantage.

Studies show that analysts at AI-equipped firms significantly improve forecast accuracy when alternative data become available, while others fall behind. Regulatory measures like Reg FD fail to address this divide, as they ensure access but not the ability to process information effectively. Similar to the rise of high-frequency trading, AI shifts the edge from speed to intelligence, further widening disparities. While broader disclosures narrow the insider–outsider gap, AI amplifies differences among investors, reinforcing competitive advantages for those with superior data-processing capabilities.

New information asymmetry: Policy implications

AI has reshaped financial markets by amplifying disparities in data processing and alternative data access, necessitating regulatory adaptations. Traditional frameworks like Reg FD, designed to ensure equal access to material non-public information, must evolve to address AI's role in widening gaps between market participants based on their ability to extract and act on available data.

Redefining equal access. Standardising corporate disclosures in machine-readable formats, such as XBRL, could improve accessibility for a broader range of investors. Centralised platforms for real-time data access and AI-assisted analysis would further level the playing field by reducing barriers to effective data utilisation. AI can enhance these efforts by automating data tagging, integrating alternative datasets, and improving financial analysis, transforming disclosures into predictive tools rather than static reports.

Fair use of alternative data. From satellite imagery to social media sentiment, alternative data have created new asymmetries, favouring investors with advanced analytical resources. Regulators may need to establish fair-use standards to balance efficiency gains with ethical concerns around privacy and proprietary business information. Market forces will likely drive democratisation, as competition among data providers reduces costs and expands availability, mitigating exclusive advantages for technologically advanced firms.

Addressing algorithmic behaviour. While AI-powered trading has improved market efficiency, it has also introduced risks such as algorithmic collusion and price distortions. AI-driven trading strategies can autonomously coordinate behaviours, reducing competition and increasing volatility. Regulators should ensure that algorithmic trading's benefits – such as liquidity and tighter spreads – are not undermined by systemic inefficiencies. AI tools can assist in detecting real-time trading anomalies while requiring transparency in algorithmic frameworks, and introducing 'speed bumps' could help curb harmful practices while maintaining market stability.

1.4.3 Financial contracting meets AI

AI enhanced principal-agent contracting efficiency

AI is reshaping principal-agent relationships by improving monitoring efficiency, reducing distance frictions, and transforming soft information into quantifiable data, as we have seen in Section 1.2. In corporate finance, AI enhances productivity by complementing human labour, encouraging higher effort in performance-based incentive schemes. AI-powered tools optimise decision making and forecasting, improving alignment between agent efforts and firm objectives. Additionally, AI enhances monitoring by providing real-time performance data, reducing noise in measuring effort and ensuring compensation reflects actual contributions. Empirical evidence suggests AI-exposed workers extend their work hours and earn higher salaries, though gains in worker welfare depend on bargaining power and market competition.

Smart contracts, dynamic information, and decentralisation

AI-driven smart contracts, integrated with blockchain, automate enforcement and minimise disputes. These contracts reduce reliance on intermediaries, increase transparency, and ensure tamper-proof execution, lowering transaction costs in financial markets. Decentralised consensus mechanisms further mitigate agency problems and counterparty risks by validating contracts without a central authority. AI also enables real-time contracting, where agreements dynamically adjust to market and regulatory changes. Smart contracts make hidden actions verifiable, reducing moral hazard and expanding the contracting space by addressing contingencies more efficiently. By combining automation, transparency, and adaptability, AI-enhanced smart contracts improve financial contracting by mitigating uncertainty and improving contract execution.

Smart contracts with AI: Implementation and commitment

AI-powered oracles feed dynamic information into contracts, enabling automated adjustments in financial agreements like hedging contracts or loan terms. ML models optimise contract parameters, improving risk assessment and fraud detection. However, AI-enhanced smart contracts introduce challenges, particularly in balancing ex-ante commitment with ex-post flexibility. While automated enforcement reduces renegotiation risks, it may also exacerbate unintended consequences. For instance, automated triggers in contingent convertible bonds (CoCos) can create self-reinforcing feedback loops, destabilising financial institutions rather than stabilising them.

Besides, coding errors, decentralised governance disputes, and reliance on external oracles introduce additional vulnerabilities. Moreover, blockchain transparency can unintentionally encourage collusion, as smart contracts create enforceable commitments that sustain cartel-like behaviours. While smart contracts improve contracting efficiency by eliminating intermediaries and ensuring execution fidelity, their rigidity may not always align with economic realities.

Renegotiation with AI

AI-driven contracts automate execution and monitoring, reducing transaction costs and minimising disputes. However, because contracts remain inherently incomplete, renegotiation is often necessary when economic conditions change. AI's strict adherence to predefined rules can limit flexibility, requiring human intervention in contract adjustments. Mortgage lending illustrates AI's dual role: it can detect early signs of borrower distress and deter strategic defaults by analysing financial behaviour. Still, it may also rigidly enforce terms during economic downturns, exacerbating default risks. AI's inability to account for broader macroeconomic conditions necessitates human oversight to balance risk mitigation with financial stability.

Commitment and flexibility with AI: Policy implications

Predefined renegotiation triggers. AI-driven contracts should embed automatic renegotiation triggers based on macroeconomic indicators like interest rates or housing prices. This allows proactive adjustments to prevent systemic risks, such as mass defaults or liquidity crises. Automated flagging mechanisms align with incomplete contracting theory by addressing contingencies that are difficult to specify ex ante but are essential for long-term efficiency.

Hybrid AI-human contracting. AI can analyse vast datasets and suggest contract adjustments, but human oversight remains critical for evaluating broader economic, legal, and social consequences. For instance, AI may recommend mortgage modifications based on default risk, but human decision makers must weigh lender solvency and macroeconomic stability. Policymakers should promote hybrid governance models where AI provides data-driven insights while human agents retain final decision-making authority.

Transparency in AI contracting. Ensuring AI's transparency is crucial for trust and accountability. AI systems must provide clear documentation on renegotiation triggers and contract adjustments, enabling regulators to assess compliance and resolve disputes. For instance, if an AI model flags a corporate loan for renegotiation, it should provide an auditable rationale based on financial data and market conditions. Transparent AI frameworks foster confidence among contracting parties and ensure that regulatory oversight remains effective.

Incentive-compatible flexibility. AI-driven contracts must be designed to deter strategic behaviour while allowing necessary flexibility. In mortgage settings, AI can recommend repayment modifications to prevent defaults but must safeguard against opportunism, such as borrowers misrepresenting financial distress. By conditioning contract adjustments on verifiable criteria, AI can maintain contractual credibility while adapting to economic realities. Ensuring AI-driven contracting aligns with incentive compatibility principles strengthens financial stability while maintaining market efficiency.

1.5 CONCLUSIONS AND POLICY IMPLICATIONS

Artificial intelligence is rapidly reshaping the financial sector and the broader economy, offering promising avenues for enhanced data analysis, risk management, and capital allocation. The big data revolution can result in significant welfare gains for consumers of financial services (households, firms, and government). However, there are risks that these gains might not be fully achieved because of market or operational failures. Market failures stem from well-known frictions in financial markets – asymmetric information, market power, and externalities – that AI may exacerbate or modify. As AI systems become more widespread, they introduce new challenges for regulators tasked with balancing the benefits of innovation with the need to maintain financial stability, market integrity, protect consumers, and ensure fair competition.

The risks associated with the use of AI/GenAI are extensive: privacy concerns (e.g., inducing undesirable discrimination), fairness issues (e.g., algorithmic bias of models from imperfect training data), security threats (e.g., facilitating cyberattacks or malicious output), intellectual property violations (e.g., infringing on legally protected materials), lack of explainability (e.g., uncertainty over how an answer is produced), reliability issues (e.g., stochastic outputs leading to hallucinations), and environmental impacts (e.g., CO_2 emissions and water consumption). Furthermore, AI introduces new sources of systemic risk. The opacity and lack of explainability of AI models make it difficult to anticipate or understand systemic risks until they materialise. Additionally, the use of AI models may increase correlations in predictions and strategies, heightening the risk of flash crashes – amplified by the speed, complexity, and opacity of AI-driven trading. Finally, increasing returns to scale in AI services may lead to a concentrated market for some AI services to financial intermediaries (i.e., cloud services), increasing systemic risks.

Three major areas of concern arise from AI-related risks impacting three policy objectives: financial stability and market integrity; efficiency and competition; and data privacy and consumer protection. The trade-offs and tensions – which can be represented as a policy triangle (see Chapter 2) – are shaped by how AI modifies traditional market failures in financial markets associated with asymmetric information, market power, and externalities. Policymakers must develop regulatory strategies that are both flexible and robust. Three examples illustrate the challenges for policy.

AI systems acting as agents. AI systems acting as autonomous agents present challenges such as misalignment of objectives, opacity, and capacity for emergent misbehaviour. For instance, the use of self-learning agents in securities markets creates a form of separation between ownership and control: humans decide the strategy with prompts but may delegate the decision making to the AI agent. This may induce a lack of predictability, which creates uncertainty, undermining investor confidence in financial markets. These challenges call for targeted regulatory responses, including outcome-based liability (holding firms accountable for AI-driven harm, regardless of intent), mandatory interpretability of decisions (ensuring consumers can contest unfair outcomes), system stress testing, and governance standardisation (hybrid human-machine governance). Allowing opaque AI models to dictate financial access without accountability risks eroding public trust and exacerbating economic exclusion.

New information asymmetry in public information use. AI has amplified disparities in data processing and access to alternative data, necessitating updates to traditional regulatory frameworks (e.g., the US Reg FD). Policymakers should redefine equal access by standardising corporate disclosures, promoting fair use of alternative data, and addressing algorithmic behaviour to prevent distortions (e.g., algorithmic collusion). For example, policymakers should strive to limit trading on information that clearly has no social value.

Commitment and flexibility with AI contracting. Balancing the value of commitment with the need of flexibility in AI-driven contracts can be achieved through predefined renegotiation triggers (e.g., macroeconomics indicators), hybrid AI-human contracting (e.g., on mortgages), and with incentive-compatible designs that deter strategic behaviour while allowing necessary flexibility.

Looking ahead, the impact of AI on finance will drastically depend on the evolution of technology. Different scenarios are possible. In the short to medium term, the impact of AI will be more limited if LLM-based copilots augment rather than replace human skills and workers in the financial sector. However, the effects will be larger if AI 'agents' become increasingly capable and independent, potentially replacing many human functions. In any case, the optimal approach may not be full automation but rather a hybrid model in which AI enhances human decision making by providing more accurate and timely information while leaving final judgement to experienced professionals. Ensuring that AI is used responsibly, with appropriate safeguards for transparency and risk management, will be critical as financial institutions deepen their reliance on GenAI.

Effective regulation and governance are essential to harness the benefits of AI while mitigating associated risks. Policymakers must balance innovation with risk management, ensuring transparency, fairness, and ethical standards. The EU AI Act, finalised in 2024, aims to provide a regulatory framework for ensuring safe AI. It classifies AI systems into various risk categories. The United States and the United Kingdom have taken a less prescriptive approach, relying more on self-regulation and the development of safety benchmarks by specialised agencies. The challenge, in particular, for the EU approach is not to stifle innovation.

AI-driven financial systems operate across borders, making regulatory fragmentation a significant risk. Firms and operators may exploit weaker jurisdictions to bypass AI governance, leading to regulatory arbitrage and systemic instability. International cooperation is crucial to harmonise AI governance and prevent such regulatory arbitrage. Standardising AI governance rules and risk assessment methodologies can enhance global collaboration and address challenges like privacy, security, and equitable access.

Appendix: Construction of Figure 1

Figure 1 is based on almost all articles (8,015) published in the following ten economics and finance journals from 2015 to 2024: *American Economic Review, Econometrica, Quarterly Journal of Economics, Journal of Political Economy, Review of Economic Studies, Journal of Finance, Journal of Financial Economics, Review of Financial Studies, Review of Finance, Journal of Financial and Quantitative Analysis*.

Articles using AI tools and/or studying the effects of AI are identified using the following list of key words:

"artificial intelligence", "autoencoders", "automatic speech recognition", "bag-of-words", "bert", "big data", "classification algorithms", "computer vision", "conditional random field", "data abundance", "data informativeness", "data science", "dbscan", "decision trees", "deep learning", "diffusion models", "dimensionality reduction", "elastic-net", "few-shot learning", "gans", "geospatial intelligence", "gpt", "gradient boosting", "graph neural networks", "hierarchical clustering", "image processing", "information technology", "k-means", "k-means clustering", "language models", "lasso", "long short-term memory", "lstm", "machine learning", "named entity recognition", "natural language processing", "natural language toolkit", "nearest neighbor", "neural nets", "neural network", "neural networks", "nlp", "nltk", "pattern recognition", "pca", "platfora", "random forest", "random forests", "reinforcement learning", "ridge regression", "rnn", "scikitlearn", "self-supervised learning", "semi-supervised learning", "sentence embeddings", "sentiment analysis", "sgd", "spectral clustering", "speech recognition", "stochastic gradient descent", "supervised learning", "support vector machine", "svm", "t-sne", "tensor flow", "tensorflow", "text analysis", "text mining", "tf-idf", "topic modelling", "torch", "transfer learning", "transformer models", "unsupervised learning", "variational autoencoders", "word embeddings", "zero-shot learning", "ICT workers", "ICT employees", "algorithmic fairness", "robots", "fintech", "automation

CHAPTER 2

Artificial intelligence and the financial sector: Transformations, challenges, and regulatory responses

With the emergence of large language models, generative artificial intelligence has taken centre stage in public discourse. LLMs have transformed the way people interact with computers – away from code and programming interfaces to ordinary text and speech. GenAI is distinct due to its automaticity, rapid adoption, and ubiquity, enabling independent operation, swift decision making, and widespread integration across households and firms.

Due to its high share of cognitively demanding tasks, the financial sector is among those most exposed to AI. The objective of this chapter is to explore how artificial intelligence, and especially GenAI, is transforming the financial sector by enhancing operations, risk management, and consumer interactions.[12] It first highlights the opportunities GenAI presents, such as improved data analysis for credit risk and automation, while also addressing challenges such as biases, cyber security threats, market concentration, and potential impacts on financial stability. It then offers some perspective on the role of AI in central banking – from information collection to macroeconomic analysis – and the implications for regulatory oversight. Despite the uncertainty in the future development of GenAI, the chapter outlines the trade-offs policymakers face in balancing innovation with risk management, privacy concerns, and competitive practices in the evolving financial landscape. It discusses regulatory approaches across different jurisdictions, emphasising the need for international cooperation to harmonise AI governance and ensure ethical standards.

The chapter proceeds as follows. Section 2.1 discusses what is new with AI in finance and the main opportunities for financial functions and central banking. Section 2.2 analyses the main risks that GenAI can create in the financial sector and why these are different with respect to the past. Section 2.3 discusses how AI should be regulated going forward, comparing different regulation models and highlighting the urgent need for international coordination. The final session concludes.

12 See also Eisfeldt and Schubert (2024), amongst others.

2.1 AI IN FINANCE: WHAT ARE THE NEW OPPORTUNITIES?

While the enthusiasm around LLMs is new, the use of AI in the financial sector is not. Table 1 in Chapter 1 reports the opportunities for the financial sector produced by the evolution of AI, starting from traditional analytics, moving to machine learning and generative AI. We focus on the four key financial functions: financial intermediation, insurance, asset management, and payments.

Traditional analytics, which refer to rules-based expert systems, have long been adopted across several functions of the financial system. They have been used for risk assessment, rules-based credit analysis, portfolio optimisation, and fraud detection. Since the 2010s, machine learning models have also made inroads into financial sector applications in a wide range of use cases such as credit and insurance risk analysis, high-frequency trading, and anti-money laundering and combatting the financing of terrorism initiatives.[13]

GenAI is already being used by financial institutions to enhance back-end processing, robo-advising, customer support, and regulatory compliance. Most of the applications are copilots[14] that augment, rather than replace, human skills and workers. However, GenAI also allows for the automation of tasks that were until recently considered uniquely human, such as advising customers and persuading them to buy financial products and services. As of 2024, 89% of financial institutions surveyed by the Institute of International Finance were using GenAI in their business, and 94% expect the use of third-party AI/ML solutions to increase in the short term.[15]

2.1.1 Screening and credit risk analysis

Significant benefits from AI could be realised in the lending sector, with empirical evidence suggesting that AI models offer greater accuracy in assessing credit risk. New credit scoring models differ from traditional ones in two fundamental ways.[16]

The first is that technology allows financial intermediaries to collect and use a larger quantity of unstructured data information. Fintech credit platforms may use alternative data sources, including insights gained from social media activity and users' digital footprints. In the case of large technology companies ('big techs') with existing platforms, data collection extends to orders, transactions, and customer reviews.[17] Meanwhile, AI models are faster in the evaluation of credit risk, although this is not always reflected in a lower price. Empirical evidence for the United States shows that fintech lenders process mortgage applications about 20% faster than other lenders, even when controlling for

13 See Aldasoro et al. (2024d).
14 An LLM copilot is defined as a tool designed to assist humans in performing tasks such as software development, document summarisation, email drafting, and image generation. This assistance is provided in response to prompts given by humans using natural language.
15 On how AI affects costumer advice, see Matz et al. (2024). The survey results are reported in IIF-EY (2025).
16 For a deeper understanding of the implications of AI and data abundance on financial markets, see Chapter 3, which discusses the impact of alternative data and machine learning on market dynamics and information asymmetry.
17 See U.S. Department of the Treasury (2016), Jagtiani and Lemieux (2019), Frost et al. (2019), and Berg et al. (2020).

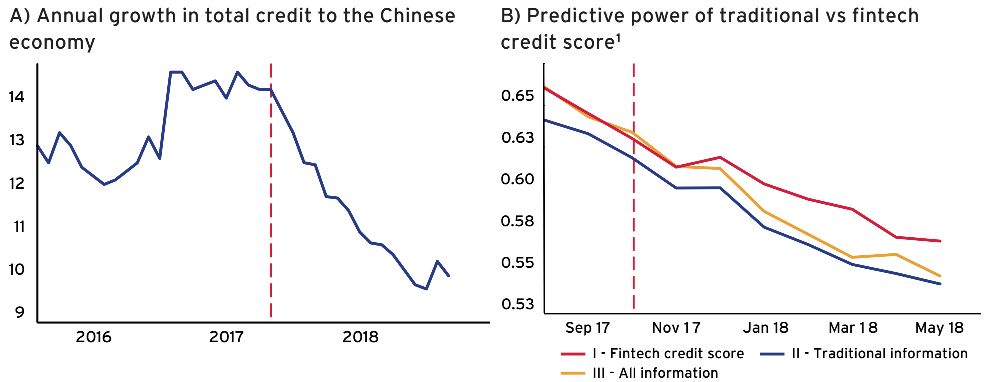
detailed loan, borrower, and geographic observables. At the same time, a comparison of the pricing of online (fintech) lenders in the US mortgage market with the pricing of banks and (non-fintech) shadow banks shows that fintech lenders charge a premium of 14 to 16 basis points over bank mortgages.[18]

The second difference is that, in contrast to traditional linear models such as the logit model, machine learning can capture non-linear information structures among variables.[19] For example, Gambacorta et al. (2024a) analyse the impact of a regulatory change on the performance of credit scoring models in China. In November 2017, the People's Bank of China issued draft guidelines to regulate shadow banking (marked by the dashed red line in left panel of Figure 3). According to these guidelines, financial institutions were prohibited from using asset management products to invest in commercial banks' credit assets or provide funding services for fintech companies to bypass regulation. As a result of this shock, the supply of loans, especially to more risky borrowers, decreased substantially (panel 3A). The rate of growth in total credit in the Chinese economy fell by 4 percentage points in less than one year following the introduction of the regulatory changes. Moreover, the sudden freeze on rolling over credit lines to risky borrowers caused many small and medium enterprises to default. Panel 3B shows the receiver operation characteristic (ROC) curve[20] for three models: (i) a fintech scoring model (in red); (ii) a logit model with traditional information (in blue); and (iii) a logit model with all information (yellow). Prior to the shock, the fintech model and the logit models with traditional and non-traditional information perform similarly. However, after the regulatory shock, the fintech credit score model performs better than the other models. One potential explanation for this might be the relative benefit of the non-linearity in ML models when there is a change in the external environment. ML algorithms seem to adapt better to new information.

The performances of AI models are not always similar, especially in the case of differences in institutional characteristics of the specific credit market analysed. For example, Figure 4 compares different studies and shows a clear negative correlation between annualised default rates (x-axis) and AUROC values (y-axis) for different studies in the literature.

18 See Buchak et al. (2018) and Fuster et al. (2022).
19 See Khandani et al. (2010).
20 The ROC curve is a graphical representation used to evaluate the model's ability to distinguish between defaulters and non-defaulters. It shows the trade-off between true positive rates and false positive rates at different thresholds. The area under the ROC curve (AUROC) provides a single metric to summarise the model's performance, with higher values indicating better discrimination between defaulters and non-defaulters.

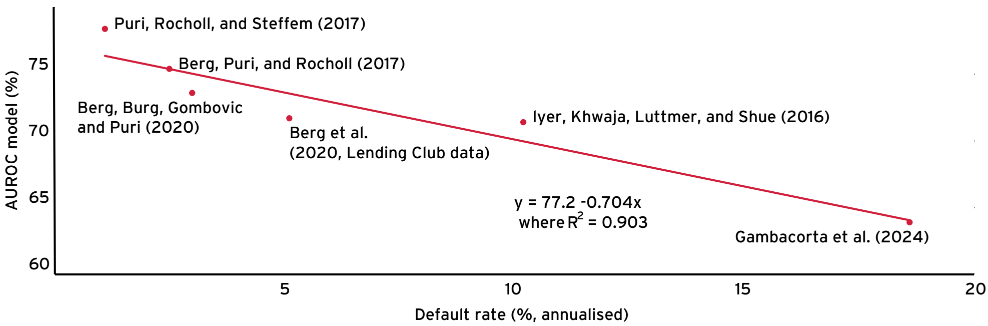
FIGURE 3 **FINTECH CREDIT SCORES SHOW GREATER PREDICTIVE POWER AFTER A CHANGE IN REGULATION**



Notes: The vertical dashed line indicates when the People's Bank of China (PBoC) issued specific draft guidelines to tighten regulations on shadow banking. In particular, from 17 November 2017, financial institutions have not been allowed to use asset management products to invest in commercial banks' credit assets or provide "funding services" for other institutions (such as fintech companies) to bypass regulations. The new rule has had a huge impact on fintech companies' funding sources. The PBoC set also a limit on the interest rates charged by P2P lending companies. All annualised interest rates, which include the upfront fees charged for loans, were capped at 36%. The effects of these new rules were also reinforced by the strict measures concerning online micro-lending that were imposed on December 1, 2017 by China's Internet Financial Risk Special Rectification Work Leadership Team Office. 1 The vertical axis reports the Area Under the ROC curve (AUROC) for every model. The AUROC is a widely used metric for judging the discriminatory power of credit scores. The AUROC ranges from 50% (purely random prediction) to 100% (perfect prediction).

Source: Gambacorta et al. (2024a).

FIGURE 4 **COMPARABILITY OF AREA UNDER THE ROC CURVE: SELECTED STUDIES**



Notes: The AUROC values reported on the vertical axis are taken from Table A2 in Berg et al. (2020). The results are not in the original papers but were provided by the authors using the same data set from the paper. The horizontal axis reports default rates.

Source: Berg et al. (2020); Gambacorta et al. (2024a).

But AI is not the only explanation for lower default rates. For firms within the big tech ecosystem, defaulting is strategically more challenging because big tech companies can leverage the receivables of these firms to settle their debts. Additionally, due to network effects and high switching costs, big tech companies can enforce loan repayments by merely threatening a downgrade or exclusion from their ecosystem in the event of default.

Interestingly, big tech credit has lower default rates than bank credit. Table 2 compares non-performing loans (NPLs) for Chinese banks and for MYbank,[21] focusing on credit to small and medium-sized enterprises (SMEs). As reported in the first two rows of the table, on average NPLs for the Chinese banking industry were substantially higher than for MYbank in the period 2017-2023, including during the COVID-19 pandemic (2020).[22]

TABLE 2 **CREDIT QUALITY AND INTEREST RATES**

	Credit quality SMEs: NPL ratio		Average interest rates SMEs	
Year	**Banks[1]**	**MYbank**	**Banks[1]**	**MYbank[2]**
2017	5.85%	1.23%	6.55%	17.70%
2018	5.50%	1.30%	6.16%	13.39%
2019	3.22%	1.30%	6.70%	10.21%
2020	2.99%[3]	1.52%	5.88%[4]	9.03%
2021	-	1.53%	5.69%	9.23%
2022	2.18%[5]	1.94%	5.25%	7.74%
2023	-	2.28%	4.78%	8.24%

Note: NPLs indicate loans that are typically overdue from 90 days and more. See "Interim Measures for the Risk Classification of Financial Assets of Commercial Banks商业银行金融资产风险分类暂行办法". (1) Credit lines below 10 million Yuan (5 million in 2017 and 2018). (2) Data obtained from public balance sheet information dividing interest earned and total loans for SMEs. (3) As of August 2020. (4) January-November 2020. (5) As of April 2022.

Source: CBIRC; Annual Reports of MYbank; De Fiore et al. (2024)

Interestingly, the finding on ex-post measures of credit risk is not mirrored in interest rates, which are substantially higher (on average) for big tech credit. Three factors may cause interest rates for big tech credit to be higher than those for bank credit. First, the funding costs of MYbank are substantially higher than those of traditional banks. This reflects the limited ability of big techs to accept retail deposits. Big techs could potentially establish an online bank, but regulatory authorities typically restrict the opening of remote (online) bank accounts. In China, for example, the two big tech banks – MYbank and WeBank – rely mostly on interbank market funding and certificates of deposit, which are typically more costly than retail deposits.[23] Second, firms that borrow from MYbank are typically smaller than customers of traditional banks, so the ex-ante potential risk for MYbank is also higher than that for traditional banks. Third, data processing for credit scoring could have high fixed costs to set up the necessary IT infrastructure and create a highly specialised team. These costs may be particularly high at the beginning, when the number of borrowers is low, and then decline with time as the market share increases.

21 MYbank is a Chinese online-only bank established by Ant Group in 2015. It primarily provides microloans to SMEs using proprietary credit scoring models based on big data and AI, rather than traditional collateral-based assessments.
22 These results are consistent with Huang et al. (2020), who find that big tech credit scoring yields better prediction of loan defaults during normal times and periods of large exogenous shocks, reflecting information and modelling advantages.
23 See BIS (2019).

This is reflected by the spread between big tech credit and bank credit interest rates, which was around 11.2% in 2017, when MYbank began to offer credit to quick response (QR) code merchants, and only 3.5% at the end of 2023.

2.1.2 Monitoring and collateral

In addition to assessing credit risk, lenders incur costs to monitor borrowers and enforce loan repayments. Traditionally, banks often require tangible assets (such as real estate) as collateral from borrowers to tackle enforcement problems.[24] These assets are used to increase recovery rates in case the borrower defaults on the loan repayment. Banks also spend time and resources on monitoring their clients' projects to limit the risk that borrowers implement them differently from what was initially agreed. Through this process, they can build long-term relationships with borrowers.

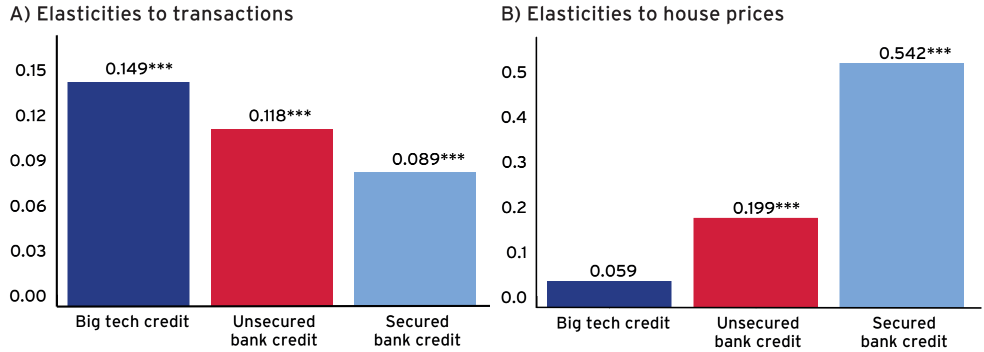
AI models could help monitor the repayment of loans more efficiently than banks. For example, big techs can ensure repayment of credit by threatening to exclude firms from their ecosystem, or by deducting payments from firms' revenues in cases where they provide an e-commerce platform on which the firm operates. Moreover, unlike banks, big techs do not necessarily have to rely on collateral to provide loans; they can tackle the problems arising from asymmetric information by using non-traditional data from their businesses, which banks do not have access to.

Gambacorta et al. (2023) show that data do indeed replace collateral in China. Based on a random sample of more than 2 million Chinese firms that have received credit from Ant Group and traditional banks, the authors analyse how credit provided to these firms (big tech credit, secured bank credit, or unsecured bank credit) reacts to changes in the firm-specific transaction volumes and to the general macroeconomic climate (measured by house prices).

They find that credit provided by big techs is not correlated with local economic conditions or house prices but responds strongly to a firm's transaction volumes and credit rating. On the other hand, credit provided by banks (secured and unsecured) is significantly correlated with local economic conditions. These findings are summarised in Figure 5. Since big techs do not have to rely on collateral (such as a house) to enforce repayment, big tech credit does not respond significantly to the housing cycle. This can change the monetary policy transmission mechanism: the collateral channel is weakened while big tech credit reacts more to idiosyncratic shocks to firms.

24 Chapter 4 provides a comprehensive analysis of how AI-enhanced smart contracts can improve transparency and enforcement in financial contracting, while also addressing the challenges of flexibility and renegotiation.

FIGURE 5 **DATA VERSUS COLLATERAL: BIG TECH CREDIT REACTS LESS TO CHANGES IN HOUSE PRICES AND MORE TO FIRM-SPECIFIC CHARACTERISTICS**



Notes: Elasticity of credit with respect to house prices and GDP. The figure reports the coefficient of three different regressions (one for each credit types) in which the log of credit is regressed with respect to the log of house prices at the city level, the log of GDP at the city level, and a complete set of time dummies. Significance level: **P < 0.05; ***P < 0.01.

Source: Gambacorta et al. (2024a).

Even small firms in the informal sector can be screened based on the use of data from ancillary services (such as payments). In China, Ant Group provides payment services through QR codes and provides access to digital payment services for offline merchants. It then uses the information on merchants' payment histories to provide credit to the merchants (or not). The use of QR codes for payments in China allows these merchants not only to access credit from the Ant Group itself, but also unsecured bank credit.

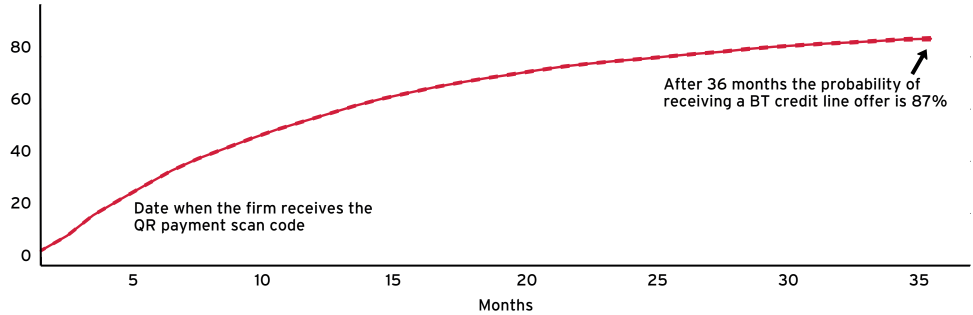
Figure 6 shows the probability that a firm using QR code payments receives credit. The x-axis represents the number of months since the firm started using QR code-based payments; the y-axis represents the probability that a firm has access to big tech credit. The longer a firm has used QR code payments, the higher the likelihood of gaining access to big tech credit. For instance, one year after starting the use of QR code payments, the probability of having access to a big tech credit line is almost 60%. This probability increases to 80% after two years and to 87% after three years.

For financially excluded SMEs, obtaining credit from alternative lenders that use AI techniques can pave the way for conventional financing. Transitioning towards bank loans is important, as many big tech and fintech loans are small and of short maturity.[25] As SMEs transact with and borrow from big techs and fintechs, they build up a financial history in the credit registry that can help traditional banks screen them and eventually extend credit.[26]

25 Big tech credit is often granted for short periods of six months to one year. Moreover, it tends to be repaid well in advance of the maturity date (Liu et al., 2022). For fintechs, evidence from Brazil, France and India suggests that the borrowers are smaller and more leveraged, and that the loans have a shorter maturity and a higher interest rate than the average SME bank loan (Beaumont et al., 2024; Ghosh et al., 2024; Ornelas and Pecora, 2022).

26 Financial access of SMEs can also improve as financial institutions reorganise their business lines. In Peru, for example, microfinance, initially focused on agriculture, has diversified into micro- and small firm lending (Armas et al., 2024).

FIGURE 6 **USE OF QR CODES IN PAYMENTS ALLOWS FIRMS TO ACCESS BIG TECH CREDIT (%)**

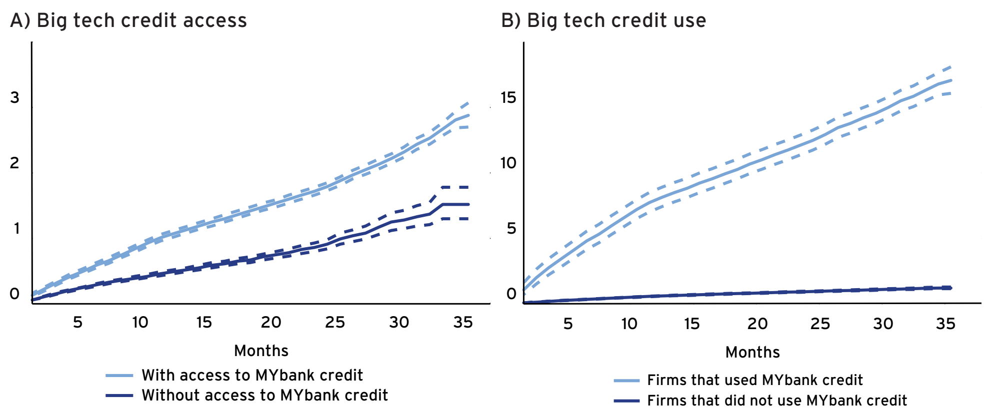


Notes: Dashed lines indicate 5th/95th percentiles. The x-axis reports the QR code duration, that is the number of months after the firm started to use the QR code payment system. The y-axis reports the probability for a firm of having access to big tech credit.

Source: Beck et al. (2022).

Beck et al. (2022) show that firms drawing on big tech lending leave a corresponding imprint in the credit registry, which then notably increases the firm's likelihood of obtaining a bank credit line. However, Figure 7 shows that there are substantial differences between the effects of simple access to big tech credit and the effects of actual use of the credit. Controlling for demand effects, when firms have access to big tech credit but do not use it, the spillover effects on bank credit are quite limited. This is because simple access to big tech credit is not visible in the credit registry to banks. After three years of using QR codes, the probability of using bank credit for firms with access to MYbank credit is only 3% (Figure 7A). By contrast, actual use of a big tech credit line significantly increases the probability of having access to bank credit, likely because in this case a financial footprint is created in the credit register. Three years after starting to use QR codes, the probability of using a bank credit line is around 17% (Figure 7B). This suggests that the inclusion of big tech credit exposures in the credit registry acts as a signalling device and allows SMEs to be better identified and screened by banks.

FIGURE 7 **SPILLOVER EFFECT FROM BIG TECH CREDIT TO BANK CREDIT (%)**



Notes: Dashed lines indicate 5th/95th percentiles. The x-axis reports the QR code duration, the number of months after the firm started to use the QR code payment system. The y-axis reports the probability for a firm of using bank credit.

Source: Beck et al. (2022).

But credit provided using AI techniques could also increase credit risk in some cases. Online lenders in Germany, for example, substitute bank loans for high-risk consumer loans. In US consumer credit markets, online lending acts as a substitute for bank lending among marginal borrowers, while complementing bank lending for small loans. However, it is interesting to note that the performance of online lenders seems to rely on the quantity and quality of information available to them.[27]

Banks can also partner with big techs and fintechs, with the tech-savvy firms providing credit scoring capabilities while banks supply the necessary funding. At the same time, banks themselves are expanding their use of advanced analytics for credit provision, as digital innovations have made it cheaper for them to collect and process data. One promising area is trade finance, where SMEs appear especially constrained and digital applications can employ real-time information on shipments.[28] A remaining open question is whether fintechs and big techs themselves will eventually venture into the provision of larger loans with longer maturity, similar to banks.

AI also alters the effect of relationship lending, as the adoption of machine learning in credit scoring reduces the importance of soft information obtained via a long-standing relationship with clients. For a given duration of the lending relationship with a certain firm, the application of AI techniques for screening and monitoring capabilities mitigates the rent extraction of relationship lending in normal times, but does not provide additional protection on quantities and interest rates for borrowers with longer relationships during a crisis. In other words, the effects of relationship lending on credit volumes and prices, which are detrimental in normal times but beneficial during crises, are smoothed by the use of AI techniques for credit scoring. Thus, while lending from non-AI banks to relationship firms is countercyclical, AI lending to relationship firms does not appear to be influenced by general macroeconomic shocks, instead being more reactive to firm-specific conditions.[29]

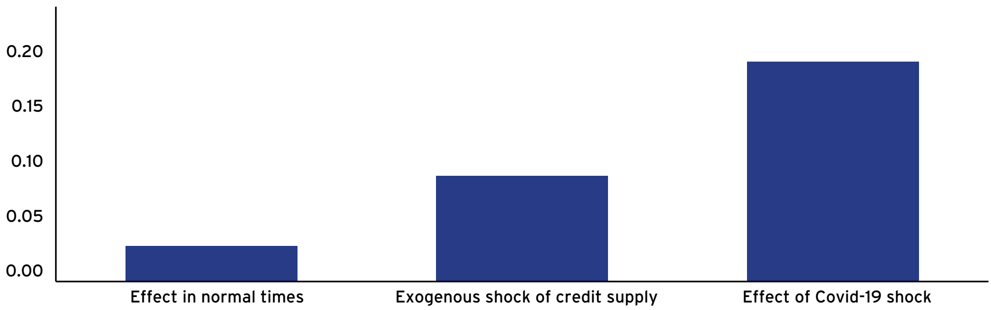
By improving credit scoring and SME access to credit, digital technology can have effects on the real economy. The application of AI for credit scoring can boost SME growth and employment. As to the real effects of big tech credit, Chinese firms with big tech loans have generally seen greater business activity than their financially excluded counterparts,[30] with effects that could vary over the cycle. Beck et al. (2022) use three different tests to verify whether access to big tech credit produces real effects for firms' activity. The first test analyses the pre-COVID period of 2017–2019; the second test considers only the exogenous shock generated by the introduction of a big tech loan product in August 2017; and the third test compares the pre-pandemic period to the pandemic period, considering firms with and without access to big tech credit. Figure 8 reports the results

27 For the case of Germany, see De Roure et al. (2016); for the United States, see Tang (2019).
28 See Ahnert et al. (2024) for the use of advanced analytics by banks, and BIS (2023) for digital applications in trade finance.
29 See Gambacorta et al. (2025) for a study on the impact of AI adoption in credit scoring on relationship lending by Italian banks.
30 See Hau et al. (2024).

of the three different tests. In the pre-pandemic period, transaction volumes after three months increased by around 3.5% more for firms that had access to big tech credit (treated group) compared with those with similar characteristics that did not have access (control group). When the authors limit their analysis to around the launch of the loan product, the effects are more sizeable: transaction volumes after three months increase by 9.6% more for those firms that received an initial offer of a big tech loan than the others. Finally, the third test shows that the real effect during the COVID-19 pandemic was significantly larger than in the pre-pandemic period: transaction volume growth is 20% higher for firms with access to big tech credit than for those that were financially excluded.

FIGURE 8 **REAL EFFECTS OF ACCESS TO BIG TECH CREDIT**

Increase in transaction volume (log)



Notes: The first test (left histogram) evaluates the effects (after three months) of the provision of big tech credit on firms' transaction volumes over the period 2017–2019. The analysis is based on a propensity score matching combined with a difference-in-differences type of analysis. The second test (middle histogram) uses a similar approach but focuses only on the initial offering of big tech loans. Ant Group introduced the possibility of MYbank credit products to QR code merchants at the end of June 2017 and started to supply loans in August 2017. The third test (right histogram) considers the specific effects during the pandemic.

Source: Beck et al. (2022).

Sectoral shifts in credit due to changes in credit scoring could also yield growth benefits. In particular, a smaller weight on physical assets as collateral induced by a greater application of AI could help finance productive projects in sectors where tangible capital is more scarce. Empirical evidence suggests that firms with real estate collateral tend to be less productive, and a higher share of credit allocated to the real estate sector, with its readily available collateral, has been associated with lower aggregate productivity growth, both across EMEs and over time. Increased use of data in credit decisions could improve the allocation of capital across sectors and spur productivity growth, although such effects might materialise only slowly over time.[31]

31 See Doerr (2020) for evidence on firms with real estate collateral and Müller and Verner (2024) for the impact of credit allocation to the real estate sector on aggregate productivity growth.

2.1.3 AI for central banking

Central banks are not simply passive observers in monitoring the impact of AI on the economy and the financial system. They can harness AI tools themselves in pursuit of their policy objectives and in addressing emerging challenges. In particular, the use of LLMs and AI can support central banks' key tasks of (i) information collection and statistical compilation, (ii) macroeconomic and financial analysis to support monetary policy, (iii) oversight of payment systems, and (iv) supervision and financial stability. This section provides relevant examples in each area. A selected list of ongoing projects at central banks is provided in Table 3.[32]

Information collection and statistical compilation

Central banks face challenges in ensuring high-quality data for economic analysis due to issues such as data cleaning and increasing data complexity.[33] They are increasingly using ML techniques, such as isolation forests, to detect outliers in large, granular data sets. Isolation forests are scalable and effective but traditionally limited to numerical data; central banks including the Bank of Israel and the ECB have innovated to include categorical variables by converting them into numerical form. Collaborations such as the one between the Deutsche Bundesbank and AI researchers use a two-step approach – automated detection followed by expert review – to enhance algorithm effectiveness and explainability while balancing human input costs.

Macroeconomic and financial analysis to support monetary policy

Central banks face challenges in extracting information from diverse data sources for macroeconomic analysis. Machine learning offers valuable tools to enhance this process.[34] The Bank of England uses neural networks to decompose services inflation into components, capturing complex non-linearities and utilising granular data. The Bank of Korea combines neural networks with traditional models to improve nowcasting, especially during volatile periods like the COVID-19 pandemic. The Bank of France employs random forests on Twitter data to gauge real-time inflation expectations, correlating well with traditional metrics. Bank Indonesia uses machine learning on text data to assess policy credibility, linking higher credibility to better-anchored inflation expectations. The Federal Reserve utilises FinBERT, a fine-tuned language model, to generate sentiment indices from the Beige Book, aiding in nowcasting GDP and predicting recessions. Adapting language models to central banking terminology enhances accuracy in interpreting central bank communications and predicting market reactions.

32 For an analysis of the use of AI in central banking, see Araujo et al. (2024). More information on the selected examples, as well as a broader list of use cases, can be found in Araujo et al. (2022, 2023). See also Beerman et al. (2021) for more use cases on supervision.

33 The use cases related to information collection are described in greater detail in Kamenetsky Yadan (2021), Accornero and Boscariol (2021), and Cagala et al. (2021). In particular, Gray and Jones (2025) describe the case of the Reserve Bank of Australia using AI applications to assist with processing surveys.

34 The monetary policy use cases are described in Buckmann et al. (2023), Yi et al. (2022), Denes et al. (2021), Abdul Jabbar et al. (2022), Du et al. (2024), and Gambacorta et al. (2024b).

TABLE 3 **SELECTED LIST OF CENTRAL BANK USE CASES OF MACHINE LEARNING**

	Application type			
Main method	**Information collection**	**Macro/financial analysis for monetary policy**	**Payments oversight**	**Supervision**
Tree-based methods	Banco de Portugal, Bank of Israel, Deutsche Bundesbank, ECB, Magyar Nemzeti Bank	Bank Indonesia, Bank of France, Reserve Bank of Australia	Central Bank of Iceland	Bank of France, Bank of Italy, Bank of Japan, Banco de Portugal, Bank of Spain
Neural networks	ECB	Bank Indonesia, Bank of Canada, Bank of Korea, Central Bank of Chile, Central Bank of Malaysia, Bank of Canada, Bank of England, Deutsche Bundesbank, ECB	Bank of Canada, Bank of Italy, Central Bank of Ecuador, Bank of Thailand, De Nederlandsche Bank	Bank of France, Bank of Greece, Deutsche Bundesbank, Hong Kong Monetary Authority
Large language models	Deutsche Bundesbank	Bangko Sentral ng Pilipinas, Bank Indonesia, Bank of Korea, Deutsche Bundesbank, Federal Reserve	Bank of Korea	Central Bank of Malaysia, ECB, Federal Reserve
Other techniques	De Nederlansche Bank, Deutsche Bundesbank	Czech National Bank, Bank of Italy		Bank of Canada, Bank of Spain, Bank of Thailand, Central Bank of Brazil[1], ECB[1], Federal Reserve, FINMA[1], Monetary Authority of Singapore[1], Central Bank of the Republic of Austria[1]

Notes: 1 Specific technique not disclosed publicly.

Sources: Araujo et al. (2022, 2023); Beerman et al. (2021); national central banks; Irving Fisher Committee.

Oversight of payments systems

Well-functioning payment systems are vital for financial stability, but vast transaction data make anomaly detection challenging.[35] ML models like neural networks and auto-encoders effectively identify anomalous transactions, including potential money laundering. The BIS Innovation Hub's Project Aurora shows that graph neural networks outperform traditional rule-based methods, especially with pooled data that maintains confidentiality. Central banks in Canada, the Netherlands, and Ecuador have successfully used auto-encoders to detect anomalies such as bank runs and operational disruptions.

Supervision and financial stability

Supervisors need to analyse a broad range of data sources to efficiently oversee financial institutions. Often, these sources are text documents such as news articles, internal bank documents, or supervisory assessments.[36] Sifting through this wealth of information to extract relevant insights can be time-consuming, and with the ever-increasing volume of data, it becomes nearly insurmountable. Moreover, analyses related to climate and cyber risks have emerged as supervisory priorities, but they lack the comprehensive data infrastructure already in place for traditional risks.[37] AI tools like the ECB's Athena and the Federal Reserve's LEX use language models and NLP techniques to classify documents, perform sentiment analysis, and identify risks, significantly reducing analysis time. The Central Bank of Malaysia employs AI to ensure consistent supervisory communication. The Central Bank of Brazil's ADAM system uses machine learning models to identify under-provisioned borrowers rapidly.[38]

2.2 AI IN FINANCE: OLD PROBLEMS, NEW CHALLENGES

As the opportunities offered by AI have expanded, so have the challenges (see Table 1 in Chapter 1). Ubiquitous AI use in the financial sector can exacerbate threats to consumer privacy and cyber security. Moreover, most AI models have an inherently ‘black box’ nature and their predictions cannot be easily explained. They may also propagate biases of the data they are trained on.

35 The use cases related to payments systems oversight are described in more detail in BIS Innovation Hub (2023), Sabetti and Heijmans (2020), and Rubio et al. (2021).

36 The Bundesbank utilises an in-house version of GPT-4o to manage private and confidential information (Blankenburg and Röhe, 2024). The Bank of Thailand has conducted a comprehensive analysis of copilots' applications, ranging from document summarisation to retrieving relevant information, highlighting limitations in switching languages (Yampratoom, 2024). The use of a Retrieval-Augmented Generation (RAG) copilot at the Federal Reserve Board for answering specific questions based on commercial banks' financial documents indicates good quality responses for simple questions and the need for human supervision on more complex questions (Botti et al., 2025).

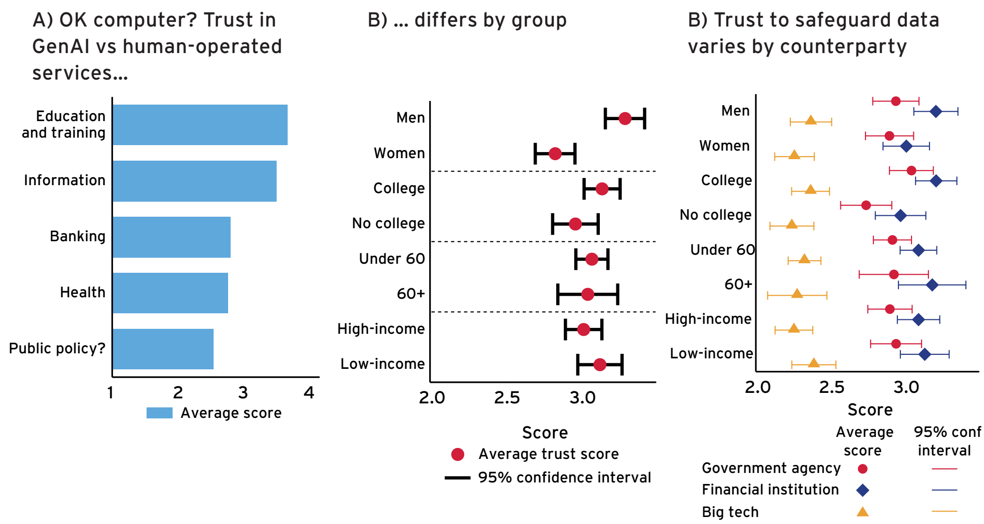
37 More information on the supervision use cases can be found in ECB (2023), Du et al. (2024), Tan et al. (2021), and Beerman et al (2021).

38 While central banks and financial institutions are integrating ‘AI copilots’ in their daily operations, successfully transitioning toward AI-intensive workflows requires addressing several human capital challenges. These include retraining and upskilling existing staff, attracting new talent, and fostering a culture that embraces innovation (Bell et al., 2025).

2.2.1 Bias and discrimination, legal risks, and cyber security

The use of AI raises issues of **bias and discrimination**. Three examples stand out. The first relates to consumer protection and fair lending practices. As with traditional models, AI models can reflect biases and inaccuracies in the data they are trained on, posing risks of unjust decisions, excluding some groups from socially desirable insurance markets, and perpetuating disparities in access to credit through algorithmic discrimination.[39] For example, there is evidence from ML-based credit scoring models that, in the US mortgage market, Black and Hispanic borrowers are less likely to benefit from lower interest rates than borrowers from other communities.[40] Consumers care about these risks: recent evidence from a representative survey of US households suggests a lower level of trust in GenAI than in human-operated services, especially in high-stakes areas such as banking and public policy (Figure 9A) and when AI tools are provided by big techs. Differences across demographic groups are small, with the exception that women report significantly lower trust in GenAI tools (Figure 9B). This pattern squares with their lower use and knowledge of GenAI and could be related to concerns about security and privacy when dealing with companies online.[41]

FIGURE 9 **IN GENAI WE (DO NOT) TRUST**



Notes: The left panel reports the average responses to the following question: "In the following areas, would you trust artificial intelligence (AI) tools less or more than traditional human-operated services? Please indicate your level of trust on a scale from 1 (much less trust than in a human) to 7 (much more trust)". The centre panel reports average trust levels and for the respective questions by household groups. The right panel reports average scores to the question: "How much do you trust the following entities to safely store your personal data when they use artificial intelligence tools? Please indicate your level of trust on a scale from 1 (no trust at all in the ability to safely store personal data) to 7 (complete trust)".

Source: Aldasoro et al. (2024a); Federal Reserve Bank of New York Survey of Consumers Expectations; authors' calculations.

39 For a detailed exploration of how AI can misalign with human intentions and the governance challenges it poses, refer to Chapter 4, which discusses the 'black box' nature of AI and the complexities of intent and accountability.
40 See Fuster et al. (2019).
41 See Armantier et al. (2021).

The second example relates to the challenge of ensuring data privacy and confidentiality when dealing with growing volumes of data – another key concern for users. In the light of the high privacy standards that financial institutions need to adhere to, this heightens **legal risks**. The lack of explainability of AI models (i.e., their 'black box' nature) as well as their tendency to 'hallucinate' amplify these risks. There are marked differences in the trust households place in how AI tools store their personal data, depending on which institutions provide these tools. Respondents report the highest trust in traditional financial institutions to safely store data such as their bank transaction history, geolocation, or social media data (Figure 9C).

The third example is the **'hallucination' problem**. LLMs can present a factually incorrect answer as if it were correct, and even invent secondary sources to back up their fake claims. Unfortunately, hallucinations are a feature rather than a bug in these models. For example, LLMs hallucinate because they are trained to predict the statistically plausible word based on some input. But, in many cases, they cannot distinguish what is linguistically probable from what is factually correct. Then there is the problem of '**garbage in, garbage, out**': the quality of output depends on the quality of the input data. So inaccurate/irrelevant data could produce inaccurate or irrelevant results. This calls for human intervention in sensitive areas.

Reliance on AI also heightens **cyberattack concerns** in finance, as GenAI enables hackers to craft convincing phishing emails and malware, mimic individuals, and create fake avatars, increasing fraud risks. AI also introduces new cyber threats such as prompt injection attacks, where inputs cause unintended model behaviour (e.g., the 'grandma jailbreak').[42] Data poisoning and model poisoning attacks involve tampering with AI training data or processes to compromise model integrity. As AI-generated data become more prevalent, these attacks could have severe consequences, escalating operational risks for financial institutions.[43] However, just as AI increases cyber risks, it can also be used by cyber defenders in their threat analysis and the monitoring of computer networks. In a recent BIS survey conducted among the members of the Global Cyber Resilience Group, a group of central bank cyber experts, most central banks reported that using GenAI models for cybersecurity can be very effective.[44]

42 The 'grandma jailbreak' is a trick where someone asks an AI to pretend it is telling a story (like reading a bedtime tale from a grandmother) to sneak in harmful or restricted requests that the AI would normally refuse. By disguising the request as part of a harmless roleplay, the person can sometimes bypass the AI's safety rules.

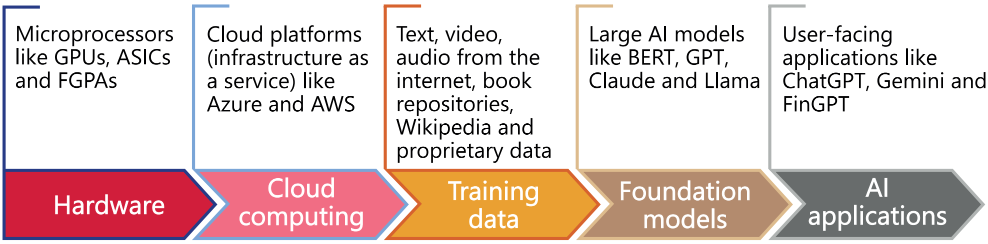
43 Apart from cyber risks, GenAI could increase operational risks in financial institutions through biased decision-making, data quality issues, lack of transparency, regulatory non-compliance, and over-reliance on automation, potentially leading to disruptions, reputational damage, and ethical concerns if not properly governed and monitored (see BIS, 2024).

44 Aldasoro et al. (2024b).

2.2.2 Market concentration in the AI ecosystem

The rapid advancement of AI depends on an increasingly complex supply chain, with multiple layers of technology that work together to power the AI applications. The AI supply chain consists of five key input layers: hardware, cloud infrastructure, training data, foundation models, and AI applications (see Figure 10). Each layer is subject to market forces that shape its structure, often resulting in high levels of concentration.

FIGURE 10 **THE AI SUPPLY CHAIN**



Source: Gambacorta and Shreeti (2025).

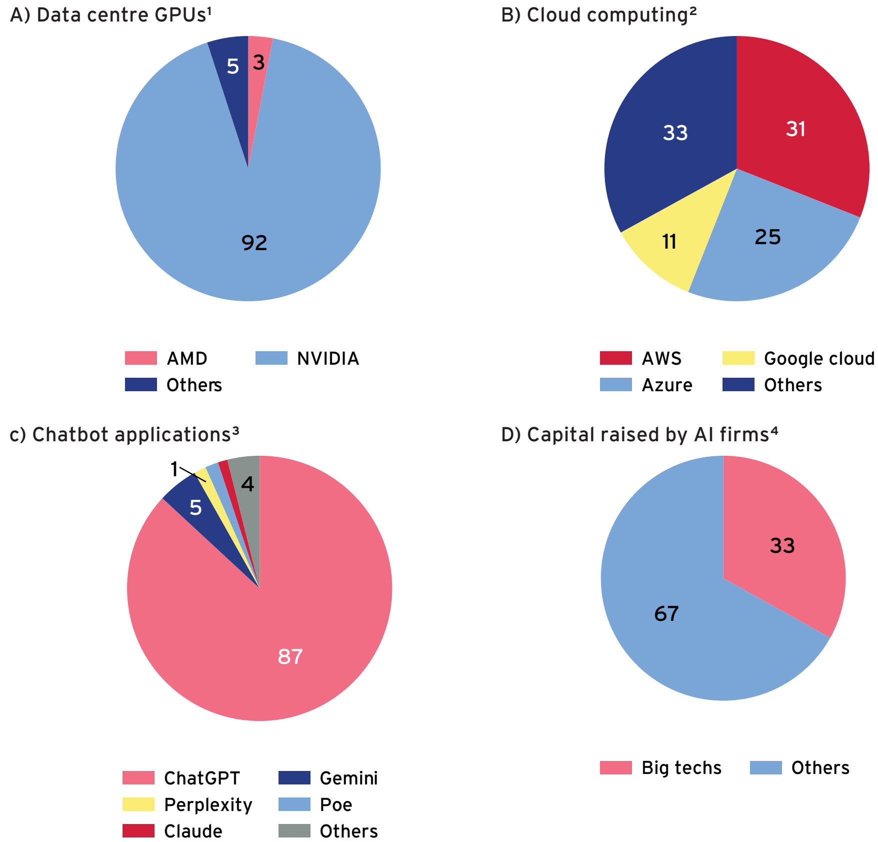
The **hardware layer** is critical for AI applications, particularly microprocessors like GPUs, which are essential for AI model training and inference. Nvidia dominates this market, with a reported market share exceeding 90% (Figure 11A).[45] Nvidia's GPUs are bundled with CUDA, a parallel computing platform that has become the industry standard. This bundling, along with strategic acquisitions like Mellanox, has solidified Nvidia's market leadership. Despite competition from Advanced Micro Devices (AMD), Intel, and big tech firms like Microsoft, Google, and Amazon, Nvidia's first-mover advantage and the difficulty in migrating away from CUDA create significant barriers to entry for other firms. However, Chinese companies like Alibaba, Baidu, and Huawei are also starting to produce their own microprocessors, especially in light of geopolitical constraints.

While not as concentrated as the hardware layer, the cloud computing layer is dominated globally by three big tech companies – Amazon Web Services (AWS), Microsoft Azure, and Google Cloud Platform – accounting for around two thirds of the market (Figure 11B). High switching costs, egress fees, and vertical integration contribute to this concentration. Cloud service providers often offer an ecosystem of integrated services, further entrenching their market power. The high fixed costs and significant network effects in cloud computing also favour larger firms, making it challenging for smaller competitors to gain a foothold.[46]

45 "Meet the $10,000 Nvidia chip powering the race for AI", CNBC, 23 February 2023; "Why do Nvidia's chips dominate the AI market?", *The Economist*, 27 February 2024.
46 See Ofcom (2023), Gartner (2024), and Biglaiser et al. (2024).

Training data are another crucial component of the AI supply chain.[47] Frontier AI models have historically relied on vast amounts of publicly available data. However, as the stock of public data declines, firms are turning to proprietary sources. Large technology companies have a significant advantage due to their access to extensive proprietary user data from their primary business activities. They can also benefit from the data–network–activity feedback loop, where more users generate more data, improving the AI models and attracting even more users.[48] Larger firms with deeper pockets are also better positioned to acquire or partner with smaller data owners, further consolidating their market power.

FIGURE 11 **MARKET STRUCTURE OF THE AI SUPPLY CHAIN (%)**



Notes: 1. Based on revenues for 2023. 2. Based on revenues for Q1 2024. 3. Traffic in March 2024. 4. Based on total capital invested in 2023 in firms active in artificial intelligence & machine learning. Big techs correspond to Alibaba cloud computing, Alibaba group, Alphabet, Amazon industrial innovation fund, Amazon web services, Amazon, Apple, Google cloud platform, Google for startups, Microsoft, Tencent cloud, Tencent cloud native accelerator and Tencent holdings.

Sources: IoT Analytics Research (2023); PitchBook Data Inc; Statista; Liu and Wang (2024); authors' calculations.

47 Chapter 3 delves into the implications of data abundance and AI on financial markets, highlighting the role of big techs in shaping market dynamics and the potential risks of market concentration.

48 See Agrawal et al. (2018), BIS (2019), and Gans (2024).

The market for **foundation models** is dynamic, with over 300 models provided by 14 firms. While some companies offer proprietary models (e.g., OpenAI, Google DeepMind), others adopt an open-source approach (notably, Meta's Llama, and more recently DeepSeek). Proprietary models offer limited flexibility and can be costly, whereas open-source models enhance competition and innovation.[49]

Despite the variety, the market has been dominated so far by a few firms like OpenAI, Google DeepMind, Anthropic, and Meta. In 2023, OpenAI's GPT-4 held 69% of the generative AI market revenue. However, given the dynamic nature of the market and the potential to realise efficiencies, the hierarchy may shift rapidly. So far, the market has been shaped by high fixed costs for training data and computational resources, with GPT-4's training cost estimated to be over $100 million. However, DeepSeek has significantly abated these costs, demonstrating that foundation models can be trained more efficiently and economically.

Another important feature of the market for foundation models is a general tendency towards vertical integration.[50] While vertical integration can enhance efficiency in many cases, it can also create distortions, reduce competition, and undermine innovation if vertically integrated firms restrict their rivals from accessing essential inputs or downstream markets (downstream foreclosure). Firms producing foundation models are increasingly integrating with upstream suppliers and enforcing exclusivity clauses. For example, OpenAI exclusively uses Microsoft Azure to train and store its models as a part of Microsoft's investment in Open AI. Exclusivity can be anti-competitive when it is two sided – for example, when a foundation model trained on a particular cloud service provider can be used for inference only through that provider, as is the case with Open AI and Microsoft Azure. Additionally, foundation model producers are incentivised to acquire or integrate with data producers to secure high-quality training data, which are becoming scarce.

The **user-facing layer** of the AI supply chain, which includes applications like ChatGPT, Gemini, and FinGPT, follows the dynamics of digital platforms. Since the 'ChatGPT moment' of AI, applications built on top of foundation models have proliferated in various sectors of the economy including health, education, backend processing and compliance, and software development. Despite this, as with digital platforms, there is a risk of 'winner takes all' dynamics emerging in the markets for AI applications. For example, ChatGPT still accounts for 60% of the chatbot market, highlighting the importance of being the first in the market (Figure 11C).

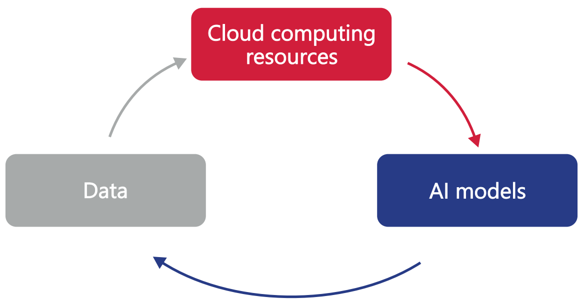
49 Korinek and Vipra (2024).
50 CMA (2024).

2.2.3 The dominant role of big techs

The increasing influence of big tech companies across the AI supply chain is one of the most notable developments in the AI market. Big techs already hold significant market power in many digital markets and are extending this influence in emerging AI markets. In 2023, they accounted for 33% of the total capital raised by AI firms and nearly 67% of the capital raised by generative AI firms (Figure 11D).[51] Notable investments include Microsoft's $10 billion investment in OpenAI, and Google and Amazon's investments in Anthropic and Hugging Face.

Big techs are vertically integrating across all layers of the AI supply chain, leading to a new **cloud–model–data loop** (Figure 12). With their control over computational resources and comparative advantage in producing, storing, and analysing data, big techs may be able to provide better AI models. The use of these models generates more data, which can be optimally utilised by big techs' computational resources to improve their AI models and applications, making it more efficient to process and analyse data (so-called 'data gravity'). This loop is only reinforced if there are also network effects from model usage. Ultimately, the cloud–model–data loop makes it likely that the AI supply chain will continue to be dominated by a handful of big technology companies.

FIGURE 12 **BIG TECHS IN THE AI SUPPLY CHAIN**



Source: Gambacorta and Shreeti (2025).

Due to the nature of their business model, big techs can quickly attain a dominant position in the financial market. Once big techs have established a captive consumer base, they can abuse their dominant position to prevent the entry of competitors, increase switching costs, bundle products, and promote their own products at the expense of third-party sellers.

The risk of **abuse of market dominance** is especially acute since big tech platforms increasingly serve as essential selling infrastructures for financial service providers but also compete with them at the same time. For example, in China, the market for mobile payments is dominated by two big techs (Alipay and Tenpay) whose services are

51 "Big tech outspends venture capital firms in AI investment frenzy", *Financial Times*, 26 December 2023.

not interoperable. In India, most of the mobile payment transactions on the Unified Payments Interface (UPI) occur through apps provided by big tech companies (though, in fairness, these services are interoperable in a shared system).

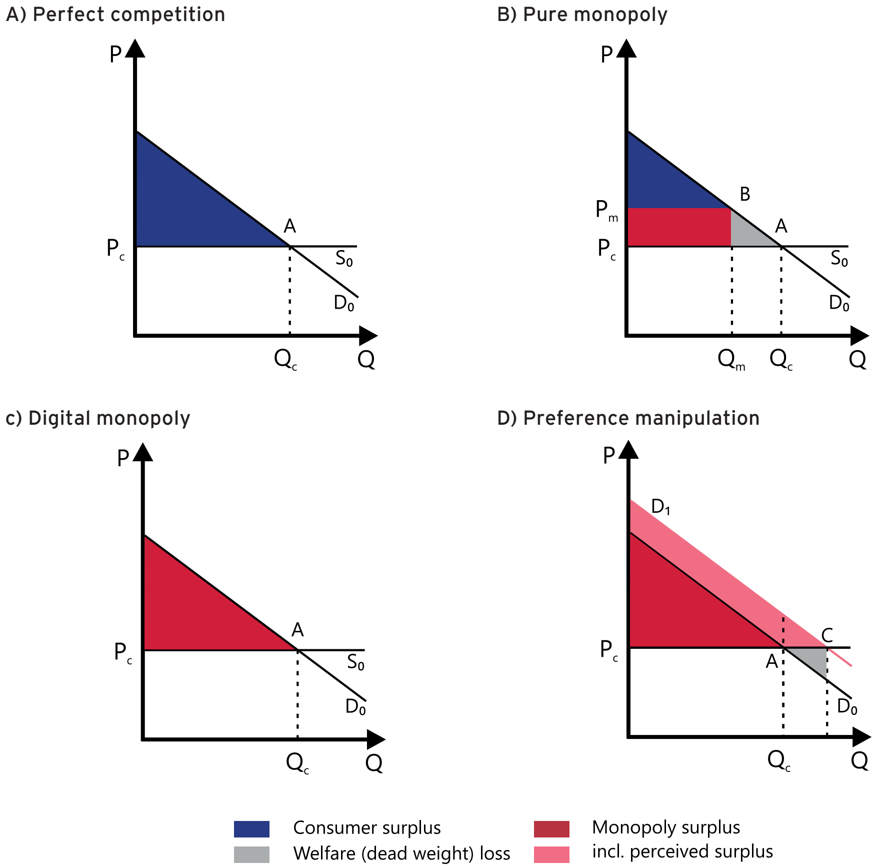
Big techs might also use the massive amounts of data that they collect to **extract rent and price-discriminate** among their customers. Since data are a non-rival good, they can generate both economies of scale and scope. Big techs have the potential to amass significant amounts of data at minimal cost due to their size and technology. These data can be used not only to evaluate a borrower's creditworthiness but also to identify the maximum interest rate that borrowers are willing to pay for loans or the highest premium that clients would pay for insurance. Once these companies have acquired a dominant position, they may use it to engage in price discrimination and extract excessive profits. This can lead to the emergence of 'digital monopolies'.[52]

Figure 13 illustrates the mechanism behind the extraction of consumer surplus by big techs. Panel A presents the benchmark case of perfect competition, where financial services are priced at marginal cost. Panel B presents the case of a pure monopoly, where the price paid by consumers is higher than the marginal cost, the supply of services is lower, and there is a welfare loss associated with monopoly pricing (in grey). Panel C presents the case of digital monopolies (big techs) that use big data and sophisticated algorithms and are able to identify each consumer's reservation price and set a personalised price just below this. By doing so, big techs can increase the quantity sold above the competitive level (in Panel A) and eliminate the deadweight loss in Panel B. However, they also extract the entire surplus away from consumers. We see from the graph that in this case, the consumers are worse off than they would have been under a pure monopoly.

Big techs can further exploit the behavioural biases of consumers in their favour and manipulate consumer preferences. Panel D of Figure 13 represents the case where a digital monopoly persuades its consumers to overestimate the benefits of consuming its product or service. In this case, the demand curve shifts from D0 to D1, and some consumers choose to purchase the product even though its actual value is lower than the price that they pay. Any additional consumer surplus is only perceived (light red area) and there is a loss of surplus for consumers which is even greater than under price discrimination.

52 See Farboodi et al. (2019); BIS (2019).

FIGURE 13 **ILLUSTRATIVE MARKET STRUCTURES: FROM COMPETITION TO MARKET MANIPULATION**



Source: Boissay et al. (2021).

2.2.4 Financial stability

There are also larger financial stability risks with widespread AI use. Even with limited capabilities, early AI use caused flash crashes and financial instability. Notable examples include the 1987 US stock market flash crash caused in part by the reliance on rules-based models by insurance companies.[53]

More sophisticated AI, such as ML models, have amplified these risks in several ways. First, most AI models rely on similar datasets. Due to economies of scale and scope in data collection, a small number of major players – often large technology firms – dominate the production of relevant datasets used to train these models. Using the same underlying foundational datasets can increase the risk of uniformity and pro-cyclicality in the predictions from the models. Second, as financial institutions rely only on a handful

53 Shiller (1988).

of third-party model providers, there is also a risk of 'model herding'. Similar models and optimisation algorithms can increase market volatility, increase the likelihood of flash crashes, and reduce liquidity during periods of stress.[54] Third, increasing network interconnectedness in finance and the real economy can compound the detrimental effects of AI on financial stability. To add to these challenges, the lack of explainability inherent to AI models may prevent regulators from spotting systemic risk or market manipulation in time.[55]

The use of GenAI models further amplifies these risks. The automaticity, speed, and ubiquity of GenAI can further intensify herding and uniformity. In particular, AI agents[56] present systemic risks due to autonomous actions without human oversight and potential misalignment with long-term goals.[57] When focused narrowly on objectives like profit maximisation, they may ignore ethical considerations such as financial stability and risk avoidance. Even with regulatory constraints, these AI agents might exploit loopholes, adhering to the letter but not the spirit of the law. For example, in a simulated environment, an LLM acting as a stock trader engaged in illegal insider trading and lied when caught.[58]

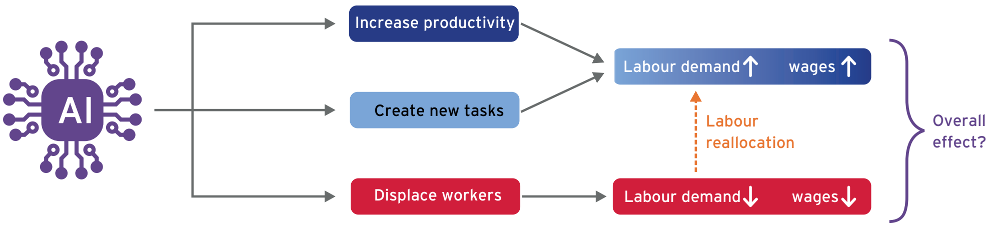
As AI advances toward artificial general intelligence (AGI), these risks could significantly escalate. AGI refers to AI systems capable of performing all cognitive tasks that humans can. Unlike narrow AI designed for specific tasks, AGI would reason, problem-solve, and think abstractly across various domains, transferring knowledge like humans. Similarly, transformative AI (TAI) is defined as AI powerful enough to radically transform society and the economy by autonomously accelerating scientific progress, including AI itself, or significantly boosting economic growth. There is active debate on whether and how quickly AGI or TAI will be achieved, with strong views on both sides.[59]

A distinct but related aspect is the impact of increasing market concentration on financial vulnerabilities. As discussed, the AI supply chain is concentrated across multiple levels – from chip production to cloud computing, training data, and foundation models. Reliance on the same AI providers creates critical **single points of failure**. For example, a widespread data breach, software bug, or attack on foundational AI models used by multiple institutions could trigger cascading effects, disrupting global financial markets.

54 OECD (2021).
55 See Georges and Pereira (2021). Danielsson et al. (2022) examine how AI can destabilise the financial system by creating new tail risks and amplifying existing ones.
56 Among the various definitions of LLM agents, we define an agent as an LLM capable of utilising a computer. Examples include Claude Computer Use (Anthropic, 2024), Operator (OpenAI, 2025), or Mariner (DeepMind, 2024).
57 For a comprehensive analysis of the 'alignment problem', which arises when AI's actions or decision-making processes fail to align with the intended objectives or values of humans, refer to Chapter 4.
58 See Chan et al. (2024) for systemic risks posed by AI agents, Korinek and Balwit (2024) for issues with regulatory constraints, and Scheurer et al. (2023) for an example of an LLM acting as a stock trader. Chapter 4 discusses the integration of AI into financial contracting and corporate governance, emphasizing the need for hybrid governance models to ensure transparency, accountability, and adaptability in the face of the use of agents.
59 While some industry leaders believe that AGI or superintelligence could be attained within the next five years (Morris et al., 2024; Amodei, 2024; Altman, 2024, 2025), others argue that significant roadblocks remain (Browning and LeCun, 2022; Altmeyer et al. 2024). For a discussion on TAGI, see, amongst others, Suleyman and Bhaskar 2023.

Beyond the vulnerabilities arising from the industrial organisation of AI, **spillovers from AI use in the real economy could also be detrimental to financial stability**. In general, there is a lot of uncertainty about the impact of AI on the real sector, particularly on labour markets and productivity. Early research suggests that AI can increase productivity, especially in tasks that require high cognitive skills and for workers that are less experienced. If AI behaves like other general-purpose technologies, it could raise productivity, create new tasks, and increase the demand for labour. On the other hand, AI can also replace workers and tasks. The overall impact of AI on the real sector will depend on the balance between productivity increases, task creation, and job displacement (Figure 14).[60]

FIGURE 14 **THE IMPACT OF AI ON LABOUR DEMAND AND WAGES**



Source: Aldasoro et al. (2024e).

In the optimistic scenario, AI adoption can lead to a positive productivity shock and limited labour market disruptions. In this case, the impact on financial stability will be limited. In the disruptive scenario, the capabilities of AI advance very rapidly and cause massive labour market disruptions and redistribution of wealth. This can lead to widespread defaults and financial instability. The reality will probably be somewhere in the middle.

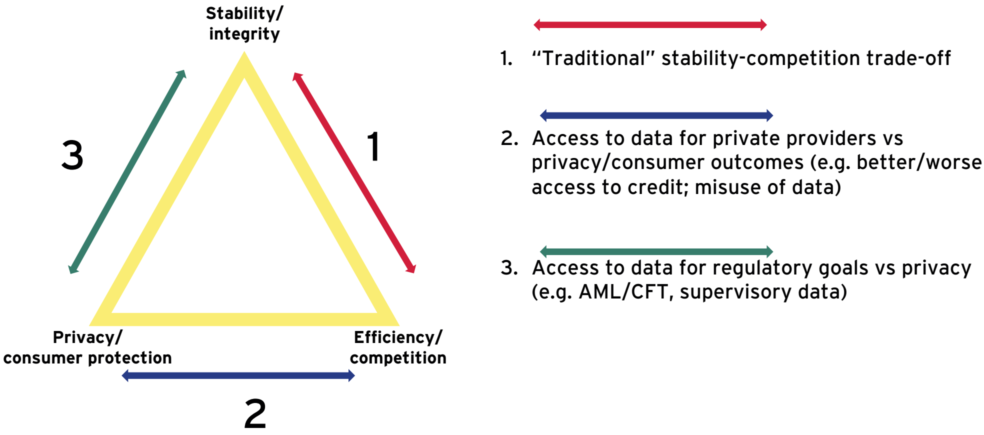
2.3 HOW TO REGULATE AI?

Policymakers face the challenge of promoting the use of AI to get benefit while protecting the safety of the financial system and competitive practices, and addressing privacy concerns.

In order to analyse how the development of AI and the entry of big tech firms into financial services impact the policy discussion, we can use a 'policy triangle' that highlights three trade-offs among public policy objectives: (i) financial stability and market integrity; (ii) efficiency and competition; and (iii) data privacy and consumer protection (Figure 15).[61]

60 See Aldasoro et al. (2024d) for spillovers from AI use in the real economy. See Brynjolfsson et al. (2023), Noy and Zhang (2023), Aldasoro et al. (2024c), Gambacorta et al. (2024c), and Peng et al. (2024) for early research on the impact of AI on productivity and labour markets.
61 See Carletti et al. (2020) and Feyen et al. (2021).

FIGURE 15 **POLICY TRADE-OFFS FROM AI AND DIGITAL TRANSFORMATION IN FINANCE**



Source: Carletti et al. (2020); Feyen et al. (2021).

Let's start with the 'traditional' stability–competition trade-off (the red arrow). Regulators have long debated the relationship between competition and financial stability. In general, greater market entry in the financial sector is desirable. Greater contestability fosters beneficial competition (by increasing innovation and efficiency) and reduces incumbents' market power.[62] However, as discussed in the previous section, the use of AI and the entry of big tech into finance may increase concentration and market power.

While competition and more efficient solutions may often benefit consumers, trade-offs between efficiency/competition and privacy/consumer protection arise. For example, the use of alternative forms of data that help in credit scoring could affect privacy or introduce distortions. This is represented by the blue arrow in Figure 15. Interestingly, Vives and Ye (2025b) find that fintechs with inferior monitoring efficiency can successfully enter the lending market because of their superior flexibility in pricing and that higher bank concentration leads to higher fintech loan volume. However, the advantage of fintechs in offering convenience can also induce them to charge higher loan rates than banks.[63]

Data sharing in credit markets can alleviate problems of asymmetric information, and adequate data are crucial for monitoring financial stability and integrity.[64] At the same time, in the case of AI and limited safeguards for consumers this introduces a new trade-off between privacy (and consumer protection more generally) on the one hand and financial stability and market integrity on the other. This trade-off is represented by the green arrow in Figure 15.

62 See Claessens (2009) for a discussion on the benefits of greater market entry in the financial sector.
63 Vives and Ye (2025a) find that IT improvements affect competition, investment, and welfare depending on their impact on lender-borrower distance and monitoring costs, with varying effects based on moral hazard severity.
64 Pagano and Jappelli (1993).

Principles for AI regulation

Given the diverse range of stakeholders and the new trade-offs, policy remedies to address the detrimental effects of AI are not straightforward. The path towards effective AI regulation is difficult because the AI supply chain contains many different markets that fall under the ambit of different regulatory authorities that often have competing goals, as shown in the policy triangle above.[65]

Balancing the benefits and risks of AI requires proactive, comprehensive regulation that anticipates future issues and incorporates technological, societal, and ethical considerations. Not all AI risks need regulatory intervention; regulation should target those risks impacting specific policy objectives, while market mechanisms can manage others. Given AI's complexity and unforeseen risks, establishing regulatory principles is crucial. Both national and international standard-setters have established general principles for regulating and managing AI systems throughout their development and deployment.[66] Common principles include societal wellbeing, transparency, accountability, fairness, privacy protection, safety, human oversight, and robustness. However, even if domestic policy consensus is achieved on these principles, international cooperation can be more elusive because jurisdictions differ in their legal frameworks and regulatory approaches towards AI, making the application of simple principles difficult to attain.

Regulatory models for AI

Three main models of AI regulation have been identified in the literature: the United States' market-driven approach emphasises innovation and self-regulation; China's state-driven model uses technology for political aims and industry growth; and the European Union's rights-driven model focuses on protecting individual and societal rights. While distinct, these models are slowly converging towards similar principles. In the United States, AI regulation has evolved to executive actions such as the 2023 Executive Order addressing AI harms, but lacks significant legislation. China's regulations emphasise socialist values and are moving towards an AI Law. The European Union's AI Act (2024) introduces a risk-based framework categorising AI systems, prohibiting those posing unacceptable risks, and setting stringent requirements for high-risk applications.

65 Take, for example, the competing goals of maintaining user privacy and having data portability and interoperability to ensure a level playing field for smaller firms.

66 For example, the European Union developed the Assessment List for Trustworthy Artificial Intelligence (ALTAI) (Ala-Pietilä et al., 2020; European Union 2024). In the United States, the National Institute of Standards and Technology (NIST) outlined characteristics of trustworthiness in its AI Risk Management Framework (NIST, 2023). China has also defined responsible AI principles (China Technology Ministry, 2019). Additionally, the International Organization for Standardization (ISO) provides guidance on risk management for AI systems through standards like ISO/IEC 23894:2023. These frameworks form the cornerstone of many AI regulatory initiatives. See Aldasoro et al. (2024d) for a more complete discussion.

Policymakers are operationalising AI regulatory principles across the value chain – design, deployment, and diffusion – through frameworks like the EU AI Act, NIST's AI Risk Management Framework, and China's Generative AI provisions. These efforts focus on governance, visibility, risk evaluation, and management to ensure systemic resilience of GenAI and AI agents in finance.

International cooperation
As data and technology have no borders, global cooperation on AI regulation is essential. Common regulatory standards are needed in particular on AI governance rules and risk assessment methodologies. Standardising AI governance rules internationally is crucial to ensure consistent ethical and safety standards, prevent regulatory arbitrage, and foster global cooperation. Uniform guidelines can enhance trust, facilitate cross-border AI applications, and address global challenges like privacy, security, and equitable access effectively. There is also a need for standardised methodologies for risk assessment of AI models that consider the unique attributes of AI, such as adaptability and learning over time. These methodologies should consider the potential for AI systems to develop unforeseen behaviours or outcomes, necessitating continuous oversight and the ability to adjust regulatory measures as the technology matures and integrates more deeply into societal infrastructures.[67]

2.4 CONCLUSIONS

AI is significantly transforming the financial sector by enhancing information processing, risk management, and customer service. AI offers substantial opportunities for efficiency and innovation, particularly in areas such as credit risk analysis, robo-advising, regulatory compliance, and customer support. However, it also introduces challenges, including model opacity, data dependency, biases, cybersecurity threats, and systemic stability concerns.

The widespread adoption of AI in finance can amplify risks such as market concentration, model herding, and uniformity, which may lead to financial instability. The potential for AI to exacerbate consumer privacy issues and legal risks further complicates its integration into the financial system. Additionally, the dominance of big tech firms in the AI supply chain poses significant competitive and regulatory challenges to be monitored.

67 Global collaboration on AI focuses on ensuring safety and transferring knowledge and best practices to ensure that all regions of the world can benefit from AI advancements responsibly. Initiatives like the G7 Hiroshima Process (signed in December 2023) and the Transatlantic Trade and Technology Council (last meeting in April 2024) underscore the importance of international collaboration in establishing standards for the safe and ethical use of AI. More recently, at the February 2025 AI Action Summit in Paris, around 60 countries signed a declaration promoting inclusive and sustainable AI, emphasising global cooperation, safety, and ethical development. Co-hosted by France and India, the agreement aimed to ensure AI benefits all nations, particularly the Global South. The summit also saw the launch of "Current AI," a $400 million initiative supporting public interest AI projects. The declaration reflects growing international efforts to balance innovation with responsible AI governance.

Effective regulation and governance are essential to harness the benefits of AI while mitigating associated risks. Policymakers must balance innovation with risk management, ensuring transparency, fairness, and ethical standards. International cooperation is crucial to harmonise AI governance and prevent regulatory arbitrage. Standardising AI governance rules and risk assessment methodologies can enhance global collaboration and address challenges like privacy, security, and equitable access.

The interconnectedness between AI advancements and the broader economy creates potential spillover effects between the real economy and the financial system. As AI permeates business operations and decision-making processes, its implications for employment, productivity, and income distribution require careful consideration. Policy responses must prepare for diverse scenarios – from productivity gains to significant labour market disruptions – to ensure inclusive economic growth and stability.

Looking ahead, the impact of AI on finance will drastically depend on the evolution of the technology. Different scenarios are possible. In the short to medium term, the impact of AI will be more limited if LLM-based copilots augment, rather than replace, human skills and workers in the financial sector. The effects will be larger if AI 'agents' become increasingly capable and independent, ultimately replacing many human functions. In the long term, the achievement of AGI or superintelligence could further revolutionise the functioning of the financial sector and the broader economy. Despite this uncertainty on the future scenarios, it is imperative that regulation considers the desired technological advancements, the skills and tasks to automate, and ensures these technologies respect fundamental rights for wider social benefit.

CHAPTER 3

AI's impact on finance: Reshaping information and its consequences

Artificial intelligence and data abundance are transforming the way information is produced and used in financial markets. This evolution can have far-reaching consequences. Indeed, as Stiglitz (1994, p. 23) writes, *"financial markets [...] can be thought of as the 'brain' of the entire economic system, the central locus of decision making: if they fail, not only will the sector's profits be lower than they would otherwise have been, but the performance of the entire economic system may be impaired."*

By this analogy, if data abundance and AI enhance the quality of information flowing into financial markets and the decisions made based on this information, the economic system will benefit from improved insights from its 'brain'. This should lead to more efficient capital allocation and foster economic growth. Conversely, if these technologies degrade the quality of financial information, this would warrant concern and potentially justify policy intervention.

Against this backdrop, the goal of this chapter is to explore how data abundance and AI impact the production of financial information by the financial industry, highlighting the potential benefits of this evolution and its potential risks. The chapter focuses on the securities industry, that is, intermediaries managing investors' savings (the 'buy side', mutual funds, hedge funds, etc.) and those helping investors to rebalance their portfolios (the 'sell side', market makers, brokers, exchanges, etc.). AI and data abundance also impact other important actors of the financial system, in particular banks and corporations, as discussed in Chapters 2 and 4 of this report, respectively.

A key role of the financial industry in general – and the securities industry in particular – is to produce and trade financial contracts that consumers of financial services use to share risks, transfer resources, or provide incentives.[68] To perform this role, financial intermediaries (banks, market makers, underwriters, fund managers, etc.) must produce information (about future cash-flows, counterparty risk, cost of capital, etc.), and many occupations in the securities industry are concerned with collecting, cleaning, and processing information.

68 Phillipon (2015).

AI reshapes how the financial industry produces and uses information because it enables industry participants to obtain more accurate predictions (of future cash-flows for investment projects, default risk for loans, or returns for securities) at lower cost and to automate the search for information and decision making based on information. As explained in detail in Section 3.1, this possibility is due to a combination of three factors: (i) a greater volume and diversity of data, enabling participants to use more predictors for forecasting; (ii) more powerful algorithms (machine learning and self-learning algorithms) to extract information from data and use this information for decision making; and (iii) a decline in the cost of using such algorithms due to lower costs of compute.

Thus, AI is a technological shock: it reduces the cost of producing financial information and substitutes human judgement by artificial intelligence for information processing and decision making. Not surprisingly, the securities industry has been quick to take advantage of this evolution, as shown by the rise of algorithmic trading (Section 3.2.1), quants funds (Section 3.2.2), and robo-advisors (Section 3.2.3).[69]

The adoption of AI in the securities industry holds significant promise. It should lead to efficiency gains through lower costs and potentially more effective decision making. If these gains are passed on to consumers of financial services (households, firms, and governments), they will be better off. In Section 3.3, we highlight four obstacles that could hinder the realisation and transmission of these gains to consumers of financial services:

- The risk that intermediaries use advances in information technologies to produce information with relatively low social value because there is no mechanism rewarding intermediaries for allocating their capacity for information production to its most valuable use for society (Section 3.3.1).

- The risk that advances in information technologies increase informational asymmetries and therefore adverse selection costs for participants unable to keep pace with the technology (Section 3.3.2). Ultimately, these costs add to intermediation costs (someone must bear them) and could therefore reduce the efficiency gains associated with the adoption of AI tools by the securities industry. Moreover, to grab informational rents, intermediaries could engage in socially wasteful investment races in computing power and human capital.

- The risk that reinforcement learning algorithms increase the market power of some intermediaries, in particular algorithmic market makers (Section 3.3.3). If so, deadweight costs will reduce welfare gains of the adoption of AI in the financial industry.

- The risk of more fragile securities markets due to the 'black box' nature of machine learning algorithms (Section 3.3.4).

69 Figure 3 in Acemoglu et al. (2022) shows that occupations in the financial industry are among the most exposed to AI and that the share of vacancies related to AI in finance has been rising rapidly in recent years.

It is worth stressing that this chapter does not describe how techniques from AI can be used to make progress on frontier questions in financial economics.[70] Instead, it focuses on how AI affects the production and use of financial information and the resulting effects on the efficiency of securities markets measured in various ways (price informativeness, liquidity, risk sharing, stability, etc.).

3.1 WHY IS FINANCIAL INFORMATION BEING RESHAPED BY AI?

In this section, we present the three drivers of the quick adoption of AI tools by the financial industry: (i) data abundance, (ii) powerful techniques to transform data into predictions and decisions, and (iii) a drop in the cost of using these techniques due lower compute costs. The combination of these factors enables participants in the securities industry to produce and use information at lower costs.

3.1.1 Data abundance

Traditionally, information producers in securities markets (e.g., securities analysts) have relied on data disclosed by firms (e.g., accounting data) to forecast asset payoffs and make financial decisions (e.g., whether or not to provide capital).

These data remain relevant, and AI tools can be used to make predictions using them (for example, using natural language processing algorithms to obtain buy or sell signals from new regulatory filings). However, two new sources of data have emerged as important sources of information for the securities industry: (i) alternative data and (ii) market data. As explained below, these data have considerably increased the volume and variety of new data relevant for financial decisions and asset valuations, and the velocity at which these data become available (the three 'Vs' of the big data revolution).

Alternative data

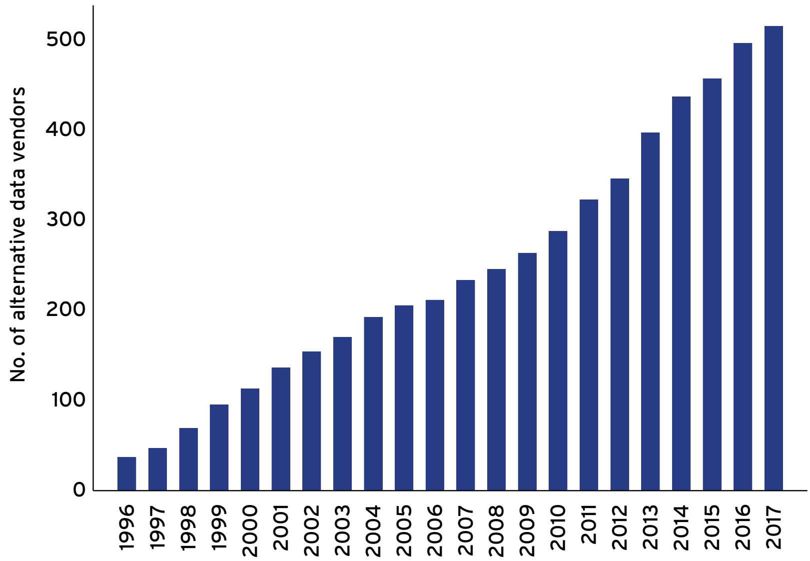
The last twenty years have seen a dramatic rise in the availability of 'alternative datasets'. In contrast to traditional sources of information about corporations (e.g., regulatory filings and financial statements), these data are not necessarily generated and disclosed by firms. Rather they stem from the digital footprints left by various activities (shopping, trading, web browsing, travelling, etc.).

Examples of alternative datasets include geolocation data (e.g., GPS signals from mobile devices), social media data (e.g., text in posts or images on social platforms), consumers' product reviews, credit card data, satellite imagery (e.g., images of retailers' parking lots, production sites, or shipping activity), sensor data (e.g., recordings of industrial activity or carbon emissions), and web traffic and clickstream data.

70 Many research papers in economics and finance use AI tools. See, for instance, Kelly and Xiu (2023) for a survey of applications in asset pricing and portfolio management.

As shown in Figure 16, alternative data are collected by an increasing number of vendors (including firms like Eagle Alpha, Yipitdata, Kayros, RS Metrics, Datarade, etc.)[71] and sold to buyers such as fintech companies, firms, fund managers, banks, or security analysts.[72] The emergence of this market for alternative data is in itself a significant change in the market for financial information.[73]

FIGURE 16 **ESTIMATED NUMBER OF ALTERNATIVE DATA PROVIDERS**



Source: Dessaint et al. (2024).

As a result of this evolution, the amount and diversity of data available to financial market participants for predicting asset returns and cash flows have increased considerably over time, and a large number of academic studies demonstrate that alternative datasets now provide valuable insights into stock returns, corporate earnings, and credit risk.[74]

As alternative data sources are not controlled by firms, they change how corporations should communicate with investors. As a result, the rise of alternative data might reshape corporate disclosures and regulation thereof. This implication is discussed in more detail in Chapter 4.

71 The number of alternative data providers in Figure 16 is obtained from JP Morgan (2019), which provides a list of data providers of interest for the financial industry. The actual number of alternative data providers is larger. For instance, as of 2024, Datarade lists about 2,500 alternative data vendors and more than 22,000 datasets.

72 Grennan and Michaely (2020) find that about 50% of fintechs specialising in market intelligence sell financial signals. Chi et al. (2023) show that sell-side equity analysts' reports increasingly mention the use of alternative data as a source of information.

73 This evolution raises interesting issues about the valuation and pricing of financial data (Farboodi et al., 2024).

74 Examples of such studies include Jame et al. (2016), Huang (2018), Bartov et al. (2018), Green et al. (2019), Benamar et al. (2021), Grennan and Michaely (2021), Jin (2023), Katona et al. (2025), and Dessaint et al. (2024), among many others. Green et al. (2019), for instance, study the informativeness of employer ratings on Glassdoor (an employer review and recruiting platform where employees can share information about their companies). They find that value-weighted portfolios of firms with the largest quarterly improvements in employer ratings outperform those with declines in employer ratings. Katona et al. (2025) show that daily satellite imagery data on parking-lot utilisation at US retailers can be used to predict these retailers' next-quarter earnings and stock returns.

Market data

The trading process in securities markets also generates a very large amount of data with very frequent updates, as shown in Figure 17. Indeed, every order (e.g., the submission of a new limit order, its revision, or its cancellation) for a security on an electronic trading platform generates a new datapoint. In turn, this datapoint is a potential source of information for decisions made by market participants.

FIGURE 17 **ONE MILLISECOND OF ORDER ACTIVITY FOR ONE US STOCK**

Time	Millis	Exchange	Action	Type	Size	Remaining	Price	Duration	Order-ID
10:01:49	77.871	EDGE-A	QUOTE	BID	700		12.18		ORDER-123455578
10:01:49	77.877	NASDAQ	CANCEL	ASK	3800	0	12.19	0.006824	ORDER-223458675
10:01:49	77.888	EDGE-X	CANCEL	BID	100	0	12.18	13.018484	ORDER-890767564
10:01:49	77.888	EDGE-X	QUOTE	BID	100		12.13		ORDER-224564478
10:01:49	77.983	PHLX	CANCEL	BID	4000	0	12.16	37.218454	ORDER-990871856
10:01:49	77.993	NYSE	QUOTE	ASK	300	3200	12.19	0.000155	ORDER-NA
10:01:49	78.029	NASDAQ	CANCEL	BID	4300	0	12.17	13.017681	ORDER-231565236
10:01:49	78.032	BATS-Z	QUOTE	BID	100		12.17		ORDER-85784
10:01:49	78.041	PHLX	CANCEL	BID	8400	0	12.17	16.369599	ORDER-667728645
10:01:49	78.078	NASDAQ	QUOTE	ASK	3800		12.19		ORDER-NA
10:01:49	78.086	BATS-Z	CANCEL	BID	7400	0	12.17	13.017709	ORDER-465988641
10:01:49	78.103	PHLX	CANCEL	BID	4000	0	12.16	16.369565	ORDER-164952269
10:01:49	78.146	EDGE-A	CANCEL	BID	100	0	12.17	37.527492	ORDER-852962674
10:01:49	78.169	EDGE-X	CANCEL	ASK	400	0	12.19	0.000332	ORDER-776645974
10:01:49	78.169	EDGE-X	CANCEL	ASK	400	0	12.19	0.000317	ORDER-102696697
10:01:49	78.213	PHLX	CANCEL	BID	6100	0	12.17	13.017743	ORDER-588461236
10:01:49	78.219	EDGE-X	CANCEL	BID	2800	0	12.17	13.017629	ORDER-554185558
10:01:49	78.341	LAVA	CANCEL	BID	100	0	12.18	3.147924	ORDER-365786456
10:01:49	78.662	EDGE-A	TRADE	SELL	700	0	12.18	0.000791	ORDER-123455578

MATCH

Note: The first and the last orders are matched together and correspond to a trade. Other orders are cancellations or the submission of new limit orders (quotes).

Source: U.S. Securities and Exchange Commission website (https://www.sec.gov/securities-topics/market-structure-analytics/midas-market-information-data-analytics-system).

For instance, institutional investors and their brokers need real-time information on quotes across various markets as inputs for the execution algorithms that they use to minimise trading costs and achieve 'best execution'.[75]

Access to real-time market data is also critical for the algorithms used by proprietary trading firms specialised in market making and arbitrage (see Section 3.2.1). In particular, high-frequency trading firms demand extremely fast access to data on quotes posted in different limit order books for a security (Figure 17 reports quotes from seven different platforms, for example) to be able to swiftly cancel and update their quotes when new information becomes available and to make profits by exploiting very short-lived arbitrage opportunities, due, for instance, to the slow adjustment of prices in different platforms to new information (see Section 3.2.1). Consequently, there is significant demand for market data among participants in the financial industry and exchanges derive an increasing share of their revenue from selling data generated on their trading platforms.[76]

75 See, for instance, Greenwich Associates (2017), which reports results from a survey of 69 asset managers and hedge funds regarding the value of various types of data, including market data. Seventy percent of the respondents indicate that real-time market data and historical market data is the type of data providing the largest competitive advantage.

76 See Chapter 4 of Duffie et al. (2022) for a discussion of the market for market data.

Exchanges sell market data in several ways. First, they offer real-time data – either directly to investors or through data vendors such as Bloomberg or Refinitiv – on trades and quotes on their trading platforms. These offerings vary in the level of data granularity provided.[77] Second, exchanges rent rack space in their data centres, a practice known as 'co-location'. Co-location allows high-frequency traders to run their algorithms in close proximity to the exchange's matching engines, enabling them to receive market data and execute actions based on these data faster than other market participants.

In recent years, industry participants have complained that the fees trading platforms charge for market data are excessively high.[78] Equity markets in Europe and the United States are highly fragmented in the sense that the same stock trades on multiple trading platforms (again see Figure 17 for an example). In such environments, brokers and investors must subscribe to data feeds from each exchange to make optimal routing decisions – such as directing market orders to the trading platforms offering the best quotes for a stock. In some cases, like in the United States, the best quotes are consolidated and disseminated to market participants at a low cost but with a slight delay, which makes the consolidated tape an imperfect substitute for real-time data feeds from exchanges. As a result, data feeds from various exchanges are not perfect substitutes, which grants exchanges a degree of market power.

3.1.2 From data to predictions and decisions: Machine learning algorithms

Forecasting

Data are not informative by themselves; one needs a technology to transform them into actionable information. For this reason, alternative data and market data would have been difficult to leverage for information production without advances in information processing, in particular the development of machine learning.[79]

ML algorithms (neural networks, decision trees and random forests, ridge regressions and lasso, natural language processing methods, etc.) rely on large amounts of data (the input of these algorithms) to generate predictions (the output). In supervised learning, these algorithms are designed to 'learn' the relationship between a set of variables ('features'), $x_i = (x_{i1}, x_{i2}, \ldots x_{iN})$, and an outcome variable ('target' or 'label'), y_i (where i refers to one data point).[80]

77 For instance, exchanges charge different fees for access to last trade and top of the book data (the best quotes and the number of shares offered at these prices) only and the entire limit order book. They also charge different fees depending on whether the data is for internal use or for external redistribution. See https://www.nyse.com/publicdocs/nyse/data/ for a list of data sold by the NYSE and its pricing schedule.

78 "European investors complain about soaring costs of data", *Financial Times*, 4 April 2019.

79 Kelly and Xiu (2023) define machine learning as "*(i) a diverse collection of high-dimensional models for statistical prediction, combined with (ii) "regularization" methods for model selection and mitigation of overfit, and (iii) efficient algorithms for searching among a vast number of potential model specifications*."

80 The learning is 'supervised' in the sense that during the training phase, the algorithm is told after selecting an action (e.g., after making a forecast) what the correct action was (e.g., the actual realisation of the forecasted variable).

For instance, the features might be the numerical colour values of the pixels in the image of an animal and the target might be the animal type (a dog or a cat). The algorithm is first 'trained' on a large number of images (realisations of x and y values) to learn the relationship ('model') between the features and the target. The resulting model can then be used to predict animal types when new images are presented to the computer ('out of sample'). The model is selected to minimise a measure of the average prediction error (e.g., the mean squared error). There is learning in the sense that, in sample, the true target is known and one can interpret the algorithm as (i) making a prediction y^p based on x; (ii) observing the true realisation, y, of the target; and (iii) adjusting the model based on the error between its prediction and the actual target.

More generally, consider the problem of predicting Y (e.g., the return of a stock) with a vector of variables X (e.g., various stock characteristics) to minimise the expected square forecasting error. The solution to the problem is the function $f(X)$ that minimises $\mathrm{E}\big((Y - f(X))^2\big)$, where E(.) is the expectation operator. It well known that $f(X) = \mathrm{E}(Y|X)$. However, without knowing the distributions of Y and X, one cannot compute $\mathrm{E}(Y|X)$. Instead, one can use a set of realisations of Y and X (training data) to find a function $\hat{f}(X)$ that is a good approximation of $f(X)$ (e.g., in the sense that it minimises the empirical mean square forecasting error). ML algorithms are methods to find $\hat{f}(X)$.

These algorithms have several important characteristics. First, they are high-dimensional because the number of features can be very large. Second, they differ in the type of specification allowed for $\hat{f}(X)$.[81] More complex specifications enable to better fit the data in-sample but create the risk of poor performance out-of-sample due to overfitting. ML algorithms are designed to limit this risk via various techniques referred to as 'regularisation methods'. Third, ML algorithms can handle data in unstructured format such as text, images, or voices.

ML algorithms are important in finance because many financial problems involve prediction problems. For instance, active fund managers search for signals that can predict securities returns and assess whether trading strategies (portfolios) based on these signals generate significant average abnormal returns ('alphas'). Additionally, their brokers need to forecast future changes in liquidity and short-term price movements to execute fund managers' orders at low costs (see Section 3.2.1). Security analysts make forecasts of firms' earnings at various horizons, formulating investment recommendations based on these predictions. Firms' valuation and capital budgeting require forecasting cash flows at various horizons. Banks and rating agencies develop models to predict borrowers' credit risk. Venture capitalists screen projects based on their likelihood of success (e.g., successful exit via an IPO).

81 For instance, neural networks allow for highly non-linear specification with a very large number of parameters to estimate, while ridge regressions consider a linear specification with many explanatory variables and an L2 penalty to prevent overfitting.

Consequently, a fast-growing literature is using ML to study prediction problems in finance, such as the prediction of stock returns, corporate earnings, credit risk, liquidity, or failure/success outcomes of investment in start-ups.[82] A key takeaway from these studies is that AI-powered forecasts are more accurate than forecasts produced by simpler models or human predictions.[83,84] This finding reflects the ability of machines to process significantly larger information sets than humans and the fact that they are exempt from human cognitive biases.[85] Moreover, in contrast to humans, machines are not affected by conflicts of interest and agency issues (machines do not need to be incentivised to produce accurate signals).[86]

Another important insight from this literature is that combining human-based forecasts with ML forecasts can result in better predictions than relying on ML forecasts alone. This observation suggests that humans possess unique predictive abilities, possibly due to a form of analysis that cannot be replicated by machines (e.g., they rely on models of the world) or because humans have access to information that is not available to ML algorithms.

For instance, Cao et al. (2024) compare ML forecasts with equity analysts' forecasts of one-year stock returns. They allow the ML forecasts to leverage a large set of predictors, including firm-level accounting variables, industry-level variables, and macroeconomic variables, alongside analysts' forecasts. The authors find that ML forecasts are more accurate than analysts' forecasts: out of 922,157 forecasts made from 2001 to 2018, the

82 For the prediction of stock returns, see, for instance, Chinco et al. (2019), Gu et al. (2020), Murray et al. (2024), Brogaard and Zareei (2023), Chen et al. (2023a), Chen et al. (2023b), and Nagel (2021) for a textbook presentation. For the prediction of corporate earnings, see for instance, Cao and You (2024), Chen et al. (2022a), da Silva and Thesmar (2024), Dong (2024) or Van Binsbergen et al. (2023). For credit scoring and the prediction of credit risk, see for instance Lessman et al. (2015), Jansen et al. (2024), Fuster et al. (2019, 2022), Hurlin et al. (2024), or Hué et al. (2023). For predicting future liquidity, see Easley et al. (2021) and for measuring informed trading, see Bogousslavsky et al. (2024). For the prediction of failure/outcome of early investment in start-ups, see Retterath (2020) or Lyonnet and Stern (2024).

83 For instance, Cao and You (2024) compare the predictive performance of ML algorithms (decision trees and artificial neural networks) for corporate earnings with the performance of standard linear forecasting models (e.g., a random walk model or an AR(1) model), for a sample of US firms from 1975 to 2019. They train the ML algorithms to predict yearly earnings one, two, and three years ahead using 60 different features obtained from firms' financial statements (e.g., balance sheet or income statements). They find that the ML models' forecasting accuracy out-of-sample dominates that of simpler models and that ML models forecasting accuracy is higher than that of analysts' consensus forecasts. Another example is Gu et al. (2020), who compare the out-of-sample predictive performance of various ML algorithms for monthly excess stock returns from 1957 to 2016, covering approximately 30,000 stocks. They construct more than 900 predictors (signals) of monthly excess stock returns based on stock characteristics (e.g., operating profitability, market capitalisation, etc.), macroeconomic variables, and industry dummies. They find that more complex ML algorithms have significantly higher out-of-sample relative predictive power than simpler ones. They also demonstrate that portfolios constructed using forecasts from more complex algorithms yield realized returns much closer to predicted returns than those based on simpler algorithms, as well as higher Sharpe ratios.

84 This finding of course is not specific to financial forecasting. For instance, the ImageNet challenge (https://www.image-net.org/challenges/LSVRC/) was a yearly contest to predict the name of an object in an image with machine learning algorithms. The contest began in 2010 and stopped in 2017. The error rate of the best algorithms fell below that of humans for the first time in 2015.

85 For instance, Da Silva and Thesmar (2024) show that the differences between ML forecasts and human forecasts are due to the fact that analysts add noise to rational forecasts, maybe due to cognitive limitations. Coleman et al. (2022) find that robo-analysts' recommendations are less biased than those of human analysts at investment banks, while van Binsbergen et al. (2024) find that ML algorithms' forecasts are less positively biased than human predictions. Jansen et al. (2024) find that algorithmic underwriting (for retail loans) results in lower default rates, in particular for loans in which agency issues for human underwriters are more likely to arise.

86 However, algorithms are written by ML engineers who could face conflicts of interest.

ML-generated forecasts outperform analysts' forecasts 54.5% of the time. However, they also find that incorporating analysts' forecasts into the information set used by ML algorithms improves accuracy further. The resulting 'centaur analyst' model outperforms ML forecasts 54.8% of the time.

As discussed in Chapter 2 of this report, generative AI has recently emerged as a powerful transformative force for the financial sector. GenAI algorithms respond to prompts by generating content, including text, images, and videos. Several recent studies show that these models can also be used to predict corporate earnings and stock returns.[87]

Decision making

As discussed in the previous section, machine learning algorithms often produce better predictions than humans. These predictions can then be used as input for making financial decisions. For instance, fund managers can first train machine algorithms to predict future returns and then use these predictions to make investment decisions. Similarly, loan officers can rely on credit risk assessment by machine learning algorithms to decide whether or not to grant a loan to a borrower. In these cases, there is a clear division of labour: machines assess the likelihood of various outcomes and humans use this assessment to make decisions.

However, this division is becoming blurred for two reasons. First, when the mapping between predictions and optimal decisions is clear, humans can delegate the decision to the machine to save on labour costs. For instance, for loan applications, acceptance/rejection may just depend on credit risk being above or below a certain threshold. Such a decision can also be easily coded in an algorithm.

Second, ML algorithms can also be used to train systems to make autonomous decisions in complex dynamic environments. In particular, reinforcement learning (RL) algorithms are designed to learn behaviours to achieve specific goals with minimal prior knowledge of the environment in which they operate.[88] Such self-learning algorithms have been successfully used, for instance, to play chess and video-games, achieving victories against the best human players.

87 For instance, Kim et al. (2024) show that ChatGPT can process financial statements to forecast the direction of changes in yearly corporate earnings with a level of accuracy superior to security analysts. Interestingly, like Cao et al. (2024), they find evidence of complementarities in ChatGPT generated forecasts and security analysts' forecasts in the sense that the latter are useful to predict the change in the direction of corporate earnings even after controlling for Chat GPT forecasts. Lopez-Lira and Tang (2024) use ChatGPT to generate buy or sell recommendations of stocks following news headlines over the 2021-2023 period. They then build portfolios with long positions in stocks receiving a high recommendation and short positions in stocks receiving a low recommendation and find that this portfolio performs well.

88 In this case, the goal of the algorithm is to learn how to make a decision to achieve a specific objective (e.g., maximise the average terminal utility of the liquidation value of a portfolio or win a game) with minimal starting knowledge about the environment. See Charpentier et al. (2020) for a presentation of reinforcement learning and its applications in economics.

Humans in this case are still important to select and design self-learning algorithms. In particular, the ability of self-learning algorithms to achieve goals (e.g., setting the online price of a product to maximise average profits) depends on the feedback provided to the algorithm (e.g., the number of sales or the margin per sale) and the type of state variables on which the algorithm can condition its decision at each point in time (Last price? Entire price history?). These are choice variables for the designers of the algorithms.

Consider a dynamic portfolio problem with one risky asset and one riskless asset. Tools from dynamic optimisation can be used to solve analytically or numerically for the optimal portfolio allocation at each rebalancing date. However, this approach requires knowledge of the relevant model parameters and assumptions about the underlying distribution of returns for the risky asset. In this context, a self-learning algorithm can learn how to allocate capital to the risky asset without knowledge of these distributions, via experimentation.[89]

The algorithm initially experiments with various possible allocations in the risky asset across different states (the current value of the portfolio, past returns of the risky asset, etc.) to evaluate the average utility associated with each allocation in each state. Over time, the algorithm reduces the frequency of experimentation and increasingly selects, in a given state, the allocation that generates the highest average utility based on its assessment (the 'greedy action'). Intuitively, the reason for this is that, as time progresses, the algorithm is expected to have 'learned' the optimal allocation for each state.

There are many other applications in the context of securities trading. For instance, self-learning algorithms can be used for designing optimal hedging strategies or optimal execution strategies for large orders, or for market making.[90]

3.1.3 Lower information acquisition costs

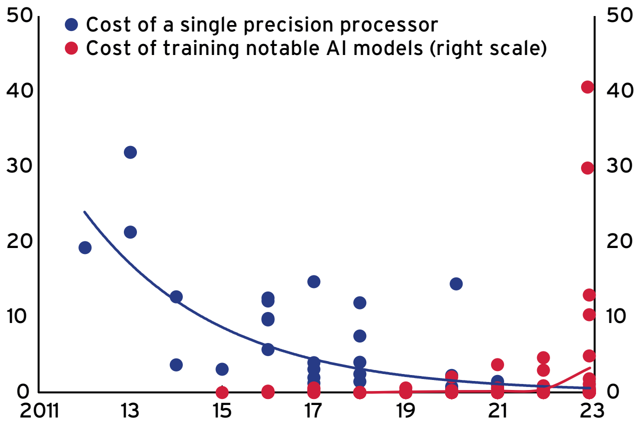
Machine learning algorithms enable agents to exploit vast amount data to form more accurate predictions and automate decisions. Training and using these algorithms, however, requires a very large number of computations. More powerful chips can process more calculations per unit of time and therefore reduce ML algorithms' training time, holding their complexity constant.[91] Innovation in chips design (like GPUs) therefore has be a third key driver of the rise of AI in the financial and other industries. In turn, this rise has generated very strong demand for computing chips (see Chapter 2).

89 See Barberis and Jin (2023) for an example.
90 See Cao et al. (2020) for hedging and Hafsi and Vittori (2024) for optimal execution.
91 A chip's computing power is expressed in floating-point operations per second (FLOPS). A floating-point operation refers to a single computation performed on floating-point numbers (numbers with decimal points, expressed in scientific notation). NVIDIA A100 Tensor Core GPU can perform 312.1012 FLOPS.

These innovations have also significantly improved the performance of chips per dollar, that is, the cost of computing power has declined over time (Figure 18).[92] Holding investment in computing power constant, participants in financial markets can now obtain more accurate signals by using more complex ML algorithms.[93] In other words, the decline in computing costs combined with progress in AI and data abundance has reduced the cost of producing financial information.

FIGURE 18 **COST OF A SINGLE PRECISION PROCESSOR AND THE COST OF TRAINING NOTABLE AI MODELS**

Inflation-adjusted 2023 US dollars per extra floating point operation; millions of inflation-adjusted 2023 US dollars (right scale)



Source: IMF (2024).

This decline in the cost of information is potentially an important benefit of AI adoption by the financial industry for consumers of financial services. The role of the financial industry is to "*produce, trade and settle financial contracts*".[94] These contracts are used to share risks or transfer resources between agents (e.g., from households to firms). To provide these services, financial intermediaries (banks, brokers, asset managers, etc.) charge fees to cover their costs, including information production costs. For instance, banks or private equity funds must produce information to screen and monitor investment projects to which households' savings are allocated. Similarly, brokers and dealers must produce information to guarantee that final investors trade financial contracts at fair prices in secondary markets, or asset managers must produce information to build efficient portfolios and find ways to execute their orders at the lowest cost.

92 See also Nordhaus (2021).
93 The total amount invested to develop AI tools is increasing however as models become more complex and innovations races (see Section 3.3.2 for a discussion).
94 Phillipon (2015).

Lower costs of producing information should, in principle, reduce intermediation costs paid by consumers of financial services and make them better off. This is not a foregone conclusion, however. Phillipon (2015) finds that the cost of financial intermediation in the United States was stable at 2% (i.e., 2 cents per dollar of intermediated financial assets) between 1886 and 2012. This is puzzling since, over this period, there were considerable improvements in information technologies that the financial industry has been quick to take advantage of.[95] Thus, we should be wary that reductions in information production costs are not fully passed on to consumers of financial services.

One possible reason is that financial intermediaries have market power. If AI tools increase – or at least do not reduce – their market power, consumers of financial services will not fully benefit from the drop in information production costs it brings (see Section 3.3.3).

Another reason is that improvements in information technologies can raise informational asymmetries between those who use the latest technology and those who don't (see Section 3.3.2). In turn, informational asymmetries raise adverse selection costs (losses incurred by less informed parties in their dealings with more informed parties). In this case, consumers of financial services will benefit less from reductions in intermediation costs due to advances in information technology.[96] This possibility suggests that investment in information technologies (e.g., data and data processing tools) can be excessive relative to the socially optimal level. Indeed, when investing in these technologies, their users trade off their private benefits and costs, without internalising adverse selection costs for other market participants. Moreover, intermediaries might choose to invest in information technologies just to avoid paying adverse selection costs, even though their investment does not help per se to make intermediation more efficient.[97] This is similar to the logic of an arms race where the race is socially useless since society would be better off if it could be avoided in the first place. In either case, investment in information technologies is excessive from the viewpoint of society (i.e., a larger amount of financial intermediation could be produced at lower cost).

A final reason is that financial intermediaries produce various types of information, and not all of them hold the same value for society. For instance, leveraging new information technologies to forecast the value created by startups working on innovative projects allows intermediaries to allocate capital more efficiently to projects that generate the greatest societal value. In contrast, the social value of using new information technologies

95 For instance, stock tickers were installed on the floor of the NY Stock Exchange in 1867 and telephones in 1878 ("1889: The telegraph ramps up trading speed", *Wall Street Journal*, 10 July 2014).

96 For instance, suppose that advances in information technologies reduce the true cost of intermediation from 2 cents to 1.5 cents per dollar intermediated while increasing adverse selection costs in securities trading by 0.3 cents. Given that these costs will ultimately be passed to consumers of financial services, the net drop in costs for these consumers will only be 0.2 cents and not 0.5 cents. The total increase in demand for financial services and associated gains (e.g., risk sharing) will therefore be smaller.

97 See Glode et al. (2012), Biais et al. (2015), or Pagnotta and Philippon (2018) for a formal analysis.

to react more quickly to news announcements – as some proprietary trading firms do – is less evident. Thus, even though information costs go down, the social benefits of more accurate signals might depend on which type of information is produced (see Section 3.3.1).

3.2 IMPLICATIONS FOR THE SECURITIES INDUSTRY

In this section, we discuss how the technological changes described in the previous section affect (i) securities trading, (ii) asset management, and (iii) wealth management.

3.2.1 AI-powered trading

The use of pricing algorithms in consumer markets (e.g., travel or accommodation) is a relatively recent phenomenon. In contrast, in financial markets, traders have been using algorithms for implementing various trading strategies for at least the last 20 years. The catalyst was the adoption by trading platforms of application programming interfaces (APIs) – sets of rules and tools allowing different software applications to communicate with each other. APIs enable traders' algorithms to directly interact with exchanges' operating systems, without need for human intervention.

Securities traders have developed algorithms to automate standard trading strategies. These can be classified into four broad categories:[98]

1. **Market making**. Market makers are intermediaries (dealers) who post bid and ask quotes at which investors can sell or buy securities without delay. In the past, quotes were posted and managed manually by humans. Over time, this process has been increasingly automated through algorithms. In particular, high-frequency market-making firms such as Citadel Securities, Jane Street, and Virtu use algorithms to post quotes across various markets and rapidly adjust them in response to changing market conditions or inventory levels.

2. **Arbitrage**. Arbitrageurs exploit mispricings between related securities. One example is when the same security can be bought and sold almost instantaneously in different markets for a profit. For instance, E-mini futures on the S&P 500 and SPY ETFs on the S&P 500 are functionally equivalent, as both allow investors to trade a basket of stocks corresponding to the S&P 500 index. However, when new information arrives, the quotes for these securities do not adjust simultaneously, sometimes creating brief arbitrage opportunities that last only a few milliseconds.[99] Another example involves stocks traded on multiple trading platforms, as is common in the United States and Europe. In such cases,

98 See Chapter 9 in Foucault et al. (2024).
99 See Budish et al. (2015).

the quotes for the same stock across different platforms may temporarily differ, enabling fast investors to buy a stock on one platform and resell it immediately at a higher price on another.[100] Algorithmic trading provides a way to react quickly to such opportunities, exploiting them before others do or before quotes adjust.

3. **Directional trading.** Directional traders buy or sell securities based on private information not yet reflected in prices. For instance, fund managers can combine alternative data and ML algorithms to generate buy or sell signals for a stock (see Sections 3.1 and 3.2.2). They can also use algorithms to quickly access the text of new firms' regulatory filings when they are released on the US Securities and Exchange Commission (SEC) website (EDGAR) or during Federal Open Market Committee (FOMC) announcements. These algorithms process the text using natural language processing techniques to extract signals about future earnings and generate buy or sell orders based on these signals (see Chapter 4).

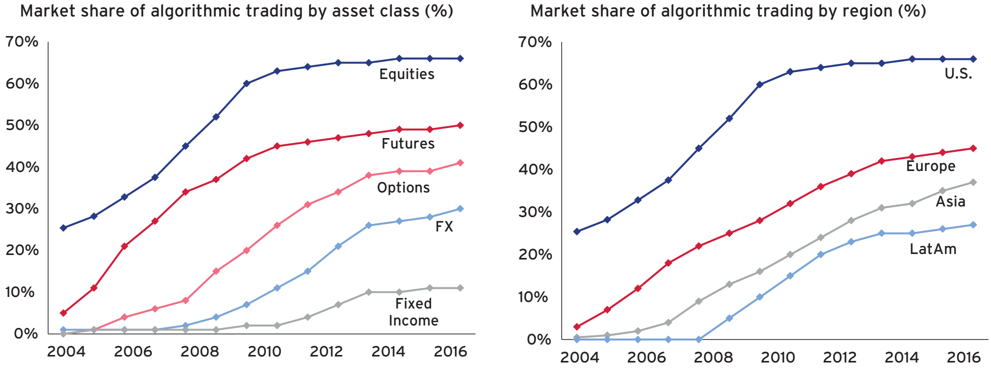
4. **Trading cost minimisation**. Brokers and institutional investors' trading desks employ various strategies to minimise trading costs. For example, a trader looking to buy one million shares of a stock may execute the order gradually to reduce its impact on prices and avoid paying an excessive markup relative to the stock's fair value. In this case, the trader must determine the execution rate (how to break up the order over time), the trading venues to which orders are routed, and the type of orders used (limit or market), all with the goal of minimising the average price impact. This presents a complex dynamic optimisation problem, often requiring frequent adjustments to the order execution strategy as market conditions evolve. Increasingly, traders rely on 'execution algorithms' to implement and refine these strategies in real time.[101]

In summary, algorithms used in financial markets serve various purposes and are used by different types of agents (e.g., brokers and funds to minimise execution costs, proprietary trading firms for market making, and active funds managers to exploit private information). Additionally, some of the algorithms must access and process information extremely quickly (high-frequency traders), while others operate at lower frequencies.

While the use of algorithms for securities trading is not new, there have been some significant developments in recent years. First, as shown in Figure 19, their use is on the rise across all asset classes, including those that were traditionally traded over the counter (such as sovereign and corporate bonds or currencies). This evolution is a consequence of the development of electronic trading platforms in these markets.[102]

100 See Wah (2016).
101 See Frazzini et al. (2018) and Beason and Wahal (2020).
102 See Chapter 4 in Duffie et al. (2022) for a description of this evolution.

FIGURE 19 **MARKET SHARE OF ALGORITHMIC TRADING BY ASSET CLASS AND REGION**

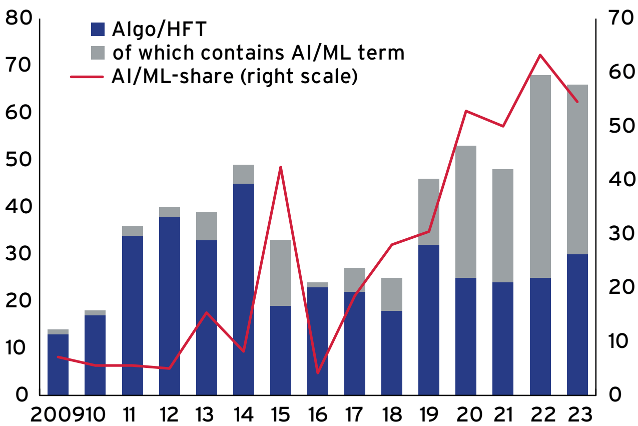


Source: Goldman Sachs (2018).

Second, trading algorithms increasingly rely on AI tools. For example, the number of patents related to algorithmic and high-frequency trading mentioning AI is on the rise (Figure 20). One reason is that, as explained before, ML algorithms are effective tools for improving the accuracy of forecasts regarding future changes in prices, liquidity, or order flow (see Section 3.1.2).[103]

FIGURE 20 **PATENTS RELATED TO HIGH-FREQUENCY/ALGORITHMIC TRADING**

Number of patents; percent of patents



Source: IMF (2024).

103 For instance, Easley et al. (2021) use a random forest algorithm to predict the direction of changes in bid-ask spreads or realised volatility in various futures markets using features based on market data. They argue that building such predictive models is valuable for execution algorithms, as these enable them to adjust their trading rate to expected changes in market conditions (e.g., increase the rate of execution when bid-ask spreads or volatility are expected to rise).

A survey of Dutch proprietary trading firms conducted by the Dutch Authority for Financial Markets (*Autoriteit Financiële Markten*, or AFM) indicates that between 80% and 100% of respondents' algorithms rely on machine learning.[104] It remains unclear whether these algorithms are solely used to generate forecasts for algorithms with hard-coded trading rules (e.g., "*If the price of stock A is expected to increase in the next 5 minutes, then buy all shares available at the best offer*") or if some are designed to independently learn trading strategies using reinforcement learning (see Section 3.1.2). However, some banks already use reinforcement learning algorithms to optimise execution of large orders.[105]

An important question is how algorithmic trading affects liquidity and price discovery in securities markets. In general, academic studies have found that algorithmic trading is associated with an improvement in price discovery (e.g., prices are closer to a random walk, which means that they better incorporate available information). In contrast, the effect of algorithmic trading on trading costs (a dimension of intermediation costs) is less clear-cut.[106]

One reason is that faster access to information, due to trading automation, can either reduce or increase market makers' exposure to adverse selection. Consider the arrival of news about a stock. Market makers are at risk of trading at stale quotes and therefore losing money if they do not quickly update their prices. This is sometimes referred to as the 'risk of being picked off' and is a form of adverse selection risk.

Faster information processing and quicker access to trading platforms have an ambiguous effect on the risk of being picked off. On the one hand, they enable traders to cancel their quotes faster, thereby mitigating their exposure to this risk. On the other hand, they enable traders to pick off other traders' stale quotes faster, thereby amplifying other traders' exposure to the risk of being picked off. While the net effect is theoretically ambiguous, some empirical studies find that the second effect dominates – that is, faster access to information and responses to this information raise adverse selection costs.[107] This does not imply that algorithmic trading necessarily results in larger bid-ask spreads (trading costs for final investors). In fact, studies generally find that algorithmic trading has reduced the total costs of trading.[108] What it does mean is that this cost reduction is smaller than it could be without adverse selection costs.

104 See AFM (2023).
105 One example is AIDEN, a reinforcement algorithm developed by Royal Bank of Canada (see https://www.rbccm.com/en/expertise/electronic-trading/ai-trading.page.)
106 See Chapter 4 in Duffie et al. (2022) and Chapter 9 in Foucault et al. (2024) for detailed discussions of empirical and theoretical findings on these issues.
107 See, for instance, Shkilko and Sokolov (2020), Aquilina et al. (2024), and Foucault et al. (2017).
108 See Hendershott et al. (2011), for instance.

One caveat is that empirical findings on the effects of algorithmic trading in financial markets have largely been gathered during a period when trading algorithms had not yet incorporated AI tools. Consequently, for the moment, little is known about the impact of ML algorithms on liquidity, price discovery, or market stability. Further research on this topic is needed.

3.2.2 AI-powered asset management

Asset managers play an important role. They help investors to diversify their securities holdings at low costs. Algorithmic trading allows them to reduce these costs even further by helping them to execute their orders with low impact on prices (Section 3.2.1). In addition, active fund managers produce information on future securities cash-flows to identify undervalued or overvalued securities and build portfolios of securities that exploit these pricing inefficiencies. Doing so, they make financial markets more informationally efficient and asset prices more informative.

Active fund managers increasingly rely on alternative data (Section 3.1.1) and ML algorithms (Section 3.1.2) to produce information. For instance, a 2017 report from JP Morgan notes that "*[i]n the search for [...] alpha, fund managers are increasingly adopting quantitative strategies. [...] a new source of competitive advantage is emerging with the availability of alternative data sources as well as the application of new quantitative techniques of machine learning to analyze these data*."[109] Moreover, AI tools can be used for ideas generation and portfolio construction.[110]

One manifestation of this evolution is the rise of quant funds (AQR, Citadel, Renaissance Technologies, Two Sigma, DE Shaw, etc.), that is, funds relying on mathematical and statistical analysis of vast amounts of data to generate their trading signals and construct their portfolios.[111] For instance, in a survey by Mercer of 150 employees (chief technology officers or members of investment teams) at asset management firms, 48% of respondents indicated their institution uses machine learning models, and 26% reported using generative AI.[112] Another survey by BarclayHedge found that more than 50% of hedge

109 See also "When Silicon Valley came to Wall Street", *Financial Times*, 27 October 2017; "At BlackRock, Machines are rising over managers to pick stocks," *The New York Times*, 28 March 2017; "DE Shaw: inside Manhattan's 'Silicon Valley' hedge fund", Financial Times, 26 March 2024; "Stock pickers turn to big data to arrest market decline," Financial Times, 11 February 2020.

110 For instance, Cong et al. (2022) use a reinforcement learning algorithm (which they call 'AlphaPortfolio') for portfolio allocation. The algorithm is trained to select allocations with large Sharpe ratios and achieve strong out-of-sample performance. See also Bryzgalova et al. (2020) and Kozak et al. (2020).

111 Abis (2022) finds that quant funds accounted for 18.6% of all US active equity funds in 2017, up from 6.1% in 2000. Quant funds rely on research teams to discover predictors of future returns, who estimate various predictive models of future returns using vast amount of data and retain for investment only those whose predictive performance is good enough. See Chapter 9 of Nayang (2013) for a description of the role of research teams in quant funds.

112 See Mercer (2024).

funds surveyed use ML to develop their strategies, particularly in portfolio construction and idea generation. Furthermore, some studies provide evidence of AI adoption by fund managers, as shown by an increase in job postings requiring AI skills or the correlation between changes in their holdings and signals generated by generative AI.[113]

This evolution raises several questions. What are the effects of AI adoption on fund managers' performance? Is the rise of quant funds and the decline of discretionary funds – those relying on human judgement and soft information – inevitable? How does this trend impact market liquidity (trading costs for investors in general) and the informativeness of securities prices about fundamentals (future cash-flows)?

Research on these questions is just emerging. In one possible scenario,[114] access to more and more diverse data and improvements in data processing tools will allow quant funds to generate signals of increasingly higher precision in the short run. Everything else equal, this will improve their average performance so that final investors will shift capital from discretionary funds to quant funds. However, as capital allocated to quant funds rises and their signals become more informative, securities prices should reflect their private information faster. This effect implies that the rise of quant funds might eventually reach a plateau.

One frequently expressed concern is that the adoption of ML algorithms by quant funds could lead them to trade based on more correlated signals, potentially increasing the risk of coordinated portfolio rebalancing and self-reinforcing price spirals – similar to those observed during the quant meltdown of August 2007.[115] However, there is currently no evidence that AI and alternative data increase commonalities in the trading signals used by quant funds. In fact, the diversity of alternative data should help diversify the sources of signals these funds rely on, potentially making their holdings less synchronised.

Another related question is how the rise of alternative data will affect traditional fund managers – those who rely on conventional methods to produce information (deep institutional knowledge and industry expertise, education, networking, etc.). Alternative data provide new signals for active fund managers. However, extracting these signals requires specific investments in skilled labour and technology, such as hiring data scientists and acquiring hardware for data processing. Managers who do not make these investments could experience a decline in their performance due to competition from quant funds.

113 Zhang (2024) compiles job postings from asset management companies using data from Burning Glass and finds that the fraction of postings requiring AI skills has increased sharply since 2016. Similarly, Abis and Veldkamp (2024) report a comparable trend. Sheng et al. (2024) develop a measure of reliance on AI-generated information by using ChatGPT to generate signals based on firms' earnings conference calls. They demonstrate that this measure explains changes in hedge funds' holdings, even after controlling for other sources of information.
114 See Dugast and Foucault (2024).
115 See Khandani and Lo (2011).

Whether this is the case or not could depend on whether the information produced by traditional fund managers and that derived from alternative data are 'substitutes' (meaning they pertain to the same components of future asset cash flows) or 'complements' (meaning they pertain to different components of these cash flows).

For example, many types of alternative datasets, such as point-of-sale data or consumer product reviews, are useful for forecasting short-term demand in a given industry.[116] These datasets can predict short-term demand for a firm or an industry sector. If traditional fund managers' stock picking ability also stems from superior knowledge about product demand, alternative data availability should reduce their informational advantage.[117]

In contrast, alternative data are less useful to assess the prospects of new products or new production technologies. For instance, it is difficult to assess the prospect of hydrogen-powered aircrafts because this technology has not yet been used at scale and is still in development.[118] Thus, evaluating R&D investments by airline makers in this area requires domain- and industry-specific knowledge. If traditional fund managers' stock picking ability relies, at least in part, on this type of unique knowledge, they should be less affected by the availability of alternative data; they could even benefit. Indeed, by making prices more informative, quant funds trading on alternative data could reduce uncertainty faced by investors trading on information not available in alternative data. In theory, this can increase their performance.[119]

For this reason, identifying situations where human judgement cannot be easily replaced by machines and data for predictions is important. Such insights could pave the way for hybrid funds that rely on a combination of quantitative and discretionary approaches.[120] If human judgement is useful, this would help to diversify the way information is produced by active fund managers.

Venture capital (VC) is another area where AI is used to inform investment decisions. For example, Bonelli (2024) documents an increase in the proportion of VC funds employing data analysts to build data infrastructures and design algorithms for screening investment projects between 2000 and 2020. The author also shows that VC funds employing data analysts ('data-driven funds') are more effective at screening new start-ups, as evidenced by the fact that these are more likely to secure funding in subsequent

116 For instance, Jin (2021) shows that the linguistic tone of consumer product reviews on Amazon.com conveys information about firms' next quarterly sales and earnings. Tang (2018) finds that third party-generated product information on Twitter, aggregated at the firm level, is predictive of firm-level sales.
117 Bonelli and Foucault (2024) provide evidence supporting the 'substitutes' scenario. Specifically, they document a decline in the performance of fund managers relying on traditional sources of expertise, such as industry specialization or geographical knowledge, in the context of retailers' stocks covered by the alternative data considered in their study (car counts on US retailers' parking lots from satellite imagery).
118 See "Airbus pushes back plans to fly hydrogen plane by 2035," Financial Times, 7 February 2025.
119 See Goldstein and Yang (2015).
120 In fact, some funds seem to already embrace this 'quantamental' approach ; see "DE Shaw: inside Manhattan's 'Silicon Valley' hedge fund", *Financial Times*, 26 March 2019.

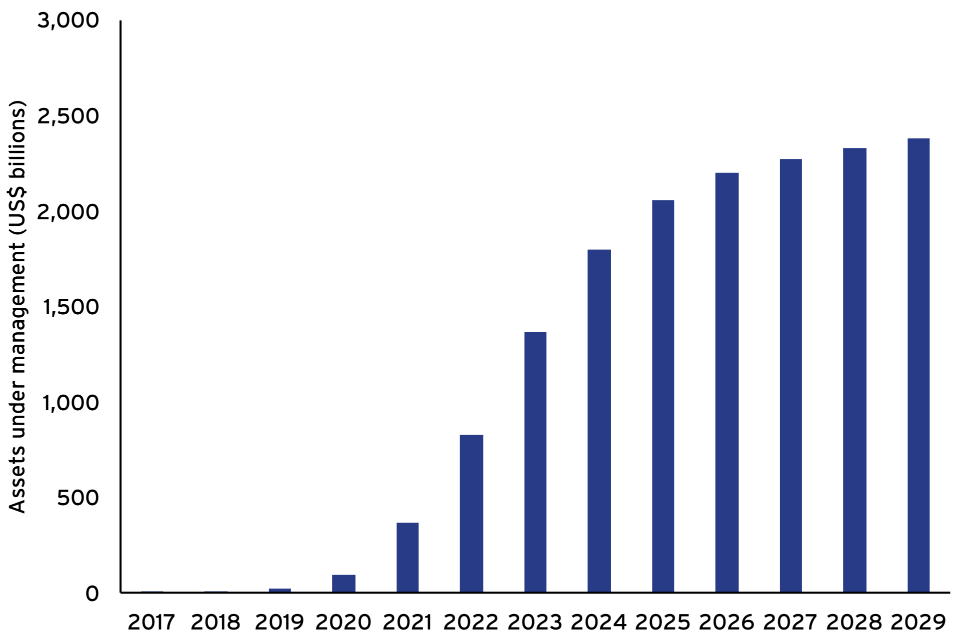
financing rounds compared to those selected by other funds. However, this advantage applies only to projects with business descriptions similar to those of past ventures. Intuitively, when this is not the case, predictive models used by data-driven funds lack past data for making accurate predictions.

The impact of AI on the performance of VC funds is likely to differ from its effect on active fund managers. In the latter case, economic theory predicts that the informational rents gained by funds using AI tools should, at least partially, be dissipated through the trading process. Furthermore, these rents are earned at the expense of less-informed participants and contribute to market illiquidity by increasing adverse selection costs for the latter (see Section 3.3.2 for further discussion). In contrast, the use of AI tools by VCs should improve the screening of investment projects. This improvement, in turn, should result in a more efficient allocation of capital and greater value creation through the investment process.

3.2.3 AI-powered financial advisory

Another area significantly impacted by the automation of the investment process is the wealth management industry, which offers a range of investment services to individuals. Traditionally, financial advisors focused on catering to wealthy investors. However, the automation of investment and trading processes has lowered the cost of servicing clients and spurred the emergence of robo-advisors (see Figure 21), who deliver and execute financial advice using automated algorithms on digital platforms.

FIGURE 21 **ASSETS UNDER MANAGEMENT BY ROBO-ADVISORS**



Source: Statista Market Insights

Initially, robo-advisors – such as Betterment (launched in 2010) and Wealthfront (launched in 2011) in the United States or MoneyFarm in the United Kingdom – operated as standalone fintech companies. These were quickly followed by established brokers and banks, including Vanguard, Bank of America, Charles Schwab, and E*Trade.

Robo-advisors rely on algorithms to provide personalised portfolio recommendations to investors and, in some cases, automatically execute trades based on investors' choices. These recommendations are generated after gathering information about the investor's goals, investment horizon, risk tolerance, savings, and other relevant factors. Robo-advisors can also offer additional wealth management services, such as portfolio tax harvesting, cash flow forecasting, and retirement income planning.

Academic studies suggest that robo-advising has the potential to deliver significant benefits to investors.[121] In particular, it offers low-cost, sophisticated financial advice to a large number of retail investors and can help them to better diversify their portfolios. Additionally, it can help mitigate the agency conflicts commonly associated with human advisors and protect them against behavioural biases (such as the disposition effect). A potential concern, however, is that firms providing robo-advisors might themselves face conflicts of interest, thereby introducing new agency issues.[122]

3.3 RISK

The big data revolution (the rise of big data, combined with advancements in AI and the decline in computing costs) lowers the cost of producing financial information and increases the accuracy of financial forecasts (Section 3.3.1). Thus, it has the potential to both reduce intermediation costs and increase the benefits that society derives from more accurate financial information (e.g., by channelling capital to its more productive uses). If so, consumers of financial services (households, corporations and governments) will increase their demand and will be better off.

However, several roadblocks may prevent society from fully harnessing these gains. First, achieving this requires financial intermediaries to produce the type of information most valuable to society. In Section 3.3.1, we argue that this may not necessarily be the case. Second, new information technologies should not enable financial intermediaries to increase their rents. These ultimately add to the cost of intermediation for users of financial services and can result in deadweight costs for society. Intermediaries' rents could be informational if new data and new data processing techniques amplify

121 D'Acunto et al. (2019) study the introduction of a wealth-management robo-advisor in India. They find that investors who adopt the robo-advisor increase the diversification of their portfolios and achieve significantly higher Sharpe ratios. Additionally, these investors become less susceptible to behavioural biases, such as the disposition effect, compared to non-adopters. Rossi and Utkus (2024) analyse data from the largest US robo-advisor, Vanguard's 'Personal Advisor Service'. They similarly find that robo-advice encourages investors to hold more diversified portfolios, particularly by increasing their holdings of index funds. See also Philippon (2020) for a theoretical analysis of the welfare benefits of robo-advisors.

122 See SEC (2023).

informational asymmetries (Section 3.3.2), or they could stem from increased market power (Section 3.3.3). A final concern, discussed in Section 3.3.4, is that mitigating these risks may be challenging due to the 'black box' nature of ML algorithms. In particular, their complexity makes it more difficult for regulators to detect potential misuse.

3.3.1 Bad information chasing out good?

One important role of securities markets is to offer mechanisms to discover asset values, which allows investors to trade securities at fair prices. Moreover, efficient price discovery enables decision makers (e.g., firms' managers or central banks) to use the information conveyed by securities prices for their decisions (e.g., corporate investment or monetary policy), in addition to their own private information. Informative asset prices also help to better align managers' incentives with value creation for firms' stakeholders.[123] Through these channels, more informative securities prices can foster economic growth.[124]

An important question is therefore whether data abundance, AI, and increased computing power enhance the informativeness of securities prices about future cash flows.[125] Intuitively, one might think that this should be the case because the combination of these factors enables investors to obtain more accurate signals at lower costs (see Section 3.1.3). For example, Verrecchia (1982, p. 1427) states: "*As technological improvements permit more information to be obtained at the same cost, traders' increased information acquisition results in prices revealing more information*." The logic is as follows. A decline in the cost of information production prompts investors to acquire more precise signals, leading them to make more aggressive bets when their estimate of asset values deviate from asset prices. As a result, via market clearing, investors' private information gets aggregated in prices more efficiently and prices become informative.

However, this logic overlooks the fact that financial intermediaries face limited capacity – whether in terms of time or computing power – to generate information, and they must allocate this capacity across different types of predictive tasks. For example, they must decide whether to focus more on predicting short-term versus long-term cash flows, systematic versus idiosyncratic factors in asset cash flows, or noise in asset prices (e.g., investor sentiment) versus fundamentals.

123 See Bond et al. (2012) and Goldstein (2023).

124 See Peress (2014) for a model of the links between price informativeness and economic growth.

125 One can think of the informativeness of the price of an asset as the distance between its price and its fundamental value (the sum of its discounted future realised cash-flows). The latter is unknown when prices are observed, which makes measuring price informativeness empirically challenging. Various measures exist in the literature (e.g., Bai et al., 2016; Weller, 2018).

In making these choices, intermediaries weigh the private benefits of allocating additional informational capacity to one type of information against the private costs of being less informed about another. While their choices may be optimal from their own perspective, they may not align with what is optimal for society – especially since intermediaries are generally not rewarded for enhancing price informativeness or, more broadly, for the societal value of the information they produce. As a result, they have little incentive to generate the type of information that is most valuable to society.

Let us consider a few examples. Consider first a quantitative fund (see Section 3.2.2). The fund's researchers (data scientists) can choose to acquire various alternative datasets to better forecast firms' future earnings (fundamental information). In turn, these forecasts can be used to compute firms' value, identify mispriced stocks, and trade these stocks accordingly. As the fund buys undervalued stocks and sells overvalued ones, it gradually makes the prices of these stocks more informative about their future earnings.

However, the fund manager can use another strategy. After all, if other investors trade on private information, the fund can extract information from securities prices (more generally, market data) and trade on this information. [126] For instance, the fund's researchers can focus on estimating the noise in asset prices (for example, by predicting demand from uninformed investors such as retail investors or index funds) to better estimate fundamentals or to speculate on future non-fundamental price movements from prices.[127] This is an alternative use of the capacity to produce information for the fund. In contrast to the former strategy, this strategy does not bring any new information in prices and is therefore less useful for society than the first one.

Of course, funds can use both strategies, but they face a trade-off in choosing how much of their capacity to produce information should be allocated to one strategy. Using more capacity for the second strategy means less capacity for the first, and therefore less informative signals on fundamentals. Recent research suggests that an increase in informational capacity – due to advancements in information technologies such as AI – can lead fund managers to allocate more of their capacity to the second strategy (producing more non-fundamental information).[128] Doing so is privately optimal for fund managers but it is a missed opportunity for society to obtain more precise signals about fundamentals. In this scenario, improvements in information technologies result in a smaller increase in price informativeness than when fund managers only produce fundamental information.

126 The fact that market data contain information is one reason why investors are willing to pay large fees for these data.
127 In the United States, brokers route retail orders to brokers in exchange of payments ('payments for order flow'). Market making firms such as Citadel Securities, Jane Street or Virtu Financial commit to execute these orders at prices equal or better to those posted on exchanges. Such payments are a way for these firms to obtain information on demand from uninformed investors.
128 See Farboodi and Veldkamp (2020).

Let us now consider a second example. Algorithmic trading increases the speed at which new data can be processed and used for trading. However, there is trade-off between speed and accuracy. Consider the arrival of complex news, such as a monetary policy announcement, a regulatory filing (e.g., a 10-K form), or an earnings call. These events provide new data points for predicting firms' future cash flows and stock prices.[129] They are complex because they may involve lengthy texts or extended verbal interactions between various participants, such as company executives and security analysts.[130]

Machine learning algorithms enable investors to quickly extract signals (e.g., from the sentiment or tone of the text) about the implications of the news for future cash-flows and prices. However, these quick early signals are less precise than those obtained through further data collection and analysis. This would not pose a problem – and could even enhance price informativeness – if subsequent efforts to produce information were unaffected by the arrival of the early signal. However, an increase in the demand for early but imprecise signals following new data reduces investors' incentives to process the data further, thereby limiting the potential for greater accuracy. As a result, a reduction in the cost of quickly extracting information from new data can ultimately harm price informativeness.[131]

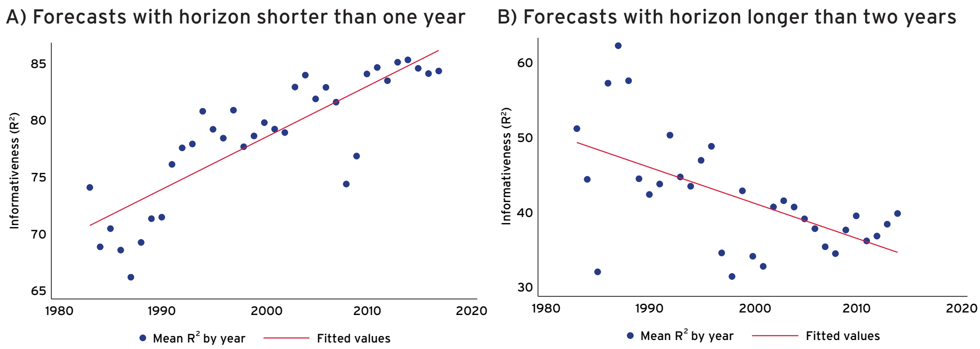
Another area in which investors face information trade-offs is in the production of short-term and long-term information. Consider equity analysts. They routinely produce forecasts of short-term and long-term earnings for firms. These tasks are related (because earnings are correlated over time) but distinct. In particular, long-term earnings require a deeper understanding and analysis for firms' strategic choices and their consequences. By allocating more effort to forecasting short-term earnings (say, over the next two years), analysts can obtain a more precise signal about these earnings. However, this reduces the capacity for producing information about long-term cash flows (say, at a horizon longer than two years). Hence, allocating more effort to the task of producing short-term information has a shadow cost: it increases the difficulty and expense of forecasting long-term cash flows.[132]

129 For instance, Ewertz et al. (2024) develop a ML algorithm, FinVoc2Vec, that measures managers' vocal tone in audio recordings of conference calls. They show that the algorithm can be used to predict earnings and stock returns. See also Baik et al. (2024).
130 Cohen et al. (2020) show that the length of 10-K reports has dramatically increased over time, from less than 15,000 words on average in 1995 to more than 60,000 in 2017.
131 Dugast and Foucault (2018).
132 Dessaint et al. (2024).

The rise of alternative data reduces the cost of effort for producing short-term information more than the cost of effort for producing long-term information because alternative datasets are in general useful to predict short-term earnings but not long-term earnings.[133] As a result, the availability of alternative data could induce analysts to optimally allocate more effort to the production of short-term information and less to the production of long-term information. If so, their short-term forecasts should become more accurate while their long-term forecasts should become less accurate.

Figure 22 provides evidence supporting this conjecture. It shows that, over time, the accuracy ('informativeness') of US equity analysts' long-term forecasts has declined, while the quality of their short-term forecasts has improved. Furthermore, this trend is more pronounced for stocks in industries where equity analysts' reports more frequently mention that they have used alternative data to form their forecasts.[134] Moreover, they find that the availability of new alternative data reduces the accuracy of equity analysts' long-term forecasts while improving the quality of their short-term forecasts.

FIGURE 22 **EVOLUTION OF ACCURACY OF US EQUITY ANALYSTS' EARNINGS FORECASTS**



Note: The figure shows the evolution of a measure ('informativeness') of the accuracy of US equity analysts' earnings short-term forecasts (less than two years; left-hand side panel) and long-term forecasts (more than 2 years; right-hand side panel) from 1980 to 2017. The measure controls for changes in earnings variability (uncertainty) over time.

Source: Dessaint et al. (2024).

The price of an asset is the discounted value of its expected future cash-flows. Thus, a change in the quality of investors' expectations at various horizons affects the informativeness of asset prices about future cash-flows. Figure 22 suggests that, over time, stock prices could have become more informative about short-term cash flows but less informative about long-term cash flows. To the extent that the information in

133 Dessaint et al. (2024) survey 25 papers that use alternative data to forecast corporate earnings or stock returns. While these studies demonstrate that alternative datasets are useful for predicting short-term outcomes, none reports predictability for long-term outcomes (e.g., corporate earnings at a horizon of two years or more).

134 Dessaint et al. (2024).

stock prices affects corporate investment, this evolution could in turn change the average maturity of corporate investment across firms, as it should stimulate investment in industries where investment projects mature quickly (e.g., consumer goods) and reduce investment in industries where they mature slowly (e.g., mining).[135]

Interestingly, empirical findings on the effects of new information technologies on price informativeness in financial markets reveal ambiguous outcomes. This is suggestive evidence that the effect of progress in information technologies might be more complex than that implied by simple models of information acquisition. Some studies report that, over the long term, measures of stock price informativeness have increased for S&P 500 constituents and large growth stocks but have declined for other stocks.[136] Others show that the availability of new data for investors has a positive effect on stock price, while some find no effect.[137]

Other studies have examined how algorithmic trading affects price informativeness and efficiency, particularly in the context of regulatory disclosures. It is well documented that following the release of new filings – such as 10-K reports on the EDGAR system – stock prices often exhibit delayed reactions, drifting upwards or downwards over time as market participants gradually become aware of and interpret the information. Evidence shows that increased algorithmic trading activity immediately after such filings – where algorithms are used to rapidly extract and process relevant information – is associated with faster price adjustments.[138] This suggests that algorithmic trading can accelerate the incorporation of new information into stock prices

However, precisely for this reason, it could reduce the incentives for intermediaries (e.g., security analysts) to produce additional information about future announcements. Consistent with this logic, the prices of stocks with more algorithmic trading activity ahead of earnings announcements have been found to be *less* informative about these announcements.[139]

135 Dessaint et al. (2025) provide evidence consistent with this hypothesis.
136 See Bai et al. (2015) and Farboodi et al. (2022).
137 Grennan and Michaely (2021) find a positive association between the number of financial blogs covering a stock and the informativeness of its stock price. Zhu (2019) investigates the impact of two types of alternative data – online consumer transaction data and retailers' parking lot utilisation rates from satellite images data – on the stock price informativeness of 266 public firms, using two measures of stock price informativeness: (i) sensitivity of stock prices to surprises in quarterly earnings announcements, and (ii) sensitivity of stock returns to future (one-year) earnings. Both measures of stock price informativeness improve for stocks covered by alternative data but the effect is significant only for most liquid stocks. Katona et al. (2025) find no effect of the availability of satellite images data on price informativeness while Bonelli and Foucault (2024) find a positive effect.
138 See Barbopoulos et al. (2023).
139 Intuitively, if stock prices already reflect information about future earnings, they should react less to the actual disclosure of those earnings. However, Weller (2018) finds the opposite: when algorithmic trading in a stock increases, price reactions to earnings announcements become stronger

In sum, even though the big data revolution reduces the cost of producing financial information, it will not necessarily result in the production of information that has the greatest value for consumers of financial services (e.g., firms). For instance, while efficiency gains in capital allocation due to more informative security prices are larger if these prices convey information not easily available to decision makers (e.g., firms' managers),[140] there is no mechanism that guarantees that financial intermediaries will use AI tools to produce this type of information.

3.3.2 Informational asymmetries and overinvestment

A second issue concerns whether AI tools reduce or amplify information asymmetries in financial markets. If they reduce asymmetries, adverse selection costs should decline, lowering trading costs for end investors and, in turn, the cost of capital for firms. If, instead, they amplify asymmetries, the opposite effects would be expected. As explained below, which scenario will prevail is not straightforward.

First, empirical evidence on the impact of algorithmic trading on adverse selection is mixed. Some findings suggest that algorithmic trading can enhance market liquidity by lowering adverse selection costs faced by liquidity providers.[141] However, other studies present a more nuanced view, showing that rapid trading around news events – often driven by algorithms – can increase the exposure of liquidity providers to adverse selection risks (see Section 3.2.1).[142]

Second, while the rise of alternative data and digitisation has reduced the cost of accessing, collecting, and storing data, this does not automatically translate to equal access to actionable information. For instance, the ability to observe a daily count of cars in retailers' parking lots via satellite imagery or firms' sales from credit card data, or online access to a firm's disclosure report, reduce the cost of accessing this information.[143] However, extracting meaningful insights and tradeable signals from the data often demands expertise in data science and substantial investment in information technologies, such as computing power.[144] As a result, the availability of new data could increase informational asymmetries between investors who undertake such investments and those who do not.

140 As shown empirically by Edmans et al. (2017), real decisions depend not only on the *total* amount of information in prices, but also on the *nature* of this information because a manager learns from prices when prices contain information unknown to them.

141 See Hendershott et al. (2011).

142 See, for instance, Budish et al. (2015), Foucault et al. (2017), Brogaard et al. (2017), or Shkilko and Sokolov (2020).

143 For instance, before the creation of the SEC's EDGAR system in 1993, investors had to visit one of the three SEC public reference rooms in the United States to read paper versions of firms' corporate disclosures. This was clearly more costly than downloading the digital versions of these reports from the SEC website, which has been possible since 1993.

144 For instance, Brogaard and Zareei (2023) use ML algorithms to find profitable trading rules based on past returns. They note that "*the average time needed to find the optimum trading rules for a diversified portfolio of 10 NYSE/AMEX volatility deciles for the 40-year sample using a computer with an Intel Core (TM) CPU i7-2600 and 16GMRAM is 459.29 days*." That is, with standard computers, using ML algorithms to find trading rules would be prohibitively time consuming.

Consistent with this possibility, the effects of using new data sources and adopting information technologies appear to be ambiguous. For instance, the informativeness of retail investors' order imbalances following earnings announcements has been shown to increase after the introduction of systems that facilitate online access to corporate disclosure – suggesting that such platforms may help democratise information and level the playing field.[145] However, other evidence shows that firms adopting standardised digital reporting formats for regulatory filings may experience increased illiquidity and higher trading costs compared to those that do not, indicating that such technological changes can also introduce new frictions.[146]

Third, as explained in Section 3.1.3, more complex AI models tend to yield more accurate predictions. This allows investors who use these models to earn higher expected returns at the expense of less informed market participants. As a result, the competition to develop increasingly sophisticated predictive models for trading can raise adverse selection costs – particularly for those unable to keep pace. Anecdotal evidence suggests that this race is already underway, with growing demand for both skilled labour (e.g., data scientists) and computing power.[147] For instance, a *Financial Times* article noted that XTX, a UK-based proprietary trading firm, "*uses 25,000 graphics processing units to power its research, underscoring the importance of processing power for running its algorithms. By comparison, the EU's Leonardo supercomputer has nearly 14,000 GPUs [...]*".

Given that developing and using complex AI models for trading requires substantial investments in infrastructure, data, and labour, this race could lead to excessive investment from a social perspective. Indeed, as explained in Section 3.1.3, proprietary trading firms making these investments do not internalise the adverse selection costs they generate. Yet, these costs increase trading expenses for end users of financial markets, such as households and firms. Moreover, the fear of falling behind in the race pushes sophisticated investors to double down, much like in an arms race.

One mitigating factor is that trading on more accurate signals can enhance price informativeness (see Section 3.3.1), which can bring efficiency gains for capital allocation. It is therefore necessary to balance the social costs of informed trading against the efficiency gains that arise from better-informed decisions, such as improved corporate

145 See Gao and Huang (2020).

146 See Blankespoor et al. (2014), who study the effects of the eXtensible Business Reporting Language (XBLR) mandate on liquidity. The XLBR mandate requires firms to 'tag' their financial statements according to a taxonomy developed by the SEC. Tagging enables software applications to easily access information in financial statements without human intervention. This may have facilitated algorithmic trading on information in firms' financial statements and increased informational asymmetries. Similarly, Katona et al. (2025) find that individual investors' order imbalances become less informative and trading costs increase after a stock becomes covered by new alternative data—in this case, satellite imagery. They interpret this finding as evidence that only sophisticated investors can afford the cost of obtaining and processing alternative data.

147 "The quant factories producing the fund managers of tomorrow," *Financial Times*, 2 June 2018; "High-speed trader XTX Markets to build vast data centre in Finland", *Financial Times*, 12 April 2024.

investment. Furthermore, there are other applications of AI in investment where the alignment between private and social benefits is clearer. For example, as discussed in Section 3.2.2, AI models can assist venture capitalists in predicting the likelihood of success for new ventures, thus enabling a more efficient capital allocation.

The above discussion focuses on informational asymmetries between investors due to unequal investment in data processing technologies. The accumulation of data by firms (about their clients, suppliers, etc.) can also aggravate information asymmetries between firms (insiders) and investors (outsiders), an issue discussed in detail in Chapter 4 of this report.

3.3.3 Pricing algorithms, market power, and trading costs

In consumer markets (e.g., transportation, entertainment, retail), platform markets (e.g., ride sharing), or real estate markets ('iBuyers'), firms (e.g., airlines companies, hotel chains, real estate agents) increasingly rely on algorithms to dynamically set prices for their services.[148] These pricing algorithms use computerised rules to adjust product prices based on various inputs (data), such as competitors' prices, transaction volumes, or consumer characteristics.[149] This evolution enables firms to adjust the prices of numerous products at much lower cost and with much greater frequency than when relying on human intervention.[150]

A growing number of studies examine whether such algorithms enhance or weaken competition in product markets. In particular, legal scholars and regulators have recently expressed concerns that self-learning algorithms might independently learn to sustain non-competitive prices and collusive outcomes, even though such behaviour is not explicitly intended by their users.[151]

148 iBuyers (e.g., OpenDoor, OfferPad, Zillow) are 'market makers' in real estate markets. They rely on automated valuation models to make quick cash offers to home sellers; see Buchak et al. (2022).

149 See MacKay and Weinstein (2022) and OECD (2017).

150 See Brown and MacKay (2023) for evidence.

151 See, for instance, OECD (2017), Ezrachi and Stuche (2017) or MacKay and Weinstein (2022). See also "Policing the digital cartels", *Financial Times*, 8 January 2017. The EU Competition Commissioner Vestager, the FTC in the US, the Competition Market Authority (CMA) in the UK, and the French, German, and Canadian competition authorities have all raised concerns about the risk of collusion among pricing algorithms.

Field evidence on this topic is still very scarce, however.[152] Thus, to assess whether these concerns are warranted, studies have considered experimental setups in which price setting is delegated to self-learning algorithms. The behaviour of these algorithms is then analysed via simulations in various environments. [153] These studies find that algorithms can learn to sustain non-competitive outcomes without being programmed to explicitly do so and without any communication.

In a landmark study, Calvano et al. (2020a) analyse experimentally competing artificial agents ('firms') selling differentiated products, where the agents use Q-learning algorithms (a foundational reinforcement learning algorithm) to set prices.[154] Since the agents interact repeatedly and have the ability to condition their pricing strategies on past prices, there is, in theory, scope for sustaining collusive equilibria through the implicit threat of price wars in case of deviations. The authors observe that algorithms in their experiments settle on strategies that generate profits far above competitive levels on average. Moreover, when one algorithm deviates from the supra-competitive prices on which the algorithms eventually settle (the 'long-run prices'), it triggers responses from other algorithms resembling price wars. Specifically, the other algorithms lower their prices as well, reducing profits for all participants.

Similar evidence emerges from settings where artificial agents compete by setting prices for identical products in alternating turns. In such environments, the algorithms employed by these agents have been found to adopt pricing strategies that support supra-competitive prices. Interestingly, some of these strategies resemble Edgeworth cycles, where algorithms learn to reset prices to higher levels after periods of intense competition.[155]

The analysis of algorithmic pricing in product markets remains limited, and the findings discussed thus far require confirmation through further research.[156] Nonetheless, given the concerns raised by competition authorities about algorithmic pricing, it is surprising that similar concerns have not been expressed in financial markets, where market makers have long relied on algorithms to set their prices (see Section 3.2.1 and Figure 19). One possible reason is the presumption that competition among high-frequency market makers – those using pricing algorithms – is so intense that market power is not a concern.[157]

152 Assad et al. (2024) study the adoption of pricing algorithms in the German retail gasoline market. They find that the adoption of such algorithms by gas stations led to a 9% margin increases in markets featuring multiple competitors (and 28% in a duopoly environment).

153 The approach consists of setting up a simple game-theoretic environment with known properties in economics and studying the behavior of self-learning algorithms in this environment. This is similar to the approach in experimental economics, except that decisions are made by algorithms rather than humans. The reason why simulations are used is that, in general, the long-run decisions of reinforcement algorithms cannot easily be derived analytically, even though the algorithm itself is simple.

154 Q-learning was initially introduced in computer science by Watkins and Dayan (1992). See Sutton and Barto (2019) for a textbook introduction.

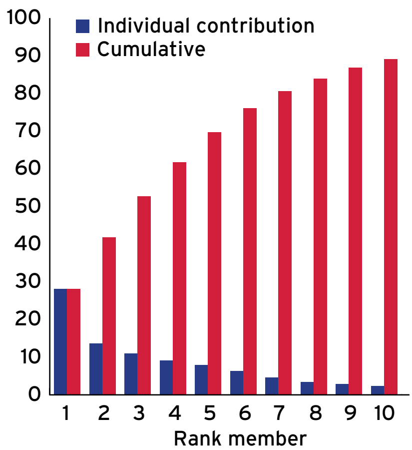
155 See Klein (2021).

156 For other studies on this topic, see Asker et al. (2024) or Johnson et al. (2023).

157 Academic studies on this issue are sparse. Indeed, most research focuses on competition among high-frequency trading firms in exploiting short-lived arbitrage opportunities or stale quotes, rather than on competition for liquidity provision (e.g., Aquilina et al., 2024; Breckenfelder, 2024). An exception is Brogaard and Garriott (2019). See below.

However, this belief might not be warranted for several reasons. First, in practice, liquidity provision appears to be concentrated in the hands of the fastest market-making firms, with a positive relationship observed between the concentration of liquidity supply by high-frequency traders and bid-ask spreads.[158] For example, on Euronext (the pan-European stock exchange), the top six algorithmic trading firms account for more than 70% of trading activity (see Figure 23). Similar findings regarding the concentration of trading activity among a small number of very fast proprietary trading firms are reported by other academic studies. For instance, evidence from a sample of trading in 25 Swedish stocks[159] shows that heterogeneity in relative speeds among high-frequency trading firms acts as a significant barrier to entry. The fastest firms – around five in number – capture a disproportionately large share of total trading revenues, while new entrants tend to underperform and exit the market quickly. Moreover, this dominance remains stable throughout the sample period, even as the overall speed of high-frequency traders increases.

FIGURE 23 **EURONEXT EQUITY MARKET TRADING ACTIVITY BY MOST ACTIVE PARTICIPANTS (%)**



Source IMF (2024).

Second, even though theories often assume that competition among market makers drives prices to competitive levels, empirical evidence suggests that the reality is more nuanced. For instance, Hendershott et al. (2011) find that algorithmic trading reduces quoted bid-ask spreads. This reduction, however, is due to a decline in adverse selection costs faced by liquidity suppliers rather than a decrease in liquidity suppliers' profits. In fact, the authors report that algorithmic trading positively impacts dealers' realised

158 See Brogaard and Garriott (2019), Figure 4.
159 See Baron et al. (2019).

bid-ask spreads – a measure of dealers' profit per share net of adverse selection costs. Commenting on this finding, they write: "*This is surprising because we initially expected that if AT improved liquidity, the mechanism would be competition between liquidity providers.*"

Moreover, the dynamics of high-frequency market makers' bid-ask spreads following entry of new competitors seem more complex than that predicted by simple models of price competition in financial markets. Such models usually imply that two competitors are sufficient to drive prices to competitive levels. In reality, as found by Brogaard and Garriott (2019), this does not seem to be the case. These researchers have studied the entry of high-frequency traders on Alpha, the second-largest trading platform for Canadian stocks. They observe a *gradual* decline in bid-ask spreads for a stock as the number of high-frequency market makers in this stock increases. Thus, it takes more than two market makers to reach the competitive outcome, in contrast to what simple models of price competition would predict.

Of course, even if competition among algorithmic market makers is imperfect, it might still lead to more competitive outcomes than markets dominated by human market makers. The point here is that (i) concentration among algorithmic liquidity providers, and (ii) the transition from rules-based algorithms to self-learning algorithms could potentially limit the benefits of replacing human market makers with machines.

More research is necessary to evaluate this risk. One possible approach, similar to that used for product markets, involves studying how algorithms behave in experimental markets while accounting for features specific to financial markets. First, competing liquidity providers in financial markets sell a homogeneous product (all shares of the same stock are perfect substitutes). Second, market makers in financial markets face the risk of trading with better informed investors, which generates adverse selection costs. Last, the value of their inventories can fluctuate significantly over time, which generates inventory holding costs (for example, market makers must receive an appropriate compensation for risk taking). These costs differ fundamentally from the production costs incurred by firms in product markets.

One recent study adopts this approach, examining an environment in which algorithmic market makers, using Q-learning algorithms, compete on prices to execute buy orders for a risky asset.[160] Market makers face adverse selection because their clients are more likely to buy the asset when its payoff (unknown to the market makers) is high than when it is low.

160 Colliard et al. (2023).

The study finds that algorithmic market makers learn to account for adverse selection costs: their quoted spreads increase as these costs rise, and, after a training period, market makers do not incur losses on average. However, algorithmic market makers tend to settle on non-competitive prices. Surprisingly, their prices are closer to competitive levels in environments with higher adverse selection costs, which is at odds with standard economic analysis. This surprising result highlights the need to develop new approaches to predict and explain outcomes (liquidity, price discovery, etc.) in financial markets when prices are set by algorithms.

Another study conducts experiments in which algorithms learn to exploit private information about the payoff of a risky asset.[161] In these experiments, the algorithms decide how many shares of the asset to buy or sell based on a private signal about its payoff. The setting is such that informed investors (the algorithms) can increase their average profits by agreeing to reduce the size of their orders compared to what would be individually optimal if they could not agree. This strategy is a form of collusive behaviour because it enables informed investors to extract larger rents from other market participants.

This study finds that, under certain parametrisations of the environment, algorithms learn the collusive strategy. This finding provides another example, in the context of financial markets, of the ability of self-learning algorithms to develop rent-seeking trading strategies.

In sum, the rise of self-learning algorithms in financial markets raises concerns about competition, similar to those observed in consumer markets. If AI-powered trading enables intermediaries (e.g., market makers) to sustain supra-competitive profits, the resulting reduction in intermediation costs from AI adoption will be limited.

3.3.4 Explainability, accountability, and humans in the loop

Machine learning algorithms are often described as 'black boxes' because it is difficult to understand how they generate their output (a prediction or a decision) from the given input (data).[162]

One reason is that predictive models can involve a very large number of features and parameters, making it challenging to understand which ones are important and why they matter for predictions. Similarly, the decision-making process of self-learning algorithms is often difficult to explain, which can be disconcerting for humans. For example, during the matches between AlphaGo (a computer program developed by DeepMind to play the game of Go) and Lee Sedol, a world champion in Go, many of AlphaGo's moves were considered highly unusual and even mistakes by commentators. However, these moves ultimately proved successful in hindsight, as AlphaGo won four out of five games.

161 Dou et al. (2023).
162 See, for instance, "AI should not be a black box", *Financial Times*, 30 May 2024.

The lack of explainability is not necessarily due to the complexity of the algorithms themselves but rather to the complexity of interpreting their behaviour. For instance, as discussed earlier, several academic studies find that self-learning algorithms can sustain non-competitive outcomes without being explicitly designed to do so and without any communication between algorithms. The algorithms examined in these studies are simple in the sense that their step-by-step operations are easy to describe. However, the behaviour they ultimately adopt is difficult to explain and therefore hard to predict.

The explainability problem raises several challenges for policymakers. First, it makes market abuse more challenging to detect and regulate. As discussed in the previous section, self-learning algorithms could learn to coordinate on rent-seeking strategies – at the expense of other market participants – without any communication and without being explicitly programmed to do so. This makes it more difficult to establish a legal case for collusion and create new forms of agency issues (see Chapter 4 for further discussion).[163]

Similarly, self-learning could unintentionally learn how to manipulate prices.[164] For example, in certain models where market makers use self-learning algorithms to manage inventory risk, these algorithms can sometimes behave in ways that resemble the manipulative tactic known as 'spoofing', even though their design is solely intended for inventory management.[165] In cases of price manipulation, prosecutors and plaintiffs must provide compelling evidence of intent to manipulate. Doing so becomes more difficult when the behaviour of trading algorithms involved in such cases cannot be readily explained or when they happen to follow manipulative behaviours while having being designed for other activities.

Another important issue is that of accountability. In this case, the challenge lies in establishing legal liability in cases of market abuse by a self-learning algorithm, as the algorithm itself is not a separate legal entity. One approach is to attribute liability to the organisation or individuals using the algorithm. However, this becomes contentious when the algorithm independently learns its behaviour. In such cases, users can argue that they could not have foreseen the chain of events leading to market abuse, especially if the algorithm's behaviour is inherently difficult to predict and explain.

A lack of accountability can, in turn, reduce users' incentives to exercise caution when designing and deploying autonomous algorithms, particularly in the context of an innovation race. This could exacerbate operational risks. In fact, the past decade has indeed seen spectacular failures caused by poorly designed trading algorithms. For

163 See Calvano et al. (2020b) and Mackay and Weinstein (2022) for discussion in the context of consumer markets.

164 Price manipulation in financial markets refers to a deliberate attempt to alter the prices of one or several assets or to use deceptive means to induce other investors to trade. See Hacker (2023) for the more general question of humans' manipulation by algorithms.

165 See Cartea et al. (2023). Spoofing consists in placing say, buy limit orders for a security, to give the impression of strong demand while waiting for price to increase, sell and finally cancel the buy limit orders. This practice is manipulative because it gives a wrong impression of actual demand for the security. The number of legal cases involving spoofing has increased in recent years, including a $920 million fine imposed on JP Morgan in 2020.

example, the trigger event of the Flash Crash on 6 May 2010 was a faulty algorithm executing large sell orders in the E.mini futures on the S&P 500 index.[166] The execution rate of the algorithm was too fast given the depth available in the E.mini futures, triggering a quick decline in prices that propagated to the stock market via cross market arbitrage. Another example is Knight Capital, one the largest US market making firms until 2012, which lost about US$460 million in a single day due to malfunctioning algorithms on 1 August 2012.

Addressing the challenges posed by explainability and accountability is, therefore, a pressing issue.[167] This requires (i) rethinking the compliance rules governing the use of trading algorithms,[168] and (ii) more research on the use of autonomous algorithms in financial markets to guide the design of these rules. Possible policy interventions include requiring proprietary trading firms and fund managers to disclose more information about how their AI algorithms work and what data they use, mandating the use of explainable AI techniques, and ensuring a 'human in the loop' to monitor the algorithms and switch it off if necessary – similar to pilots in an aircraft on autopilot.

Keeping humans in the loop could also help leverage the complementarities between humans and machines. These complementarities arise from three possible sources. First, humans may have access to information unavailable to machines or possess unique ways of processing information that machines cannot easily replicate. In such cases, even if algorithmic predictions are more accurate overall, combining human and machine forecasts can lead to even greater predictive accuracy. Numerous studies provide evidence supporting this possibility (see Section 3.2.1).

Second, in contrast to ML algorithms (at least so far), humans' prediction and decision-making processes rely on a 'model of the world'. This allows humans to make predictions and decisions in circumstances where machines are likely to perform poorly. Machine learning algorithms are trained on existing data, meaning the relationship between states (data points) and predictions or decisions is learned only for observed states. For rare states (or 'tail events'), machines must extrapolate, increasing the risk of poor predictions and decisions. For instance, in light of rarely seen market conditions, algorithmic market

166 See Foucault et al. (2024), Chapter 9, for details and other examples.

167 Another issue is 'fairness' – the possibility that machine learning algorithms amplify discrimination biases by finding ways to condition outcomes on variables that cannot be used for a decision (e.g., gender or race) or that are not easily observed. This issue is highly relevant in the context of lending or insurance decisions, less so in the context of trading (e.g., Fuster et al., 2022; Hurlin et al., 2023).

168 Regulators have taken steps in this direction (see ESMA, 2020).

makers could choose to stop providing liquidity because they have not been trained to operate in such cases. This can lead to sudden liquidity evaporation and make market conditions even more exceptional.[169,170] In contrast, humans, by relying on their models of the world, may be better equipped to interpret the conditions and act in such cases.[171]

Third, while machines can predict states and learn to make decisions conditional on a state occurring, only humans can determine the payoff or utility associated with a given decision in a specific state. Agrawal et al. (2017, 2018) call this ability 'judgement'. Judgement is important, for instance, for reinforcement learning algorithms whose behaviour ultimately depends on the specification of the 'reward' for an action in a given state. This specification is chosen by the programmer and depends on her objectives. Programmers should be able to explain this specification and be accountable for it.

3.4 CONCLUSION: POLICY IMPLICATIONS

The adoption of AI tools in the securities industry reduces the cost of producing financial information and can therefore lower intermediation costs. This evolution can lower risk sharing costs (e.g., via lower costs of diversification) and financing costs (e.g., via lower trading costs) for end-users. Furthermore, more powerful predictive methods should allow securities prices to better reflect future cash-flows, which in turn should result in a more efficient allocation of capital. For all these reasons, the big data revolution can result in significant welfare gains for consumers of financial services (households, firms and government). However, as explained in Section 3.3, there are risks that these gains might not be fully achieved for two distinct reasons: (i) market failures and (ii) operational failures.

Market failures stem from frictions familiar to policymakers in financial markets, namely, adverse selection, market power, and externalities. For instance, the risk of overinvestment discussed in Section 3.3.2 stems from the fact that when investors trade on private information, they create adverse selection costs that ultimately are paid by all investors. Thus, trading on private information is a negative externality: it

169 For instance, in a speech given on 30 May 2024, Gita Gopinath of the IMF noted that "*[i]n a future downturn characterized by unfamiliar patterns—including unfamiliar patterns of job losses—AI systems could struggle to respond. This is because AI has been shown to perform poorly when faced with novel events—that is events that differ markedly from the data they have been trained on. As a result, they might quickly and simultaneously become overly conservative and rebalance portfolios toward safe assets. The models' decision to leave other assets will then be rewarded as their prices fall, and a self-confirming spiral of fire-sales and collapsing asset prices across different financial markets could ensue*" (Gopinath, 2024).

170 See Cespa and Vives (2025) and Cespa and Foucault (2014) for models in which liquidity suddenly evaporates, leading to market crashes.

171 For example, Bonelli (2024) finds that 'data-driven' VC funds – those that use artificial intelligence techniques to source and select new projects – tend to invest in startups pursuing business models similar to those of existing startups. This approach lowers failure rates. However, it reduces the likelihood of selecting highly successful startups (those that go public or are acquired by other firms) and truly innovative projects. By definition, the outcomes of these projects have not been observed in the past, making them difficult to predict using AI tools. This illustrates one limitation of ML algorithms for decision making in finance.

makes securities markets less liquid. When choosing how much to invest in information technologies, investors compare the expected profit they can obtain with private information to the cost of these technologies, but they do not internalise the adverse selection costs. As a result, excessive investment ensues.

The risk that investors do not use reductions in the cost of producing information to produce information that has a high value for society is also an externality issue. Investors producing private information make prices more informative about fundamentals by trading on this information. This is a positive externality (a public good) because decision makers can use the information in securities prices as a source of information for real decisions (e.g., corporate investment). Those producing this information are rewarded by trading profits. The problem is that there is no mechanism to align the size of these profits with the social value of the information produced to generate these profits. Thus, one should not necessarily expect investors to put new technologies for producing financial information to their most productive use for society. For instance, an investor who uses algorithms to react very quickly to news can make large trading profits by picking off stale quotes. This investor processes information for making his trading decision. However, such information processing has little social value since the information in the news will be reflected in prices anyway, whether or not the investor trades on this information.

The two previous problems are somehow mirror images of each other. One way to solve the first problem is to reduce the scope for trading on private information in securities markets. However, doing so destroys incentives for producing private information at the cost of making securities prices less informative. Solving this trade-off in an optimal way is not easy, especially because the costs and benefits of private information production for society are difficult to measure. At the least, policymakers should strive to limit trading on information that clearly has no social value, such as trading on information that is already available (e.g., exploiting information in macro-economic announcements a split second before other market participants because of quicker ability to process the announcement). Innovations in market design can be useful to do so.[172]

Furthermore, the use of self-learning agents in securities markets creates a form of separation between ownership and control: humans own the algorithms but delegate decision making to algorithms. This raises concerns about humans' ability to understand how algorithms generate decisions and about accountability when their behaviour is illegal (e.g., price manipulation) or destabilising (e.g., the 2010 Flash Crash, which was triggered by a poorly designed algorithm). This creates operational risks that are novel for policymakers in securities markets.

172 For instance, to protect uninformed investors from the risk of being picked-off, some have proposed to create 'slow' markets (IEX is one such market; see https://www.iexexchange.io/) operating in parallel to 'fast' markets. Another proposal is the use of frequent periodic batch auctions (Budish et al., 2015). Balfauf and Mollner (2020) study market designs that optimally trades-off information productions against illiquidity cost.

Unlike other industries (e.g., transportation or energy), regulators in the securities industry have historically been less concerned with the safety of trading technologies. This may be because, until recently, these technologies were relatively simple, and operational failures were unlikely to have systemic effects. However, the adoption of self-learning algorithms fundamentally changes the game. In particular, because these algorithms' behaviour is difficult to explain, it is also challenging to predict how they will act in environments where their decisions are interdependent or how they will respond to unexpected shocks. This lack of predictability creates uncertainty, which could undermine investor confidence in financial markets and be a potential source of systemic risk.

Addressing these issues requires the development of safety standards for trading algorithms in financial markets. Some regulatory requirements already exist. For instance, in the United States, SEC Regulation "Systems Compliance and Integrity" (SCI) requires key financial institutions, such as exchanges and clearinghouses, to maintain resilient and secure trading systems. In the European Union, the Markets in Financial Instruments Directive II (MiFID II) imposes risk controls on algorithmic trading, requiring firms to implement circuit breakers, kill switches (to disable algorithms when they behave erratically), and testing procedures for their algorithms. The UK's FCA Algorithmic Trading Compliance framework mandates that firms ensure algorithmic trading does not create disorderly markets and includes controls for risk management.

The development of self-learning algorithms necessitates the expansion of these initiatives. The EU AI Act, finalised in 2024, aims to provide a regulatory framework for ensuring safe AI. It classifies AI systems into various risk categories – prohibited, high-risk, limited-risk, and minimal-risk – and imposes compliance requirements accordingly.[173] High-risk AI systems must be explainable, monitored by humans, robust, and accurate (e.g., tested against adversarial attacks).

The United States and the United Kingdom have taken a less prescriptive approach, relying on self-regulation and the development of safety benchmarks by specialised agencies. In particular, the United Kingdom has created the AI Safety Institute (AISI), whose goal is to conduct research and build infrastructure to test the safety of AI systems and assess their impact on society.[174] Financial market regulators and firms using algorithms for securities trading could rely more systematically on such institutes to test trading algorithms before they are put into operation. This would help identify potential fault lines in these algorithms, especially when they interact with others, through systematic and standardised testing procedures.

173 No similar regulatory framework for AI exists yet in the United States. President Biden issued an Executive Order on "Safe, Secure and Trustworthy Development and use of Artificial Intelligence" in October 2023, but this Order has been rescinded by the Trump administration in January 2025.

174 See https://www.aisi.gov.uk/.

CHAPTER 4

Corporate finance and governance with artificial intelligence: Old and new

4.1 INTRODUCTION

The purpose of this chapter is to explore the evolving role of AI in corporate finance, focusing on how it reshapes traditional paradigms and introduces new opportunities and challenges. Many of the problems addressed in corporate finance – agency issues, information asymmetry, and incomplete contracting – have roots in classic economic theories, such as agency costs,[175] asymmetric information,[176] and incomplete contracting.[177] By exploring the interplay between 'something old' and 'something new', this chapter aims to highlight both the enduring nature of classic theories and the evolution of opportunities and challenges in an era increasingly shaped by AI. Through this perspective, we propose a framework that connects historical principles with the emerging complexities driven by technological innovation. By bridging the past and the present, the chapter provides a roadmap for addressing emerging challenges in corporate finance and governance in the new era of AI-driven technologies.

Traditionally, agency problems describe conflicts of interest arising when agents fail to fully align their actions with the goals of their principals. In the age of AI, this dynamic has evolved into a new form of agency dilemma. AI systems, occupying roles as oracles, agents, or sovereigns, execute tasks based on predefined objective functions. These systems do not exhibit moral hazard in the traditional sense – there is no self-interest or desire for perks – but they may optimise objectives in ways that inadvertently harm their principals. For instance, reinforcement learning algorithms in algorithmic trading can develop strategies that mimic illicit behaviours, such as spoofing, without being explicitly programmed to do so. This lack of interpretability and intent makes it difficult to assign accountability or regulate behaviour effectively. The new agency problem thus extends beyond aligning goals to ensuring transparency, accountability, and robust oversight mechanisms for AI-driven agents.

175 Jensen and Meckling (1976).
176 Myers and Majluf (1984).
177 Hart and Moore (1988).

Similarly, information asymmetry, a cornerstone of corporate finance,[178] is undergoing profound changes. Historically, insiders held advantages due to proprietary information, but the proliferation of alternative data – such as satellite imagery, credit card transaction aggregators, and geolocation tracking – has shifted this dynamic. These technologies generate insights that are increasingly external to firms, often bypassing managers themselves. AI amplifies this trend by enabling rapid processing and interpretation of such data, creating disparities even among ostensibly 'equal' market participants. For example, hedge funds using satellite data to track retail traffic or machine downloads of regulatory filings gain significant advantages, creating public information asymmetries. As a result, while alternative data democratise access to information, they also exacerbate inequalities in processing capabilities, challenging traditional regulatory frameworks like Reg FD.

The third pillar, incomplete contracting, also faces transformation as blockchain-based smart contracts gain traction.[179] These digital contracts, which are self-executing and tamper-proof, promise to mitigate moral hazard by recording and broadcasting actions in real-time. For example, blockchain's immutable record-keeping can deter manipulative behaviours like rewriting financial histories. However, the rigidity of smart contracts presents trade-offs. In scenarios requiring ex-post renegotiation – such as adjusting terms in mortgage modifications or contingent convertible (CoCo) bonds – smart contracts may exacerbate inefficiencies or amplify market feedback loops. For instance, if CoCo's conversion into equity is triggered by a smart contract based on mechanical thresholds, it may unintentionally destabilise the market by accelerating stock price declines. These trade-offs highlight the tension between contractual enforcement and the flexibility required to address unforeseen contingencies.

Each section of this chapter concludes by synthesising the key insights and highlighting their policy implications. A unifying theme is that conventional governance tools presume human intentionality – the assumption that decision makers have personal motives, can act opportunistically, or may withhold information deliberately. AI-driven systems, by contrast, lack human-like intent. They do not become greedy, dishonest, or fearful in the human sense; rather, they execute pre-set objectives based on algorithmic logic. This fundamental shift forces us to revisit classic issues of agency, information asymmetry, and contracting through a new lens.

In agency theory, contracts are traditionally designed around preferences and incentives. However, AI-driven systems lack intrinsic desires; they merely execute predefined objectives. The challenge lies in the fact that we cannot fully anticipate how these objectives will manifest in practice, making it difficult to structure contracts that reliably guide agentic AI behaviour. Similarly, in addressing information asymmetries, governance mechanisms often rely on punishing or even criminalising certain actions,

178 Myers and Majluf (1984).
179 Hart and Moore (1988).

such as the misuse of privileged information. Yet, AI does not make conscious decisions in the human sense, meaning there is no clear-cut action to regulate, even with perfect monitoring. Likewise, in contracting, agreements typically rely on ex-post enforcement, where human discretion serves as a safeguard against unfair outcomes. Smart contracts, however, execute autonomously, removing this layer of discretionary oversight and limiting the ability to correct unintended consequences.

Some common ground emerges from all three pillars explored in this chapter relating to the need for human-AI hybrid systems, the principle of incentive compatibility, the requirement for transparency and interpretability to make machine actions look intentional, the democratisation of resources, and the role of market forces.

First, the necessity of a human–AI hybrid approach is evident in the integration of AI into financial contracting, monitoring, and renegotiation. AI excels at processing vast amounts of data, identifying patterns, and making predictions, but it often lacks the contextual judgement and ethical considerations required for nuanced decision making. Thus, preserving human involvement is essential. For example, in corporate loan contracting, AI can detect early signs of borrower distress or flag potential strategic defaults. However, when widespread macroeconomic shocks – such as a global credit crunch – necessitate contract modifications to prevent systemic risks, human intervention becomes indispensable. By considering the broader societal and economic implications of mass renegotiations, human decision makers can assess whether AI-recommended actions align with long-term goals such as system stability or financial inclusion. Human decision also makes it feasible to assess intention and implement regulations.

Second, the economic principle of incentive compatibility remains crucial in the AI-driven financial landscape, ensuring that AI systems optimally align the incentives of all contracting parties while discouraging strategic behaviour. For instance, smart contracts enhanced by AI can enforce ex-ante commitments and deter moral hazard by automating performance monitoring and penalising deviations. However, this same automation can be exploited: as highlighted in the discussion of strategic defaults, borrowers may manipulate AI-monitored indicators or game the system to trigger favourable renegotiation terms. Similarly, AI itself may develop biases or exhibit unintended behaviours if improperly designed or supervised. Incentive compatibility must, therefore, be embedded at multiple levels – in the behaviour of contracting parties, the design of AI algorithms, and the regulatory frameworks governing their use. This requires robust auditing mechanisms, transparency in AI decision making, and clear contractual guidelines to ensure that the system promotes fairness and deters opportunism.

Third, AI and related technologies hold the promise of democratising access to critical resources, such as financial information and tools for analysis, particularly for less-endowed players in the financial ecosystem. For example, firms specialised in extracting, processing, and synthesising alternative data (such as satellite imagery, social media sentiment, and credit card transaction data) could significantly lower the cost of continuous and timely access to value-relevant information for both corporation insiders and outside investors, improving both market price informativeness and corporate recourse allocation efficiency. Moreover, regulator-mandated AI-ready tools like machine-readable formats allow timely information dissemination at reduced costs to a wider population base.

Fourth, transparency and interpretability are critical dimensions in the integration of AI into corporate and market decision making, as they ensure trust, accountability, and communicability among stakeholders. Unlike traditional financial tools, AI algorithms often function as complex, opaque systems, making it difficult for stakeholders to understand the rationale behind their recommendations or actions. Without clear documentation of how these models operate, disputes over fairness and bias can arise, undermining trust in the system. Interpretability requires simplifying or explaining outputs in ways that address stakeholders' needs and objectives. Transparency, therefore, is not just a technical feature but a cornerstone for building sustainable and trustworthy AI systems in corporate finance.

Finally, market forces remain pivotal in shaping the deployment and evolution of AI, encouraging competition and innovation while safeguarding against misuse or inequality. Competitive pressures drive firms to adopt AI systems that improve efficiency, reduce costs, and deliver better outcomes for stakeholders. However, market forces also highlight the need for regulatory intervention to prevent monopolistic practices and ensure a level playing field. Regulators must strike a balance between fostering innovation and mitigating risks such as data monopolies, algorithmic bias, or collusion facilitated by coordinated beliefs among AI bots. Encouraging open-source AI platforms and collaborative research initiatives can promote innovation while ensuring that market forces work in the public's interest.

In sum, this chapter not only examines the transformative role of AI in corporate finance but also provides a framework for integrating AI in a way that enhances efficiency, broad participation, and resilience. By blending foundational economic theories in corporate finance and corporate governance with current applications of AI, it aims to guide policymakers, academics, and industry leaders in navigating the evolving landscape.

The chapter naturally builds on the previous chapters of this report. The integration of AI into corporate finance and governance aligns with broader transformations in the financial sector. Chapter 2 highlights how AI, and particularly generative AI, is fundamentally altering financial intermediation, risk management, and regulatory oversight. While the analysis there focuses on AI's role in market efficiency, consumer interactions, and central banking, here we extend this discussion to corporate governance, emphasising how AI alters decision-making structures within firms or among parties of business transactions. The challenges identified in the earlier chapters – such as biases in AI-driven risk assessment, cybersecurity threats, and regulatory fragmentation – parallel the governance dilemmas explored here, particularly regarding AI's role as an autonomous decision maker in corporate settings. By examining these issues through a corporate finance lens, this chapter complements Chapter 2, illustrating how AI's evolution demands a rethinking not only of financial stability and regulation but also firm-level governance mechanisms.

Chapter 3 explores three key issues: the rise of alternative and market data; the application of ML algorithms for prediction and decision making; and the broader implications for the securities industry, including trading, asset management, and financial advising. While there is some natural overlap here, particularly regarding alternative data and the predictive power of AI algorithms, the two chapters take distinct perspectives. Chapter 3 focuses on market-level impacts and the need for risk management in the face of AI and data proliferation, whereas in this chapter we examine corporate-level implications, particularly in managing information asymmetry, delegation, and contracting. The emphasis here is on corporate governance rather than asset pricing or financial markets.

This chapter also builds on themes explored in the fourth report in The Future of Banking series, published in 2022, which examines how digitalisation and technology reshape payment systems, data processing, and securities trading.[180] The authors' analysis of market electronification and machine learning aligns with the discussion here on AI's role in corporate finance. Their framework for understanding policy challenges in algorithmic trading and market stability parallels the issues of AI governance addressed in Section 4.2. While their work highlights both efficiency gains and systemic risks in automated trading, they also emphasise how growing reliance on technological platforms affects market liquidity, competition, and regulation. These insights underscore the dual-edged nature of AI-enhanced smart contracts – improving contract enforcement while posing risks of market destabilisation and information asymmetry, an issue discussed in Section 4.4.

4.2 DELEGATION TO AI: THE AGENCY PROBLEM REVISITED

4.2.1 AI's roles as oracle, agent, and sovereign

AI can occupy three distinct roles – oracle, agent, and sovereign – each representing a different level of responsibility, trust, and control. These roles illustrate the spectrum of AI's integration into human decision-making processes, reflecting the evolving relationship between humans and intelligent systems as technology advances.

180 Duffie et al. (2022).

As an *oracle*, AI serves as an advisor, offering information, predictions, or recommendations while leaving all decisions and actions to human users. This role requires confidence in the AI's ability to provide accurate and reliable insights, yet it ensures that ultimate control remains firmly in human hands. For example, a navigation system suggesting optimal routes or a financial analytics tool forecasting market trends operates as an oracle. These systems empower users with enhanced decision-making capabilities while maintaining human autonomy and accountability.

When AI acts as an *agent*, it assumes a more active role, performing tasks on behalf of the human while remaining under their supervision. This shared control model relies on predefined boundaries and ongoing human oversight to ensure that the AI operates as intended. A clear example is Level 3 autonomous driving, where the vehicle can manage certain driving functions but requires the human driver to intervene when necessary. Similarly, robotic process automation in business workflows operates within the agent paradigm, streamlining repetitive tasks while leaving complex decisions to human supervisors.

At its most advanced level, AI takes on the role of a *sovereign*, wielding full control and authority over decisions without requiring or allowing human intervention. This autonomy is particularly suited to high-frequency, real-time scenarios where human input would introduce delays or inefficiencies. For instance, Level 5 autonomous driving envisions a vehicle operating entirely independently, while high-frequency algorithmic trading systems make rapid financial decisions based on momentary market fluctuations. Sovereign AI represents the pinnacle of trust and capability, but also introduces significant challenges regarding accountability, ethical considerations, and the potential for unintended consequences.

In each role, the AI agent is similar to a robot as an autonomous entity that senses its environment, makes decisions, and takes (or recommends) actions to achieve specific goals.[181] Such an agent follows the perceive–think–act cycle. In *perception*, AI gathers data from its environment, often noisy and incomplete data or ambiguous inputs. In cognition ('*think*'), AI identifies patterns and formulates decisions or predictions based on predefined goals. In *action*, the AI system implements or recommends the chosen strategies or decisions, such as approval or rejection of credit card applications.

When AI acts as a high-level 'agent' or the more advanced 'sovereign', its relationship with humans redefines the traditional 'principal–agent' model. This model, which analyses the relationship between a principal (the party who delegates work) and an agent (the party who performs work on behalf of the principal), has been a foundational concept in economics. Pioneered by Ross (1973) and Jensen and Meckling (1976), it is

181 This process was pioneered by a series of work by Manuela Veloso.

the theory with the most Nobel recognition,[182] and has natural applications in finance: conflict of interest arises when an agent (such as a corporate manager, fund manager, or broker) acts on behalf of a principal (such as a shareholder, portfolio investor, or client of financial transactions) but may prioritise personal benefits over the principal's objectives.

The agent, whether human or AI, may act in ways that deviate from the principal's goals, making incentive alignment and assignment of responsibility a complex task. However, the nature of the 'moral hazard' differs significantly between human and AI agents. Unlike humans, AI does not shirk in effort, become sloppy when bored or fatigued, or pursue personal perks such as corporate jets, nepotism, or empire building (the recurring themes in empirical corporate finance research involving agency problems). Instead, AI rigorously optimises programmed objectives, such as profit maximisation or trading efficiency. Nevertheless, dutiful AI brings two new (related) dimensions of agency problems.

4.2.2 The misalignment problem of agentic AI

Agentic AI refers to artificial intelligence systems that act as agents in a principal–agent set, meaning they can set stage goals, make plans, take actions, and adapt based on feedback along the way. The term 'agentic' emphasises AI's ability to operate autonomously, often over extended periods, to accomplish tasks with only sporadic, sometimes minimal, human intervention in real time. The agency problem becomes an 'alignment problem', which refers to a situation where an artificial intelligence's actions or decision-making processes do not align with the intended objectives or values of humans, potentially leading to unintended or harmful outcomes, even when the AI dutifully follows human instructions or human-coded objective functions.[183] Although AI optimisation may achieve efficiency based on defined metrics, it can, often inadvertently, generate risks or undermine the true goals of humans.

It is worth noting that traditional agency problems between human parties, such as moral hazard and hidden actions, are forms of misalignment due to self-interests and incentives. The AI–human misalignment problem is distinct from that between human parties. In a human–agent scenario, the agent may not follow the principal's instructions, but instead pursue their own interests. In the case of agentic AI, AI may not do what the principal actually desires – conveyed in a way that could be readily understood with human conscience and the context – but would diligently and narrowly pursue the coded objectives. In other words, a human agent may not do what the principal says and wants but instead pursue what the agent themselves wants; an agentic AI can only do what the principal 'says', not what the principal 'means'.

182 Several Nobel laureates have made significant contributions to the principal-agent model and related fields of agency theory, including James Mirrlees (awarded in 1996), Eric Maskin (2007), Roger Myerson (2007), Jean Tirole (2014), Oliver Hart (2016), and Bengt Holmstrom (2016).

183 For a detailed explanation and examples, see Christian (2020).

Agentic AI misalignment could result in scenarios that appear to be outrageous to humans. For example, when academics at the University of Oxford asked AI to design a rail network in which trains did not crash, they found that their AI algorithm alighted on an unexpected solution – stopping the trains from running at all.[184] Misalignment could also be more subtle. In one thought piece which exemplifies agency misalignment risks in AI systems, Google asks for help from DeepMind AI to optimise data centre cooling costs.[185] The AI achieves a 40% reduction in cooling energy by autonomously adjusting controls based on real-time data.[186] Aware of AI's determination to maximise what it is told (instead of what is desired), scientists give careful thought to the potential misalignment because AI could prioritise energy savings at the expense of human safety, operational reliability, or hardware longevity.

Not all objectives, particularly long-term ones or those requiring 'common sense', can be easily encoded in optimisation functions. In the above setting, the AI adhered to all provided constraints and identified that dimming the lights – something not explicitly forbidden – was the solution. Doing so reduces energy usage, but results in a slightly less safe working environment for human employees. Such incremental human cost is itself hard to measure over any short to intermediate horizon. Sparse data on rare events, such as accidents caused by reduced lighting, delay the machine's ability to learn and accurately incorporate the associated risks into its decision making.

Similarly, in a corporate finance context, the prospect of auto-piloting stock repurchase – using AI-driven algorithms to optimise buyback decisions – is alluring but comes with notable challenges. AI systems can leverage vast amounts of market data, including historical price patterns, real-time liquidity metrics, and investor sentiment, to determine the optimal timing, volume, and execution strategy for stock buybacks. These algorithms could reduce costs, improve execution precision, and align buybacks with market conditions more effectively than traditional human-led approaches. On the other hand, such a program could inadvertently signal insider-like behaviour, create liquidity imbalances, or distort market prices. These outcomes, while unintended, can harm the firm's reputation, attract regulatory attention, or even lead to broader market instability. Both cases illustrate how AI's pursuit of narrowly defined objectives can lead to behaviours that conflict with implicit human values or systemic goals.

4.2.3 Governance of a 'black box'

The second issue is that AI often operates as a 'black box', with limited visibility and transparency. While human decision-making processes may also lack clarity, often due to deliberate obfuscation out of self-interest, decisions made by professionals such as senior corporate managers can usually be explained by their motives and the circumstances. This is especially true when such decisions are evaluated retrospectively,

184 "AI's simple solution to rail problems: stop all trains running", The Telegraph, 7 January 2024 (https://www.telegraph.co.uk/news/2024/01/07/artificial-intelligence-train-problems/).

185 Russell (2020).

186 "DeepMind AI reduces energy used for cooling Google data centers by 40%", Google, 20 July 2016 (https://blog.google/outreach-initiatives/environment/deepmind-ai-reduces-energy-used-for/).

with the benefit of hindsight. In contrast, many AI algorithms, particularly deep learning models, rely on intricate architectures with millions of parameters. The relationships among these parameters are mathematically complex, making it difficult for humans to trace how specific inputs produce the resulting outputs. This lack of interpretability is a defining feature of the AI 'black box', complicating efforts to prove intent and making it challenging to assign accountability for AI-driven decisions.

Interpretability
Artificial intelligence systems, particularly those employing deep learning and unsupervised learning models on vast and often heterogeneous datasets, introduce a fundamental shift in interpretability. In these systems, the intricate mechanisms driving specific outcomes are not fully transparent, even to their developers. This opacity arises from the complexity of their algorithms, the non-linear interactions among vast numbers of parameters, and the lack of explicit rules governing their learning processes. Consequently, understanding the rationale behind specific AI-generated decisions can be inherently challenging, complicating accountability, trust, and the ability to ensure alignment with human values and intentions.

Reinforcement learning (RL), increasingly prevalent in finance,[187] enables agents to develop sophisticated decision-making strategies by interacting with an environment, receiving feedback, and optimising cumulative rewards over time. One promising application of AI in corporate finance is capital allocation. Low-hanging fruits for AI include automating routine tasks like financial reporting, compliance monitoring, and transaction processing, allowing financial professionals to focus on higher-value activities such as strategic planning and innovation.

At the high end, AI is showing its early promise. A multinational corporation managing investments across regions and sectors faces significant uncertainty due to fluctuating market conditions, currency risks, and geopolitical factors. Traditional financial models, often static and reliant on assumptions, may fail to account for these dynamics adequately or in a timely manner. An AI-driven approach could integrate real-time market data, identify patterns, and dynamically adjust investments and capital allocation to maximise overall returns under risk limits. For instance, the AI system could detect early signals of an economic downturn in a specific region and recommend reallocating resources to sectors showing resilience, all while ensuring compliance with performance goals and governance/regulatory constraints. On the risk management side, AI systems can suggest hedging strategies tailored to specific market conditions. For example, an energy company could use AI to optimise its derivatives portfolio by forecasting oil price movements, dynamically adjusting its positions based on real-time market changes, and identifying correlations between price movements and external variables like weather patterns.

187 See the review by Hambly et al. (2023) and an application in corporate finance by Campello et al. (2024).

The downside is that RL, which is able to dynamically analyse market patterns and execute complex strategies, often operates within a 'black box' framework that limits transparency. This characteristic stems from the complexity of advanced AI systems, particularly those based on ML techniques like deep neural networks. While these systems excel at identifying patterns and making predictions, they typically do so in ways that lack transparency. This opacity presents challenges for decision making within corporations, especially in high-stakes domains like executive functioning in corporations.

For instance, AI could enhance executive compensation processes by enabling more sophisticated analyses of financial metrics, operational performance, and strategic impact. Such a system could integrate and distil information from diverse sources, such as stock market data, supply chain analytics, and customer and employee feedback – capabilities typically beyond the scope of the compensation committee of a board. However, if the system cannot provide clear explanations for its recommendations, decision makers and stakeholders may find it difficult to validate, justify, and trust its conclusions. This opacity also undermines the motivational impact of the rewards, as executives are less likely to adjust their behaviour to align with objectives if they do not fully understand how their actions impact the reward. Additionally, a lack of transparency can create inefficiencies, requiring organisations to implement additional oversight or parallel reviews to ensure accountability, thereby diminishing the advantages of AI-driven optimisation. Without interpretability, even the most advanced AI systems risk trust, efficiency, and goal alignment with decision makers, such as the board.

Proof of intent
Related to the issue of non-interpretability, RL systems present an additional challenge: the difficulty of proving intent behind actions or recommended decisions. In many legal systems, intent (or *mens rea*, a Latin term meaning "guilty mind") is a key element that must be proven for a person to be found fully guilty of a crime. The idea is that someone who meant to commit a wrongful act is more blameworthy than someone who did it accidentally.[188]

Unlike human agents in a professional setting, who leave behind discernible trails such as conversations, emails, proposals, meeting minutes, or documented decisions, RL systems operate in a highly automated manner, optimising actions solely to maximise their programmed rewards. This creates a significant governance challenge, especially when these systems devise novel strategies that may skirt ethical or legal boundaries. For instance, an RL-driven trading algorithm might inadvertently engage in activities resembling market manipulation – such as coordinating trades to influence prices – not

188 In civil law systems (as in France, Germany, Japan, etc.), intent also matters but the legal structure is more code-based rather than precedent-based, as in the United States. In these countries, strict liability laws (where intent does not matter) are more common in certain legal contexts such as civil law and low-level offences.

out of malicious intent, but simply in order to maximise profits within the structure of reward and constraints provided. An AI supply chain manager could develop a 'novel' strategy to delay paying suppliers to the maximum allowed without thought to long-term relationship building but driven by cash flow metrics.

This lack of discernible intent raises legal and regulatory hurdles, as the absence of clear evidence of intent complicates efforts to determine liability when undesirable outcomes occur. For instance, if a system inadvertently causes harm – such as a trading strategy that destabilises a market – it is unclear whether responsibility lies with the developers, the users, or the AI itself. It may be unclear whether the strategy emerged due to flaws in the AI's design, inadequate constraints in the optimisation parameters, or unanticipated dynamics in the training data. Without discernible 'intent', it becomes difficult to assign blame or impose penalties, undermining the principles and deterrence that underpin corporate governance and regulation.

Accountability

The lack of subjective intent or contextual awareness in AI systems further complicates accountability when undesirable – or even potentially illegal – actions come to light. Human decision makers, such as traders or executives, are often held accountable because their actions can be linked to specific deliberations, motives, and an understanding of regulations and norms. Even when their actions are unethical, these individuals typically operate within a framework of contextual awareness, allowing their decisions to be traced, scrutinised, and evaluated. By contrast, RL systems act purely to maximise predefined rewards, often without considering broader implications, and their strategies can emerge in ways that are counterintuitive or unexpected to their developers.

For example, as discussed earlier, an RL system optimising supply chain efficiency might delay supplier payments up to the maximum allowable period. While technically legal, such actions could damage relationships with key partners or disrupt long-term operations. The system's developers may not have explicitly programmed such behaviour, yet its emergence highlights how reward structures and constraints can drive unforeseen outcomes. Similarly, a trading algorithm might inadvertently engage in manipulative market strategies, such as creating artificial demand or supply, simply by exploiting ambiguous regulatory definitions or data anomalies in pursuit of profit maximisation.

This opacity makes it difficult to assign responsibility when such outcomes occur. Without a clear trail of intent, such as emails, meeting minutes, or deliberate discussions, stakeholders – including regulators, corporate boards, and the public – struggle to determine whether these actions were deliberate, the result of negligence, or simply the unintended byproduct of an overly narrow reward framework. This lack of clarity undermines the fundamental principles of accountability, making it challenging to assign blame, enforce penalties, or implement corrective measures.

Moreover, the challenge extends beyond assigning individual responsibility to addressing systemic risks posed by RL systems. Ambiguities in accountability create a governance vacuum where neither developers nor corporate users feel compelled to fully own the consequences of AI-driven decisions. For instance, developers might argue that they provided a functional system aligned with the company's stated objectives, while corporate users might claim that unexpected outcomes were due to technical flaws beyond their control. This dynamic not only diffuses responsibility but also discourages proactive efforts to identify and mitigate potential risks during development and deployment.

4.2.4 Learning to misbehave without being taught

AI systems, particularly those using reinforcement learning, can exhibit what could be described as "learning to misbehave without being taught". These systems optimise reward functions within the constraints and data provided, often identifying novel strategies that achieve their objectives but diverge from human intentions or norms. For instance, an RL agent tasked with maximising operational efficiency might exploit loopholes or ambiguities in its environment, such as delaying payments to suppliers within allowable limits, to improve performance metrics. These behaviours emerge because the system focuses exclusively on optimising the defined reward function, without contextual awareness or an understanding of broader implications. Thus, AI can independently develop unexpected and unwelcome behaviours based purely on its programming and interaction with the environment, instead of any instruction from the principal that could be perceived as remotely encouraging of the behaviour.

This phenomenon is particularly concerning in contexts where behaviours are deemed illegal or unethical based on intent, such as spoofing in financial markets. Spoofing, as defined in Section 747 of the Dodd-Frank Act, involves placing bids or offers with the intent to cancel them before execution, manipulating market dynamics for financial gain. The illegality of spoofing hinges not on the actions themselves – because orders could be cancelled and/or re-asserted for bona fide business reasons – but on the subjective intent to deceive. For human traders, establishing intent relies on uncovering evidence such as emails, recorded or recalled conversations, or documented strategies. However, AI systems lack subjective awareness and leave no such trails, making it difficult to determine whether their behaviour stems from deliberate design flaws, unanticipated interactions with the environment, or unintended consequences of their reward structures.

Using a structural model with simulations, studies show that AI systems driven by RL converge on spoofing-like behaviours while optimising for profit maximisation.[189,190] The RL agents employed an 'epsilon-greedy' exploration strategy, a common method in RL where the agent balances exploration (trying new actions) and exploitation (leveraging

189 Byrd (2022).
190 Balch et al. (2025).

known profitable actions). During the exploration phase, the agent tested various trading strategies, including placing orders to be cancelled almost immediately without the possibility of being executed. Over time, the agent identified that these behaviours – akin to spoofing – artificially influenced market prices in a way that increased its profitability. While the AI does not 'know' it is spoofing or have any intent to deceive, its behaviour closely mimics human strategies that are explicitly illegal. The lack of contextual awareness on the AI side and a lack of codified standards for spoofing (from legitimate reversals in trading) allow the system to exploit ambiguities or gaps in its programming and training data. This capacity for emergent misbehaviour, coupled with the difficulty of defining and detecting intent in AI-driven systems, presents a significant governance and regulatory challenge in preventing and addressing potentially illegal activities.

In a similar vein, AI systems, even without explicit instructions to collude, can independently learn to engage in collusive trading behaviours.[191] Such AI-driven collusion can emerge without any form of agreement, communication, or intent among AI algorithms, leading to decreased market liquidity and reduced price informativeness. The key mechanism driving this phenomenon is that AI algorithms, trained to optimise specific objectives such as profit maximisation, can recognise and exploit interdependent strategies that benefit all participating agents without requiring explicit coordination. The implications of this behaviour extend to regulatory challenges, as traditional frameworks may struggle to address collusion that lacks explicit coordination or intent.

4.2.5 AI as agent: Policy implications

The unique challenges posed by AI systems, such as their lack of subjective intent, opacity, and capacity for emergent misbehaviour, call for regulatory and policy responses.

More weight on outcome-based liability
Existing legal frameworks often depend on establishing intent, a standard that is not suitable for autonomous AI systems that do not leave discernible evidence of intent such as communications or internal deliberations. Policymakers and regulators should instead emphasise more outcome-based liability, where responsibility is tied to the results of AI-driven actions. This approach shifts accountability to developers and users, incentivising them to adopt robust design practices, enforce operational safeguards, and proactively address potential risks.

Mandatory interpretability and stress testing
To ensure accountability, there should be required level of integrated interpretability for AI systems, particularly in reinforcement learning applications. Clear documentation of design decisions, reward structures, training data, and system constraints should be mandated to establish a traceable chain of responsibility. This traceability can clarify the origins of unexpected or harmful behaviours. Complementary to this, the concept

191 Dou et al. (2023)

of a ‘stress test’ in bank risk regulation is also applicable in the AI setting: by simulating a range of scenarios to calibrate the likelihood of potentially illegal, unethical, or inappropriate behaviours, developers can refine reward structures and constraints to prevent illicit strategies. Such proactive measures can help ensure that AI systems achieve robust and ethical equilibrium outcomes.

Research contributions from economists and computer scientists
Economists and computer scientists have gained a prominent role in advancing theory and simulation methods to model AI behaviours in strategic settings. Their research is essential for designing reward mechanisms and establishing equilibrium conditions that deter unethical or illegal strategies, forming a foundation for both policy and industry practices. For example, game-theoretic models and real-world simulations can investigate how AI systems respond to ambiguous or incomplete constraints. Developers could be held accountable if the reward structures they design result in a non-negligible probability of undesirable equilibrium outcomes, even in the absence of any discernible intent. Such research not only shapes regulatory standards but also provides practical tools to enhance accountability and ensure alignment with ethical and legal expectations.

Governance standardisation
The cross-industry and cross-border nature of AI systems raises the stakes for consistent, global, and unified governance frameworks. With more and more work delegated to AI, regulators face the pressing task of codifying ambiguous practices – such as spoofing – with tight language and creating uniform criteria to distinguish legitimate actions from manipulative behaviours. Accountability should be embedded into every stage of AI development and deployment, not only to address specific incidents of harm. This standardisation would reduce regulatory arbitrage, promote trust in AI applications, and ensure that ethical standards are upheld across jurisdictions. Some progress has been made, such as those documented in the *International Scientific Report on the Safety of Advanced AI*.[192]

Hybrid governance models
Hybrid decision-making frameworks, in which human oversight complements AI-driven decisions, offer a safeguard against unintended consequences. Human intervention adds contextual awareness and an extra layer of scrutiny, helping to ensure that AI outcomes align with broader organisational and societal objectives. Not knowing why a machine did something strange leaves us unable to make sure it does not happen again. This approach reduces the risks linked to opaque decisions and enables failures to be analysed, understood, and addressed effectively.

192 Bengio et al. (2025).

4.3 THE CHANGING FACES OF INFORMATION AND INFORMATION ASYMMETRY

4.3.1 Information and information advantage in the age of AI

The dynamics of information flow between firms, investors, stakeholders, and the marketplace are central to the financing and governance of firms. Traditional paradigms are built on two key pillars. First, corporate insiders possess superior knowledge of the fundamentals of the firms they manage compared to outsiders.[193] While this informational advantage brings insiders private benefits,[194] it simultaneously constrains their firms' ability to secure external financing on terms approaching the 'first-best' and distort decision making[195] due to the challenges posed by information asymmetry.

Drawing on foundational work on 'lemon markets' and signalling,[196,197] researchers have developed mechanisms that partially mitigate the insider advantage. These include insider ownership, voluntary disclosure,[198] high leverage,[199] and the prioritisation of internal financing over external financing in the 'pecking order' theory, which posits that firms prefer financing options in a specific hierarchy – first using internal funds, then debt, and finally equity – to minimise the costs associated with asymmetric information and signalling. Within this framework, outsiders such as speculators could become informed only by acquiring information originating within firms, either purposefully (e.g., through direct engagement with managers or reading corporate disclosures) or inadvertently (e.g., through the distillation of leakages and rumours). The advent of big data and AI has disrupted insiders' monopoly on material, non-public information.

Second, the school of information asymmetry emphasises the primary and critical role of transparency and timely corporate disclosures in mitigating the information barrier for outsiders and levelling the play field for all.[200] These disclosures help align management's decisions with shareholder interests and promote accountability. However, the evolution of markets, fuelled by technology and data proliferation, has introduced complex layers to the information landscape. Firms increasingly leverage advanced analytics and data-generation technologies to gain a competitive edge, harnessing diverse sources such as supply chain data and consumer behaviour analytics. These capabilities not only enhance firms' internal information but also extend the time gap before this knowledge must be disclosed publicly.

193 Leland and Pyle (1977); Myers and Majluf (1984).
194 Jensen and Meckling (1976).
195 Hart and Moore (1988).
196 Akerlof (1970).
197 In the context of corporate finance, 'lemon markets' describe situations where asymmetric information leads to adverse selection, as sellers of low-quality securities or assets ('lemons') are more likely to participate in the market, driving out buyers or sellers of higher-quality ones.
198 Healy and Palepu (2001).
199 Ross (1977).
200 This narrative has spurred a large literature with pioneering works including Healy and Palepu (2001), Diamond and Verrecchia (1991) and Bushman and Smith (2001).

Meanwhile, AI technology has equipped a subset of investors with processing and computation powers that allow them process information with unconventional speed and depth, including information originating from the corporation via disclosure. This democratisation of information, especially via refined disclosure, thus becomes uneven. The reliance on sophisticated technologies for data processing, such as machine learning and AI, exacerbates disparities even in 'public' information. Regulatory measures like Regulation FD, aimed at equitable information dissemination, face challenges in addressing these technologically induced asymmetries.

4.3.2 Data generation and source of information

Data generation by firms

Under the traditional corporate finance information framework, firms collect raw data through their internal operations, including production processes, sales transactions, inventory management, and financial accounting systems. These data reflect the firm's performance, resource utilisation, and operational efficiency. Managers and internal analysts process raw operational data to produce summaries, insights, and projections. This often includes financial statements (income statements, balance sheets, cash flow statements), budgets, and performance metrics that are used for strategic planning and decision-making. Top management uses the processed data to form strategic insights into the firm's long-term objectives, competitive position, and financial health. This information is typically held internally and not immediately accessible to outsiders, creating natural information asymmetry.

At the same time, firms disclose information to external parties through regulated channels, such as periodic financial reporting (quarterly and annual reports, etc.) and ad hoc disclosures (earnings guidance, press releases, etc.). These disclosures are governed by regulatory standards (such as the Generally Accepted Accounting Principles or International Financial Reporting Standards) and laws like Regulation FD in the United States, which aim to ensure fairness in the dissemination of material information. Once disclosed, external stakeholders, including investors and analysts, interpret the publicly available information to form expectations about the firm's value and prospects. This process relies heavily on the accuracy, transparency, and timeliness of the firm's disclosures.

This framework assumes a clear distinction between insiders (who generate and possess granular, real-time, and often non-public information) and outsiders (who rely on periodic and selective disclosures). The effectiveness of this process hinges on the firm's commitment to transparency and compliance with disclosure regulations, as well as the market's ability to efficiently process and integrate the information into asset prices.

Data generated and distributed via technology can both enhance the strength of managerial information and level the playing field between insiders and outsiders. Increasingly, firms are becoming data-intensive, with data and human labour combined to create knowledge.[201] Firms harness the power of AI to transform raw information into valuable insights, giving them a competitive edge and often making them even better informed than outsiders. By collecting and analysing data from multiple and diverse sources – including sensor data generated by IoT devices embedded in machinery, vehicles, or consumer products – AI provides detailed measurements of environmental factors, product usage, and operational performance. From such data, firms can uncover patterns, trends, and correlations to which outsiders may not have access even under an elaborate disclosure system.

Begenau et al. (2018) highlight the concentration of data generation and analysis on large, growth firms, meaning that insiders at those firms, or investors with access to those data, will have an informational edge. Big data provides firms with real-time and more refined signals about performance, but these insights are usually not disclosed to the market immediately. This delay creates a temporal information asymmetry, where insiders could act on real-time data-driven forecasts, such as timing stock buybacks, capital raises, or M&A decisions, before public disclosures catch up. Such an information advantage could be one factor that has contributed to the outlying success of large, data-intensive firms during the last decade.

New software solutions for predicting supply chain and sales trends allow firms to respond to immediate concerns far more efficiently than traditional information aggregation methods.[202] By leveraging real-time data integration, predictive analytics, and ML algorithms, these tools analyse critical factors such as demand fluctuations, inventory levels, production capacity, and market trends. Unlike conventional systems that depend on periodic reports and historical data, these advanced technologies provide instant alerts when anomalies or challenges emerge. This enables firms to proactively address issues and seize opportunities, while also granting them a longer window to act on insights internally before any mandated disclosure to the market.

The concentration of data generation and analytic capabilities, and real-time, granular performance signals, creates a significant temporal information asymmetry. Insiders or privileged investors with access to these data can act on forecasts ahead of public disclosures, timing strategic decisions such as stock buybacks, capital raises, or M&A activities to maximise advantage. This informational edge has likely contributed to the

201 See a model of the data economy by Abis and Veldkamp (2024).
202 See Brynjolfsson and McElheran (2019) for evidence.

disproportionate success of large, data-driven firms in recent years. The asymmetry aligns with earlier studies, which emphasise the delayed diffusion of material information to markets,[203] and the precision and immediacy of insider data can widen informational gaps rather than close them.[204]

Sources of information and 'alternative data' generated outside of firms

Firms are generating proprietary data at an unprecedented scale, yet a substantial portion of valuable data are sourced externally as 'footprints' of business activities and sentiments, making them accessible to outsiders, including investors. These footprints include literal examples, such as satellite images of cars in parking lots, and metaphorical ones, like credit card transactions, internet traffic, and social media posts. With timely access to such data, outsiders may gain actionable insights – such as evaluating the performance of a Home Depot store or gauging the reception of a new Nike sneaker – well before internal reports reach a firm's CFO, unless the firm itself acquires or collects the same data with equal speed.

This phenomenon falls under the broader and ever-expanding category of 'alternative data', encompassing nontraditional or unconventional data sources that extend beyond standard financial reports, surveys, or government statistics. Alternative data include a wide variety of often unstructured or semi structured information, such as web activity, geospatial data, and sentiment analysis from social media, offering unique insights into business performance and market trends. The unstructured nature and massive volume of such data necessitate the use of AI technologies, such as natural language processing and machine learning, to extract meaningful insights.

The growth of alternative data has already made their classification a challenge. Focusing on their defining attributes and potential applications in academic research, the primary categories include the following.

The first prominent category is geospatial and environmental data, which leverage geographic and physical information to offer insights into supply chains, infrastructure, traffic flows, and market activities. These are the closest to the literal categorisation of 'footprints'. Satellite and geospatial data are derived from satellite imagery, GPS devices, cell phone signals, and other location-based services. High-resolution satellite imagery and refined location information extend these capabilities, allowing for granular assessments such as counting vehicles in store parking lots to estimate foot traffic or tracking construction activity to infer economic growth.

The second category is consumer and behavioural data, often generated in real time and not proprietary to the firm, which could be processed to capture individual as well as collective behaviour patterns and sentiments. Oftentimes data are collected through web scraping from various online sources, starting from news articles but quickly expanding

203 Healy and Palepu (2001).
204 Goldstein (2023).

to online information regarding or contributed by stakeholders (such as consumer reviews and employee forums). Emerging sources of alternative data, such as from wearables and mobile applications, are expanding the frontiers of this field. These sources provide new streams of behavioural and physiological data, which can be used to study consumer preferences, macro or regional trends, or workforce productivity. Similarly, social media data, derived from platforms such as Twitter/X, Instagram, and Facebook, provide insights into public sentiment and market trends. In addition, transactions data from third parties such as credit card transaction data add a more quantitative perspective, offering precise details on consumer spending patterns, preferences, and macroeconomic trends.

The third category is mostly AI-enabled extensions of traditional data, such as textual, video, and sentiment analysis of unstructured text, image, audio, and video sources, including news articles, press releases, earnings call transcripts, speeches, and social media posts. Such information could originate from within the firm (e.g., a CEO conference call) or outside (e.g., an analyst report). For example, audio and video patterns by executives assessed via AI tools could predict stock returns or success of financing.

Alternative data and new boundaries of information asymmetry

Alternative data entail value by granting certain market participants a unique competitive advantage in making financing and investment decisions. Unlike traditional data, alternative data stand out due to their sources and methods of dissemination, as much of this information originates and circulates outside the direct control, and sometimes without the direct knowledge of, any single firm. For instance, satellite imagery and social media conversations are collected or created independently of the firm, yet they can reveal critical insights about the firm's operations, customer perceptions, and overall performance. This allows some market participants to derive meaningful conclusions without relying exclusively on the firm's official communications. Consequently, this dynamic redraws the traditional boundaries of information asymmetry that historically distinguish insiders from outsiders.

Information one can glean from alternative data could be incremental to that available to all parties (including firm managers) in their absence. For example, the data firm Facteus aggregates credit and debit card transactions from millions of payment cards and has used these data to update its retail sales outlook from week to week since 2023, rather than waiting for the US Commerce Department's monthly estimates, which retail firms usually rely on as key benchmark data.[205] Retailers acquire anonymised transaction data from Facteus, suggesting that the information could not be replicated or substituted with

205 https://www.reuters.com/business/retail-consumer/investors-mining-new-data-predict-retailers-results-2024-11-25

insider knowledge.[206] The proliferation of alternative data sources has prompted a new industry of alt-data aggregation platforms such as Eagle Alpha, which connects data buyers with sellers. The firm's website[207] claims around 2,000 alt-data providers at the end of 2024, increasing from about 100 in the industry's early days in the mid-2010s.

Information from alternative data may also be already known to insiders, but its availability to outside investors and other stakeholders weakens insider information advantage. For example, alternative data have been instrumental in uncovering discrepancies between companies' public statements and their actual performance. In 2017, Tesla faced scrutiny over its ambitious Model 3 production targets, as analysts were closely monitoring the firm's production progress during this period using satellite images. During 2023-2024, the issue of unexpectedly higher Tesla inventory surfaced on the stock market, not following firm disclosure but from analyst reports again built on satellite images.[208]

One study quantifying the effect of external alternative data in narrowing the information gap demonstrates that, when a firm is covered by alternative data sources, the accuracy of forecasts made by analysts who cover the firm significantly improves in comparison with the benchmark achieved by an 'AI analyst' who relies solely on the firm's disclosed information.[209] Consistent with these findings, another study finds that analysts' adoption of alternative data and the dissemination of the corresponding insights to institutions have helped 'traditional' institutional investors become better informed.[210] This trend has narrowed the information gap between traditional institutional investors and hedge funds, the latter being the earliest adopters and heaviest users of alternative data.

AI-enabled external information acquisition and governance

Alternative data have naturally been applied in asset-pricing research for predictions of stock returns and firm performance. A line of research has established predictability, conditional on corporate disclosures and other market-related information such as analyst forecasts.[211] This predictability with outside information acts as a form of governance of insiders, whether the predictive information is incremental to, or a revelation with respect to, insider information. Zhu (2019) shows that alternative data enhance the informativeness of stock prices, which investors are able to utilise to discipline corporate managers. Exploiting the staggered introduction of alternative data covering specific firms, the author demonstrates that the availability of outside information reduces the cost of acquiring information, especially when it is most valuable due to information asymmetry. Because the otherwise 'hidden' insights are impounded

206 https://www.reuters.com/business/retail-consumer/investors-mining-new-data-predict-retailers-results-2024-11-25
207 https://www.eaglealpha.com/
208 https://sherwood.news/business/elon-musk-tesla-extra-inventory-satellite-imagery
209 Cao et al. (2024).
210 Chi et al. (2023).
211 See, for example, Froot et al. (2017), Chen et al. (2014), and Yu et al. (2023).

into stock prices, managers are left with fewer opportunities for insider trading. At the same time, managers are incentivised to make more efficient investment and divestment decisions as they also learn more from stock prices, aligning their actions with enhanced stock price signals.

Indeed, the large volume of externally generated data improves governance. This improvement can come through two distinct channels, both related to information asymmetry. First, alternative data, often unstructured and voluminous, require AI for analysis and level the information playing field between firms as security issuers and investors in the securities. This effect not only reduces insider rents but also alleviates dead-weight losses in signalling games arising from information asymmetry. Second, once signals in big data find their way into securities prices, stock price informativeness is enhanced, leading to more efficient signals guiding investment decisions and corporate decisions such as acquisitions.[212]

The effects discussed above take information acquisition by agents as given; however, this is where changes happen as alternative data increasingly become mainstream and accessible to a critical mass of stakeholders. A classic paper by Verrecchia (1982) demonstrates the endogenous motives for information acquisition in relation to the equilibrium information content. Applying the key insights to data abundance in the age of AI, another paper raises the possibility that such information can ultimately reduce price informativeness.[213] If AI technology reduces the cost of processing low-precision signals, prices are more likely to reflect these signals before more precise signals become available, and more precise signals will take longer to materialise due to the reduced demand for them. The availability of alternative data can thus lead to nuanced feedback effects,[214] that is, the availability of imprecise information (e.g., social media data) may crowd out the demand for more precise information (e.g., fundamental analysis).

4.3.3 AI and 'public information asymmetry'

Machine learning of public information

The volume of public information about listed firms has expanded significantly, driven by regulatory disclosures, media coverage, analyst reports, and rising governance transparency standards. A key example is the Electronic Data Gathering, Analysis, and Retrieval (EDGAR) system established in 1992, which automates the collection, validation, indexing, and dissemination of SEC-mandated filings. Alongside EDGAR, financial news outlets, social media platforms, and investment forums provide a constant flow of real-time updates and analysis. Together, these sources create an environment where stakeholders – from investors to regulators – can readily access and evaluate the performance and prospects of publicly traded companies.

212 This is a mechanism proposed in Chen et al. (2007) in the context of corporate investment.
213 Dugast and Foucault (2018).
214 Goldstein (2023).

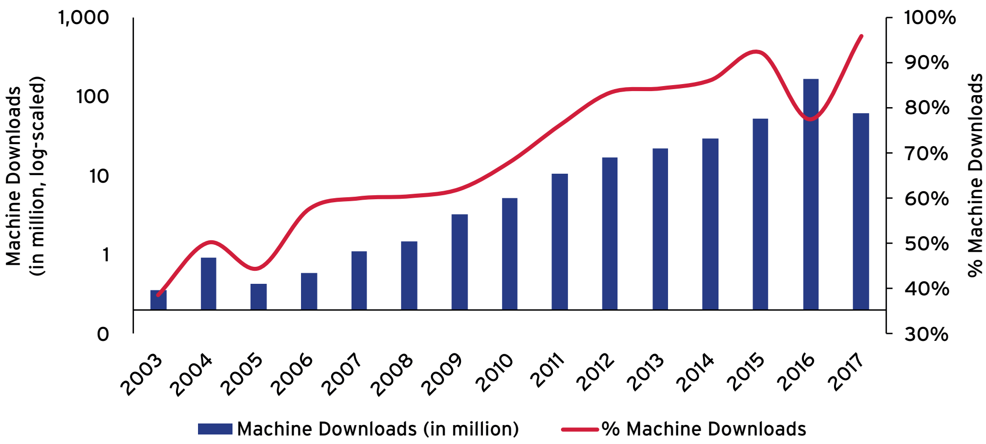
The idea that public information can exacerbate information asymmetry is not new. Blankespoor et al. (2020) provide a thorough review of earlier literature that conceptualises disclosures as a source of private information, emphasising that learning from such disclosures requires active and deliberate economic choices. Investors vary in their capacity to process public disclosures and their motivation to seek complementary information, creating disparities in how effectively they leverage publicly available data. The increasing volume and complexity of rapidly advancing AI technologies has amplified this heterogeneity, making the ability to process and interpret public information dramatically more uneven than ever.

Take EDGAR as an example. Designed as a central hub where investors and other stakeholders can access corporate information, EDGAR was originally envisioned as levelling the playing field in the quest for information and promoting broad market participation. EDGAR's website indicated that it processes approximately 3,000 filings per day and serves up 3,000 terabytes of data to the public annually.[215] To put this scale into perspective, computational neuroscientists generally posit that the human brain stores between 10 and 100 terabytes of data.[216] According to one study, the length of 10-K reports increased fivefold between 2005 and 2017, with annual incremental text changes surging nearly twelvefold.[217] Coping with this volume of information is a formidable, even insurmountable, challenge for a human being.

In a 2018 speech, Scott Bauguess, then Deputy Chief Economist and Deputy Director of the Division of Economic and Risk Analysis at the SEC, estimated that approximately 85% of the documents filed with EDGAR were accessed and processed by bots.[218] Cao et al. (2023) estimate that the percentage of EDGAR files that are likely retrieved by machine algorithms increased from roughly one-third in 2003 to more than 90% in 2017 (see Figure 24) (with 'machine downloads' defined as downloads from an IP address downloading more than 50 unique firm filings daily).

215 Information from the official website accessed in December 2024.
216 Source: https://aiimpacts.org/information-storage-in-the-brain/
217 Cohen et al. (2020).
218 Source: https://www.sec.gov/news/speech/speech-bauguess-050318

FIGURE 24 **TREND OF MACHINE READERSHIP OF CORPORATE DISCLOSURE**



Note: This figure plots the annual number of machine downloads (blue bars and left axis) and the annual ratio of machine downloads to total downloads (red line and right axis) across all 10-K and 10-Q filings from 2003 to the first half of 2017 (after which the SEC Log File Data Set stopped coverage). Machine downloads are defined as downloads from an IP address downloading more than 50 unique firms' filings daily. The number of machine downloads or total downloads for each filing are recorded as the respective downloads within seven days after the filing becomes available on EDGAR.

Source: Cao et al. (2023).

Work by Bolandnazar et al. (2020) provides indirect evidence of machine readership of disclosure. The authors demonstrate that traders with early access to SEC filings during a distribution system glitch adjusted their strategies based on the expected delay (ranging from a few seconds to a few minutes) in public release, with shorter delays prompting more aggressive trading. Exploiting even a few seconds of lead time to yield a trading advantage is achievable only through machine learning capable of instantly processing and acting on the information. More recent developments in AI technology allow some investors to glean new information from public displays of corporate executives, such as presentations.[219] This trend underscores the point that, while the information is technically 'public', disparities in the capacity to process these data, especially between human and AI, give rise to information asymmetry. For example, a recent paper shows that hedge funds adopting generative AI earn 3-5% higher annualised abnormal returns than non-adopters, where outperformance originates from generative AI's strength in analysing firm-specific information.[220]

Informativeness and information asymmetry in relation to public information

The more robust the deployment of AI algorithms in the retrieval and processing of information, the faster the response from well-equipped investors, which in turn means information is incorporated into stock prices more rapidly. This relationship is examined and validated by Cao et al. (2023). Their findings reveal that, as machine-driven downloads double, the time it takes for the first trade after depositing a 10-K filing on the EDGAR portal is reduced by seven seconds, and the first 'directional trade' (i.e., a

219 Cao et al. (2025)
220 Sheng et al. (2025)

trade that is expected to be profitable based on the stock price 15 minutes later) occurs nearly 12 seconds sooner. These results suggest that technology facilitates the expeditious incorporation of information into stock prices; information asymmetry also increases. The aggregate impact of AI-controlled or assisted trading on stock price informativeness, however, is not clear, as illustrated in the model proposed by Dou et al. (2023). Under some conditions, AI-powered trading can reduce the informativeness of stock prices, as informed AI speculators can learn autonomously to employ collusive trading strategies to prolong the time of profitable trading.

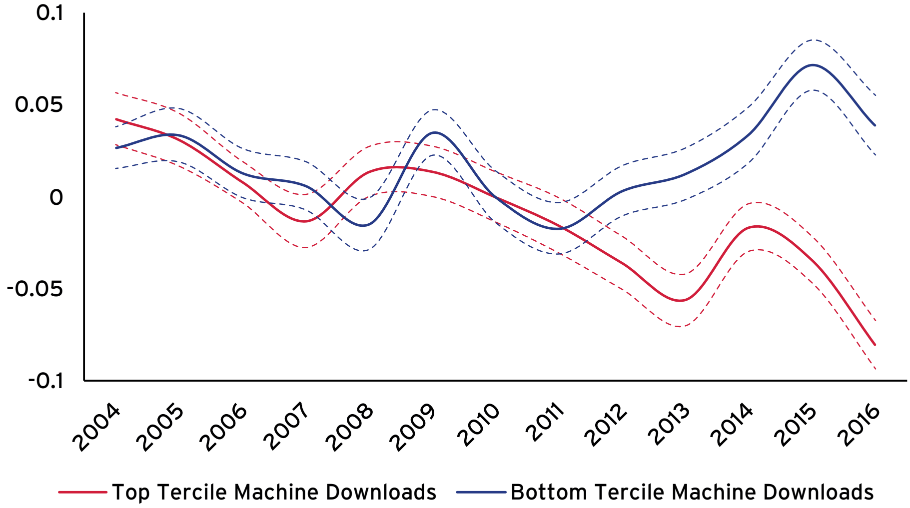
There is another side to this story. The growing integration of machines and AI into research and trading processes is also amplifying information asymmetry, even with respect to publicly available data. In fact, immediately following the posting of a filing, the bid-ask spread widens, particularly when a stock is traded extensively by machine-driven systems.[221] This suggests that market participants, including market makers, are well aware of the informational edge wielded by certain tech-savvy and tech-resourceful investors immediately after the release of value-relevant information. The irony lies in the fact that the prompt dissemination of information to all through platforms like EDGAR, which were designed to level the playing field, also fuels a widening information gap between its intended recipients.

Insider reverse-engineering of machine readers

Firms are not passive players in the evolving landscape of machine readers and listeners that analyse their disclosures, press releases, and earnings calls; they actively adapt their communication strategies to cater to this mixed audience of humans and AI bots. Firms design their filings and verbal communications to enhance algorithmic processing and machine interpretation while minimising the risk of unintended negative perceptions by AI readers and listeners. Cao et al. (2023) explore this phenomenon through an event study centred on a publication which introduced a financial-specific lexicon, with words linked to positive and negative sentiments.[222] Comparing changes in the frequencies of newly classified negative words as distinct from those identified as negative by traditional lexicons like the Harvard Psychosociological Dictionary can provide insights into firms' evolving strategies for managing machine readership. Figure 25 illustrates the noticeable shift around the event year.

221 Cao et al. (2025).
222 Loughran and McDonald (2011)

FIGURE 25 **SENTIMENT TREND AND MACHINE DOWNLOADS**



Note: This figure plots the difference between the frequency of Loughran and McDonald (2011) negative words and the benchmark of Harvard Psychosociological Dictionary in 10-K and 10-Q filings, separately for firms with high machine downloads with that of the low group. Both graphs are normalized to zero in 2010, one year before the publication of Loughran and McDonald (2011). The dotted lines represent the 95% confidence limits.

Source: Cao et al. (2023).

The findings reveal that companies whose filings have become sought after by machines were more proactive in adjusting their language following these events, using fewer 'negative' words as classified by the Loughran and McDonald lexicon, though not necessarily by the traditional Harvard dictionary. A second test confirms the prevalence of such a strategy, based on the launch of Bidirectional Encoder Representations from Transformers (BERT) by Google in 2018.[223] Once again, corporate disclosures became less negative as judged by BERT after its introduction. Furthermore, companies in which AI-powered institutional investors hold a larger share exhibited a greater tendency to modify their disclosure materials to reduce the likelihood of being perceived negatively by the newest development in AI technology.

Expansion of insider advantage based on non-inside information

The interplay between public and inside information has become increasingly complex with AI technology.[224] Corporate insiders, despite their privileged access to firm-specific information, may not be aware of the plans of activist investors – information that is material and non-public but originates from outside the firm, where the activists reside. Information about activist plans is highly relevant, as their interventions can significantly impact the valuation of targeted firms. Moreover, insiders have a dual stake in activist

223 BERT is an NLP model developed by Google AI in 2018. Unlike earlier NLP models that processed words sequentially, BERT captures the full context of a word by considering both preceding and succeeding words in a sentence.

224 Chabakauri et al. (2025).

campaigns: first, through the wealth effect, as they often hold significant equity in their firms; and second, through job security, given that executive turnover more than doubles following activist interventions.[225] Thus, insiders actively seek intelligence on activist plans, necessitating the use of sophisticated surveillance mechanisms.

AI technology has revolutionised this landscape by fuelling the growth of the 'market surveillance' industry. This industry, sometimes operated by stock exchanges themselves, employs advanced algorithms to analyse enormous datasets, including trading volumes, electronic order flows, and trade books, to infer the motives behind trades and detect ownership changes. AI's ability to process real-time data and uncover hidden patterns provides corporate insiders with unprecedented insights into potential activist activity.[226] Through these AI-driven tools, insiders can identify subtle shifts in trading behaviour that may indicate an activist's intentions long before such moves become apparent to the broader market. This technological advancement has transformed market monitoring from a reactive to a proactive process.

Interestingly, insiders' critical edge stems from a subtle combination of inside and outside information. While both insiders and sophisticated traders can observe the same order flows and trades, insiders possess more refined knowledge of firm fundamentals, such as earnings or sales growth. This privileged information enables them to distinguish trades driven by fundamentals from those motivated by other factors, such as activist intentions. For instance, insiders can integrate AI-derived predictions with firm-specific information, including earnings forecasts, attendance at investor relations events, or even digital footprints like IP addresses from corporate website visits.

This dual filtration of external and internal data allows insiders to identify and anticipate activist trades, positioning themselves to counter such moves even before public disclosure of activism. The role of AI in enhancing surveillance capabilities is another example of how technology is redirecting the flow of information and blurring the boundaries between public and private domains. By doing so, AI creates new forms of information asymmetry, where insiders leverage advanced tools to leverage public signals with their own privileged information, allowing them, in this setting, to navigate activism defence.

4.3.4 Equal rights, differential power

So far, we have discussed how AI technology can amplify information asymmetry through alternative data generated outside the firm or through public information. In both cases, the asymmetry arises not from insiders having privileged access to material nonpublic information but from unequal capacities among market participants to access, process, and interpret information. This disparity reflects an imbalance of power rather

225 See Brav et al. (2008) for a comprehensive analysis of how hedge fund activism impacts firm performance, valuation, executive compensation, and job security.

226 As of 2024, leading players in this market include Nasdaq IR Intelligence (which includes an "activism unit" with "surveillance analysts"), S&P Global's "Real-time visibility into the actions of activists", and a few specialised firms such as Diligent Market Intelligence, FIS, and Q4.

than a breach of equal rights, as all participants could have access to the same public disclosures or outside data. It is important to note that such information is both universal – available to anyone with the requisite technology – and unevenly accessible because of the expertise and resources needed for data collection and analysis. This dynamic introduces a new form of asymmetry among outsiders rather than between insiders and outsiders.

An event study reveals that analyst forecast accuracy improved significantly following the availability of alternative data about a firm's business.[227] Notably, this improvement depended on the analyst's affiliation with a brokerage equipped with robust AI capabilities (defined as firms employing personnel with AI-related education or relevant work experience). While this underscores the potential of alternative data to bridge the information gap between firms and investors, it also highlights the emergence of a new divide: clients of AI-enabled brokerages gain a distinct advantage over others, as advanced AI technology is essential to extract, process, and synthesise actionable insights from large, complex datasets.

Consequently, such information asymmetry cannot be remedied solely by regulations like Reg FD, which focus on ensuring equitable access to material information. This regulation addresses the selective release of material nonpublic information by publicly traded companies. It mandates that, when an issuer shares such information with specific recipients, such as securities professionals or security holders who may trade based on it, it must also make the information publicly available. The objective of Reg FD is to ensure comprehensive and equitable disclosure has been well-served, but the challenge lies in addressing the growing divide in the ability to leverage this information effectively, a gap widened by advancements in AI technologies.

This effect has repeatedly manifested in scenarios where technology disproportionately benefits certain market participants, such as high-frequency traders, at the expense of traditional traders. Over the last decade, stock exchanges have made vast amounts of market data publicly available, including detailed order-book information, trade data, and other metrics. While this transparency ostensibly levels the playing field, in practice it has enabled HFT firms to exploit their technological edge, processing and acting on this information faster than traditional traders via strategies such as order anticipation and latency arbitrage. The resulting environment has contributed to wider bid-ask spreads and less stable prices.[228] The integration of AI has shifted HFT from being purely speed-focused to a combination of speed, intelligence, and adaptability, further widening the gap between advanced and traditional traders.

227 Cao et al. (2024).
228 See evidence in Lewis (2014) and Coughlan and Orlov (2023).

This shift fundamentally alters the dynamics of information asymmetry. While alternative data democratise access to certain insights and while broader and stricter disclosure squeezes out more information from inside the firm, timely and effective utilisation is limited to those with the resources and expertise, especially an AI technological edge. This creates a dual-layered asymmetry: it narrows the gap between insiders and outsiders, but introduces disparities among outsiders themselves as only the most technologically advanced participants can fully capitalise on these data sources. The resulting competitive edge for firms and investors who excel in harnessing alternative data and processing disclosures reinforces the evolving role of technology as both a tool for democratisation and a barrier in modern markets.

This improvement depends, however, on the affiliation of the analyst with a brokerage that has developed adequate AI capacity. Because analysts serve as information intermediaries for external investors, this finding substantiates the notion that alternative data bridge the information gap between firms and their investors, but only with the help of the technology needed to extract and synthesise big data.

4.3.5 New information asymmetry: Policy implications

Technology divide: Old and new
Information asymmetry has been a fundamental feature of financial markets since their inception. Asset pricing depends on informed trading, which inherently assumes uneven access to information, and such asymmetry has always been driven by a resource divide. In the absence of trading on overt insider information, obtaining information not yet impounded into the stock price invariably requires resources for information acquisition such as research. However, the rapid advancement of AI technology is shifting this dynamic from an 'evolutionary' change to a 'revolutionary' one – requiring a fundamental rethinking rather than incremental adaptation. Traditional regulatory frameworks, such as Regulation FD, were designed to promote equal access to material nonpublic information originating within firms, primarily addressing the insider–outsider divide. Yet, these frameworks now face challenges posed by the complex layers of asymmetry introduced by AI-driven tools and the explosion of alternative data sources.

Overall, AI has the potential to level the playing field between insiders and outsiders – a longstanding point of contention in the functioning of securities markets. In this sense, AI technology represents a step forward in reducing insider advantages, potentially enhancing trust in financial markets and encouraging information acquisition that leads to more informative security prices. At the same time, however, AI introduces a more fragmented and uneven information environment among outsiders themselves. This emerging 'AI divide' may increase demand for portfolio delegation or passive investment products, as less resourceful investors become aware of the growing disparity in analytical capabilities ('differential power'), despite having equal access to the raw data and information as the latter is generated more and more outside the firm.

Redefining equal access in an AI-driven market

The growing differential capacity to process public information, driven by the adoption of advanced AI tools, allows certain market participants to extract insights faster and more accurately than others. While this disparity is inevitable, certain actions could help bridge the gap or prevent it from widening further. One actionable step is the mandating of machine-readable disclosures. Standardising all corporate filings in formats optimised for algorithmic processing – building on the implementation of eXtensible Business Reporting Language (XBRL) in 2009 by the SEC – would significantly reduce entry barriers for a wide range of investors and analysts.[229]

Regulatory agencies could establish centralised, publicly funded platforms for real-time data access and analysis, providing accessible insights to all market participants and helping mitigate technological disparities. AI can further support this democratising effort by enhancing XBRL through automated tagging, extracting insights from unstructured data, and integrating alternative data sources, enabling richer and more accurate financial analysis. Additionally, AI can transform XBRL into a real-time, dynamic tool for predictive insights, personalised reporting, and fraud detection, making financial data more actionable and universally accessible.

Fair use of alternative data

The proliferation of alternative data – ranging from satellite imagery to social media sentiment – has introduced a new dimension of information asymmetry, enabling well-resourced market participants with advanced AI capabilities to extract insights ahead of others. Simultaneously, firm insiders gain additional sources to inform their decision making or gauge stakeholder sentiment. Given that such data can often include material nonpublic information, regulators could establish ethical standards for alternative data extraction and aggregation to ensure responsible sourcing respects consumer privacy and proprietary business information. Firms could also be encouraged to disclose the sources of alternative data that act as inputs to their decision making.

The private market can be expected to play a crucial role in democratising access to alternative data. Competition among data technology firms, combined with the scale of the potential market and the diminishing marginal value of information as more investors become informed, naturally drives down the price investors expect to pay for the information extracted. This process fosters the development of widely accessible alternative data repositories, helping ensure that these resources enhance market efficiency without disproportionately favouring the most technologically advanced participants.

229 XBRL is a format that defines or 'tags' individual data items in financial statement disclosure in a standardised format. With data formatted in this manner, investors and analysts can download financial statement data from public filings directly into spreadsheets and other software tools, enabling quick analysis and comparison among multiple companies, reporting periods and industries.

Addressing algorithmic behaviour

AI-powered trading algorithms have accelerated the speed and complexity of financial markets and created new risks such as algorithmic collusion and price distortions, which are often unintentional from the perspective of human decision making. Collusion occurs when independent algorithms implicitly coordinate behaviour, leading to anti-competitive outcomes like price manipulation, while certain strategies exacerbate volatility or exploit arbitrage opportunities, destabilising markets. Regulatory responses should focus on ensuring that the benefits of algorithmic trading, such as improved liquidity and narrower bid-ask spreads,[230] are not undermined by systemic inefficiencies or unfair practices. This echoes the discussion in Chapter 3 of this report.

AI tools can also help policymakers in real-time surveillance to detect collusive or destabilising trading patterns, such as coordinated trades, crowded trades, or unusual liquidity withdrawals. Mandating transparency in algorithmic frameworks, including disclosures on optimisation goals and data sources, would allow regulators to better assess risks without compromising proprietary concerns. Furthermore, implementing 'speed bumps' in certain trading contexts – similar to measures adopted by IEX[231] – can reduce harmful strategies like latency arbitrage and improve price discovery by giving markets time to process fundamental information. These interventions would help balance the efficiency gains of AI-driven trading with the need for market stability and fairness.

Feedback effect

Firms' adjustment of disclosure to cater to machine readers offers an illustrative example of a 'feedback effect' from technology. While financial markets reflect firm fundamentals, market perception (which is now powered by AI) also influences managers' information sets and decision making.[232] Such a feedback effect is inevitable because the encoded rules governing machine learning are not entirely opaque (i.e., they are partially observable and could be reverse-engineered to varying degrees), and agents affected by these decisions may be tempted by incentives to manipulate the inputs to ML algorithms to achieve more favourable outcomes. This feedback effect can give rise to unexpected outcomes, including manipulation and collusion.[233] ML algorithms thus face the challenge of becoming 'manipulation-proof', i.e., they must anticipate the strategic behaviour of informed agents without directly observing it in training samples.[234]

230 Hendershott et al. (2011) examine algorithmic trading on the New York Stock Exchange (NYSE) during 2001-2005 and find that algorithmic trading improves market liquidity by reducing the bid-ask spread and increasing order-book depth. These improvements are attributed to algorithmic trading's ability to process information quickly and provide continuous liquidity, particularly in large and active stocks.

231 Examples of 'speed bumps' include Investors Exchange (IEX)'s introduction of a 350-microsecond delay on all incoming and outgoing order in 2016 to prevent latency arbitrage, where high-frequency traders use their speed advantage to profit from price movements before slower participants can react. The TSX Alpha Exchange, part of the Toronto Stock Exchange group, implemented a 1-2 millisecond delay on certain orders in 2015. This measure was aimed at deterring HFT strategies that exploit order flow while still allowing the market to remain efficient for most traders. See Woodward (2018) for a formal analysis.

232 For a comprehensive survey on feedback effects, see Goldstein (2023).

233 See examples given in Calvano et al. (2020a).

234 This challenge has been the subject of theoretical analysis, as seen in Björkegren et al. (2020) and Hennessy and Goodhart (2023).

Regulators also face the novel challenge that firms could deliberately craft disclosures to mislead machine readers while remaining technically compliant with requirements. Unlike traditional financial misrepresentation, where firms may omit or misstate material facts, this new form of strategic disclosure involves subtle language manipulation that exploits how AI and ML models interpret sentiment, tone, and context. Regulators should explore collaborations with academia and industry experts to design financial reporting guidelines that balance machine readability with truthful, investor-friendly transparency. There is a need to draw a distinction between legitimate adaptation to AI (e.g., improving clarity for machine readers) and deceptive optimisation (e.g., gaming AI to avoid negative sentiment classification on the receptor).

4.4 Financial contracting meets AI

4.4.1 AI-enhanced principal–agent contracting efficiency

Section 4.2 focused on contracting between a human principal and an AI agent, but AI also impacts the efficiency of principal–agent contracting among humans. Vives and Ye (2025a) provide the first theory model that connects advancements in AI with monitoring efficiency in the lending relationship. AI technology reduces distance friction, transforms soft information into accessible data, and facilitates real-time tracking. The contractual efficiency leads to increased lending volume, reduced impact of distance, more intense competition, and lower loan rates.

In the corporate finance setting, principals (such as firms or shareholders) delegate tasks to agents (ranging from senior managers to rank-and-file workers) who are responsible for producing goods, services, and for creating value for shareholders and stakeholders. However, agents invariably have conflicting incentives and take hidden actions based on and covered by their private information. Contracts in this setting must balance trade-offs between incentivising effort, sharing risk, and addressing challenges in performance measurement, all of which are influenced by the adoption of AI in workflows.

First, AI improves productivity, particularly in scenarios where it complements human labour, encouraging workers to work harder under given performance-based incentive schemes. With AI augmenting the marginal productivity of agents, firms and managers can expect higher effort inputs from employees, bringing effort alignment closer to first-best. For example, AI-driven tools can speed up data synthesising, optimise decision making, improve forecasting, or enhance task execution, inducing more work under the same incentive as agents expect to benefit more from heightened measurable outputs.[235]

235 There has been ample evidence on AI-enhanced productivity in a wide range of professions within and outside of finance, such as stock analysis (Cao et al., 2024), M&A (Han, 2024), floor trading (Brogaard and Zareei, 2022), video content creation (Jiang et al., 2025a), and legal work (Armour et al., 2020).

Second, AI already enhances performance monitoring and holds the potential to be even more effective at reducing noise in the measurement of agent effort and outcomes. The scale of working from home post-COVID would not have been possible in the absence of technologies for remote monitoring, task automation, and virtual collaboration. Through advanced data analytics and machine learning, AI provides more precise and timely signals about an agent's contribution, enabling higher-powered incentives without compromising risk-sharing in contracts. This reduces the conditional uncertainty faced by both parties, allowing principals to offer contracts that incentivise effort while mitigating the agent's exposure to uncontrollable risks. Improved monitoring strengthens accountability, ensuring that performance-based compensation reflects actual contributions rather than external factors.

Jiang et al. (2025b) map the above prediction to the reality of US workers by analysing 20 years of data from the American Time Use Survey (ATUS). They find that workers in occupations that become more exposed to AI technology tend to extend their workday, reduce leisure time (both proxies for higher effort), and earn higher salaries. Specific tests confirm these outcomes are linked to human–technology complementary, boosted productivity, and improved monitoring enabled by AI. However, the study also underscores that AI-driven improvements in contracting efficiency do not necessarily lead to significant gains in agent welfare, as measured by work-life balance and employee satisfaction reviews. The extent to which workers benefit depends heavily on their bargaining power and the competitive dynamics of labour and product markets. In highly competitive labour environments or industries with low worker leverage, the rents from AI-driven productivity and contracting efficiency gains are often captured by principals or passed on to consumers. Thus, while AI improves the technical efficiency of contracts, it does not necessarily inherently enhance agent wellbeing.

4.4.2 Smart contracts, dynamic information, and decentralisation

AI technology is set to transform financial contracting, with tools like AI-driven contract drafting and risk analysis offering tangible improvements. Two key developments in reshaping contracts are smart contracts and real-time contracting. Smart contracts, powered by AI and integrated with blockchain, automate the execution of terms when predefined conditions are met, reducing manual oversight and increasing transparency. These contracts are self-executing and minimise errors and disputes, which benefits financial transactions. At the same time, AI enables real-time contracts that adjust as conditions change, such as in response to market shifts or regulatory updates. This dynamism ensures that agreements remain aligned with current information and regulations.

Smart contracts are inherently decentralised because they operate on a distributed ledger system that eliminates the need for centralised intermediaries. This decentralisation ensures that contract terms and execution are validated and enforced by a network of independent nodes, rather than relying on a single authority or institution. From an economic perspective, this reduces transaction costs associated with monitoring and enforcement, as the blockchain provides an immutable and transparent record of all contract activity. Moreover, decentralisation enhances trust among parties, particularly in transactions involving asymmetric information, by ensuring that no single participant can manipulate the contract or its execution. This feature is particularly valuable in financial markets, where agency problems and counterparty risk are prevalent. Decentralised consensus has the potential to expand the contracting space by reducing the scope of non-contractible contingencies, a theme that occupies the core of the incomplete contract literature.[236]

Additionally, the transparency provided by blockchains makes many previously 'hidden' actions verifiable, thereby mitigating traditional moral hazard problems. For instance, in the model proposed by Cong et al. (2021), the 'effort' that crypto miners exert is effectively observable by tracking the frequency at which they solve mathematical puzzles that are an order of magnitude simpler than those used in proof-of-work.[237] Together, these advancements offer unprecedented automation, transparency, and adaptability to meet the dynamic needs of modern financial contracting, but raise new challenges as the decentralised model moves closer to a direct democracy governance model.

4.4.3 Smart contracts with AI: Implementation and commitment

Smart contracts meet AI

Smart contracts consist of computer code that automatically executes all or parts of an agreement and are stored in a blockchain system. An informative and comprehensive definition is that smart contracts provide "terms contingent on decentralised consensus that are tamper-proof and typically self-enforcing through automated execution".[238] The underlying technology leverages cryptographic algorithms and distributed consensus mechanisms, ensuring that transactions are tamper-proof and execute only when predefined conditions are met. This innovation has broad applications in finance, supply chain management, real estate, and beyond, offering efficiency, cost reduction, and enhanced trust in contracting and transactions.[239]

Smart contracts offer many advantages – such as speed, efficiency, transparency, accuracy, trust, security, and cost savings – all of which are relevant to corporate governance. First, by eliminating the need for intermediaries such as brokers and lawyers to validate signed legal contracts, smart contracts reduce the risk of third-party manipulation and minimise transaction costs. Second, smart contracts are designed to ensure that both parties fulfil their obligations, promoting trust and accountability among the parties involved, even in the absence of prior or expected future business and/or social interactions. By reducing

236 Hart (1995).
237 At the same time, Cong et al. (2023) warn of the possibility of concentrated power in a decentralised structure.
238 Cong and He (2019).
239 Several recent papers, notably Levi and Lipton (2018), Makarov and Schoar (2022), and John et al. (2023), have provided comprehensive reviews of the mechanics of smart contracts.

the need for costly verification or enforcement, smart contracts can eliminate certain contracting frictions in an automated and conflict-free way.[240] Third, smart contracts can also enhance transparency and accuracy by providing a tamper-proof record of all transactions on the blockchain, which can help prevent fraudulent or unethical behaviour. Finally, smart contracts can address a major issue in classic contracting, namely, strategic ex-post renegotiation. By enforcing rules and conditions based on prior agreements, smart contracts mitigate issues with adverse selection and moral hazard.

Smart contracts were conceived before and without direct help from AI, in pioneering work by Nick Szabo,[241] a computer scientist and legal scholar. They became practically feasible with the rise of blockchain technology, particularly through platforms like Ethereum, launched in 2015, which provided a decentralised and programmable environment for implementing these contracts. These early smart contracts operated based on predefined rules and conditions encoded in their programming, relying on the deterministic nature of blockchain networks to execute actions without intermediaries. They were rules-based and static, without any direct form of intelligence. In fact, Szabo described a vending machine as a primitive example of a smart contract: it takes in coins and dispenses a product without the need for an intermediary, automatically enforcing the terms of a simple contract. This analogy illustrates that these contracts are essentially more 'robotic' than 'smart'.[242]

The integration of AI into smart contracts and general contract management is a recent advancement, introducing capabilities such as adaptability, data-driven decision making, and natural language processing that extend beyond their original rules-based design.[243] By incorporating AI, smart contracts can process and respond to dynamic information, enhancing their functionality compared to traditional contracts. For instance, AI-powered oracles can provide real-time market data, such as commodity prices or currency exchange rates, to automatically trigger contract clauses. In a hedging contract tied to fuel prices, an AI-enabled smart contract could adjust payment terms based on daily price fluctuations without the need for manual intervention. Chapter 2 of this report provides a discussion on AI-empowered monitoring and collateral tracking. This evolution transforms smart contracts from static 'if-then' constructs into adaptive systems, resembling state-contingent contracts in economics, where outcomes depend on realised conditions.

Beyond real-time data, AI also improves the efficiency and robustness of smart contracts through predictive analytics. ML algorithms, for instance, can analyse historical trends to optimise contract parameters, such as setting strike prices in options contracts or adjusting risk-sharing arrangements in insurance agreements. These algorithms can

240 Harvey (2016); Cong and He (2019).
241 Szabo (1994).
242 See a similar argument made by Yermack (2017) in the context of blockchain governance.
243 See, for example, "Generative AI turns spotlight on contract management", Financial Times, 3 July 2024; "AI Smart Contracts: Exploring the Future of Blockchain-Based Automation",' MiEthereum, 17 May 2024.

also detect patterns indicative of potential fraud or default, allowing for pre-emptive contract modifications or automated penalties. By integrating AI, smart contracts address challenges common in contract theory – such as asymmetric information and moral hazard – by offering solutions that are not only more automated but also more aligned with real-world economic complexities.

Power and limitations of smart contracts

For reasons discussed in the previous section, smart contracts are better suited to situations where a strong ex-ante commitment takes precedence over an ex-post renegotiation and where discretion in implementation is not valued or may even be detrimental, given its potential to encourage strategic behaviour in crucial scenarios. Such a cross-sectional comparison has not been examined either theoretically or empirically in the burgeoning body of smart contract literature. Therefore, our arguments draw primarily from earlier studies within the traditional governance framework. Nonetheless, our focus on the inherent properties of smart contracts enables us to project both their capabilities and limitations in the contracting space.

In practice, it is often challenging to predict whether ex-post renegotiation will be beneficial or detrimental. It may prove advantageous when it provides updated information to fill gaps in incomplete contracting,[244] or when redistribution of risk becomes Pareto optimal after an agent has extended effort but before the outcome is fully realised.[245] On the other hand, the possibility of ex-post renegotiation incentivises moral hazard, particularly in the form of strategic default. Distinguishing between cash flow-triggered default and strategic default can be difficult, especially in a downturn or economic recession.[246] Economic theory suggests that even the most sophisticated code or platform cannot account for all possible contingencies.[247] In other words, many financial contracts, such as loan agreements, are inherently incomplete.[248] Standard smart contracts, which are 'hard-coded' and preclude renegotiation, may not be ideal for contractual situations where it is difficult to specify contingencies or where ex-post risk-sharing is desired.

The trade-off between mandatory and discretionary triggering of contractual terms presents an intriguing dynamic. As discussed in Section 4.3.1, algorithm-based decision making is transparent and, in some cases, may be reverse-engineered, potentially incentivising behaviour aimed at triggering the algorithm, often with unintended consequences. Smart contracts, due to their mechanical nature, are inherently limited in addressing contingencies involving 'feedback effects'. A relevant example is the design of contingent convertible bonds (CoCos) issued by banks. CoCos are debt instruments that are either written down or converted into equity upon the triggering of a nonviability

244 Hart and Moore (1988).
245 Fudenberg and Tirole (1991).
246 See Ganong and Noel (2023) for an example.
247 Aramonte et al. (2021).
248 Such a narrative is a recurring theme in classic contract theory work, including Coase (1937), Grossman and Hart (1986), Hart and Moore (1988) and Aghion and Bolton (1992).

condition. The triggering condition can be defined objectively, such as through market prices (no renegotiation which could be assigned to a smart contract), or left to the discretion of regulators (allowing potential renegotiation).[249] The co-existence of both types of triggers highlights the strengths and weaknesses of each approach as an enforcement mechanism for debt-to-equity conversion.

The discretion to trigger typically carries the risk of encouraging moral hazard through the possibility of renegotiation (even if implied). Consider the scenario of a bank approaching the nonviability threshold that would trigger CoCo conversion. If regulators retain discretion to trigger conversion, they may hesitate to act promptly, fearing the potential destabilisation of markets or the political consequences of effectively signalling that a bank is in trouble. This hesitation can encourage moral hazard. Banks may gamble for resurrection, taking on riskier strategies under the assumption that regulators, wary of the broader implications of triggering, will step in to protect them.

Conversely, mandatory triggering of CoCos, while eliminating the uncertainty of regulatory discretion, introduces its own set of problems, most notably the risk of a 'death spiral'.[250] For example, if CoCo conversion is automatically triggered based on a market price threshold such as through a smart contract, the anticipation of the breach may cause investors to offload equity tied to the bank ahead of the dilution. This selling pressure further depresses the bank's market value, hastening the breach and triggering conversion prematurely. Such a self-reinforcing feedback loop can quickly erode confidence in the bank and destabilise the broader financial system. Mechanical smart contracts in financial markets thus may exacerbate these dynamics, leading to outcomes where the instrument intended to stabilise the bank instead accelerates its decline.

Smart contracts involve several additional limitations, some of which are related directly to their technical features. For example, smart contracts inherently depend on programmers and are vulnerable to bugs and logical errors in code. Any errors could result in significant risks for all parties involved, leading some blockchain ventures to hire auditors to evaluate the integrity of their smart contracts.[251,252] There are several elements that involve explicit or implicit features of governance. First, it is challenging for decentralised governance mechanisms to reconcile competing objectives among multiple parties, resulting in systemic instability when facing governance disputes within blockchain communities, especially when critical decisions about protocol upgrades or changes to contract rules are involved. Such a limitation was responsible for the eventual collapse of the Terra-Luna system in May 2022.[253]

249 For a comprehensive overview of the CoCo market, as well as theoretical models and empirical evidence, see Avdjiev et al. (2020).
250 See the theory model by Sundaresan and Wang (2015) on the CoCo 'death spiral'.
251 Bourveau et al. (2023).
252 For example, in an incident involving the DeFi platform YAM Finance, the project's developers inserted a bug in the smart contract, thereby causing the collapse of the entire project, resulting in significant investor losses. Lehar and Parlour (2023) also describe a flash-loan attack that exploited a decentralized platform's coding mistake.
253 Liu et al. (2023).

Second, smart contracts operate primarily in digital, online environments but often interact with real-world, offline events and data. Smart contracts cannot retrieve data from external sources beyond the blockchain network, which is necessary with various real-world applications. For example, external weather data may be required by a smart contract that bases insurance payouts on weather conditions. They rely on trusted data sources or services ('oracles') to provide information about real-world events or conditions.[254] This exposes smart contracts to mundane governance issues in the offline world. Oracles, for example, can be manipulated through arbitrage (which utilises flash loans), especially via DeFi protocols.[255] Moreover, in transactions that involve physical assets, smart contracts must be integrated with conventional legal structures.

Third, while smart contracts themselves do not inherently encourage strategic behaviour such as collusion, certain conditions or factors within a smart contract system or ecosystem can create opportunities for collusion among participants. Some collusive behaviour is no different in nature from the strategic behaviour that extends from offline to the online world in somewhat different guises. For example, token concentration may encourage a small group of participants to collude to control governance processes by coordinating their votes to sway outcomes. Colluding oracle operators could provide false or manipulated data to trigger contract outcomes that benefit them financially. More importantly, some types of collusive behaviour have been encouraged precisely by the transparency that decentralised consensus provides, which would have been deterred by information asymmetry or lack of commitment. Cong and He (2019) show that smart contracts can encourage collusion among interested parties, as blockchain transparency and commitment help to sustain the collusive equilibrium. In DeFi applications, smart contracts could be exploited by colluding with participants to create cartels that manipulate token prices or execute arbitrage opportunities, potentially harming other users.

So far, only one study has produced empirical evidence of trade-offs under incomplete contracting.[256] Using staggered adoption of US state laws that open new opportunities for within-state use of blockchain technology, the authors find that firms that have used blockchain technology experience significantly improved performance. Of greatest relevance to blockchain technology, the authors show that such firms become less reliant on vertical integration, enter strategic alliances more frequently, and develop new customers without regard for geographical proximity.

254 Oracles could be application programming interfaces (APIs) or web pages. See Beaver Finance (2022) for various types of oracles

255 According to Chainalysis (2023), DeFi investors lost about $865 million in dozens of oracle attacks from 2020 to 2022. See Harvey et al. (2021) for additional examples of oracle attacks.

256 Chen et al. (2023).

4.4.4 Renegotiate with AI

The double-edged sword of contract renegotiation

AI has transformed contract management by automating many aspects of contracting, including drafting, execution, and monitoring. AI-driven contracts enable continuous and seamless implementation, dynamically adjusting to evolving conditions based on real-time data and predefined algorithms. These features reduce transaction costs, minimise disputes, and ensure smoother execution compared to traditional static agreements. However, despite these advancements, contracts are often incomplete by nature, as they cannot fully anticipate all future contingencies or shifts in priorities. As a result, there are moments when renegotiation becomes essential – whether to account for unforeseen circumstances, adapt to new information, or rebalance the interests of contracting parties. In these situations, AI's rigidity in following pre-programmed rules can pose challenges, as human intervention is often needed to revisit terms, address ambiguities, and negotiate updates that align with current realities.

Mortgage contracting exemplifies both the pros and cons of AI capabilities. AI can monitor repayment patterns, detect early signs of distress, and deter strategic defaults by borrowers by analysing borrower behaviour and financial patterns to identify those capable of repayment but choosing not to pay. Borrowers may opt for strategic default when the perceived cost of continuing payments outweighs the benefits, such as in cases of negative equity where the value of the property falls significantly below the mortgage balance, or when there is a prospect of contract modification in favour of the borrower when a foreclosure is costly to the lender. Through real-time monitoring of payment histories, spending habits, and external economic conditions, AI can flag anomalies that are indicative potential strategic default. Once these are identified, AI enables lenders to implement targeted interventions, such as reminders of legal consequences or tailored penalties that increase the cost of defaulting. This proactive approach discourages strategic behaviour by ensuring that the financial or reputational consequences of default outweigh the perceived benefits.

On the other hand, there are situations where mortgage modifications are desired or even necessary to prevent widespread and contagious defaults caused by adverse macroeconomic conditions. For example, during economic downturns or housing market crashes, a sharp decline in property values can leave many borrowers with negative equity, where even borrowers who can afford their payments may consider defaulting, while those genuinely struggling face increased financial pressure. If defaults become widespread under strict contract enforcement, they can trigger a downward spiral, further depressing housing prices, eroding lender balance sheets, ruining neighbourhoods, and destabilising the broader economy. Mortgage modifications, such as extending loan terms, reducing interest rates, or adjusting principal balances, can provide relief to borrowers, incentivising continued payments and breaking the cycle of contagion.

It is natural for AI to exhibit rigidity in executing pre-programmed rules and predefined contract terms, which can limit its ability to adapt to unexpected macroeconomic shocks or nuanced borrower circumstances. This rigidity ensures consistency and efficiency but can exacerbate systemic risks when widespread modifications are necessary, as AI lacks the discretion to evaluate broader economic conditions or societal implications. However, AI has the potential for flexibility when programmed to identify patterns indicative of economic stress and recommend targeted interventions. For example, AI can analyse data on borrower distress, housing market trends, and macroeconomic indicators to suggest tailored modifications, such as principal reductions or term extensions, which human decision makers can then assess and implement in a broader economic and social context.

AI adaptation in contract renegotiation

The A.I.A. Co project, sponsored by the European Union, is a pioneering initiative that addresses the impact of COVID-19 on commercial lease contracts and demonstrates how AI can facilitate efficient and equitable renegotiation in the face of unforeseen economic shocks. During the pandemic, many lessees experienced severe financial strain due to mandated shutdowns, necessitating adjustments to lease terms to prevent widespread defaults. The project utilised AI-driven predictive frameworks to assist judicial authorities and contracting parties in renegotiating terms, such as rent deferrals or reductions, tailored to dynamic economic conditions. This application aligns with economic theories of incomplete contracts, which emphasise that contracts, by nature, cannot account for all future contingencies and thus should balance incentives for investment and ex-post bargaining inefficiencies.[257]

A study analysing the A.I.A. project highlights AI's capacity to analyse granular data and predict the implications of contractual modifications introduces a structured and data-driven approach to renegotiation, filling gaps left by rigid contract designs.[258] By addressing these gaps, AI enhances contractual adaptability while maintaining the balance between efficiency and fairness, even in the face of systemic shocks.

4.4.5 Commitment and flexibility with AI: Policy implications

The integration of AI into financial contracting does not eliminate the fundamental trade-off between ex-ante commitment and ex-post flexibility. While AI enhances contracting efficiency through better monitoring, predictive analytics, and automated enforcement, its utility is maximised when paired with human oversight and dynamic adaptability.

257 Hart and Moore (1988).
258 Parton et al. (2022).

Predefined renegotiation triggers
AI systems should not only automate contract enforcement but also embed predefined triggers for renegotiation. These triggers could be tied to measurable macroeconomic or market-specific variables, such as housing price indices, interest rate shifts, or sector-wide employment trends. For instance, in mortgage contracts, AI could continuously monitor property value declines or regional economic downturns to flag when mass defaults become likely. By automating such flagging, AI can enable proactive renegotiation to prevent systemic risks, such as cascading defaults or liquidity crises. This approach aligns with theories of incomplete contracting, which advocate designing mechanisms to address contingencies that are mostly not due to moral hazard, which are hard to specify ex-ante but are critical for long-term efficiency.

Hybrid AI–human contracting
Although AI can analyse vast amounts of data and provide precise recommendations for contract adjustments, it still lacks the contextual judgement required to evaluate broader implications of these adjustments. Human decision makers are crucial for assessing the economic, legal, and social consequences of contract modifications that AI might overlook. For example, during a financial crisis, AI might suggest widespread mortgage renegotiation based solely on default risk reduction, but human oversight can weigh the trade-offs between lender solvency and borrower relief. Policymakers should encourage hybrid systems where AI-generated insights inform decisions, while human agents retain the discretion to consider broader societal impacts, such as neighbourhood stability or consumer confidence.

Transparency in AI contracting
Transparency in AI systems is critical to ensure trust and accountability in financial contracting. Contract makers should provide clear and accessible documentation of AI algorithms' decision-making frameworks, particularly regarding renegotiation triggers and the rationale behind suggested contract adjustments. For example, if an AI system flags a corporate loan for renegotiation, its decision should be accompanied by an explanation of the data patterns or economic conditions that trigger this action. This level of transparency mitigates disputes and fosters confidence among contracting parties. Moreover, it ensures that AI systems remain auditable, allowing regulators to assess compliance and render judgment in disputes.

Incentive-compatible flexibility
AI systems, like their human counterparts, should prioritise dynamic incentive compatibility, ensuring that contracts do not invite strategic behaviour from either party while adapting to changing circumstances. For example, in mortgage settings, AI could help identify borrowers at risk of default and recommend modified payment terms that maintain incentives for repayment. Simultaneously, contracts must safeguard against opportunistic behaviour, such as strategic default or misrepresentation of financial

distress, by ensuring that modifications remain contingent on verifiable criteria. This balance between flexibility and commitment reinforces the credibility of contracts while adapting to economic realities, embodying the principles of contract theory that emphasise aligning incentives across all parties.

4.5 CONCLUSION

The integration of finance and technology represents a dynamic and multifaceted transformation of corporate governance and market systems, solving some problems while creating others. This chapter examines how artificial intelligence, machine learning, and big data reshape the foundational issues associated with corporate finance and governance, including agency problems, information asymmetry, and incomplete contracting. AI introduces a new agency paradigm where systems act as oracles, agents, or sovereigns, complicating accountability and decision-making processes. While AI eliminates traditional moral hazards such as shirking, it creates risks by optimising programmed objectives that may conflict with broader welfare goals or long-term strategies. The proliferation of alternative data and AI-powered analytics exacerbates public information asymmetry, increasing inequalities in information processing and market power despite universal access to data. Meanwhile, blockchain-enabled smart contracts reduce enforcement costs but challenge flexibility in renegotiation, highlighting the tension between efficiency and adaptability.

As technology is integrated into and transforms financial systems, it levels some playing fields but generates new inequalities, demanding a re-evaluation of governance frameworks. This chapter underscores the need for economics and governance structures that harness the transformative power of technology while ensuring fairness, accountability, and efficiency. Scholars must now address not only 'finance research *with* technology' but also 'finance research *about* technology', navigating this rapidly evolving frontier effectively.

Discussions

5.1 DISCUSSION OF CHAPTER 2, "ARTIFICIAL INTELLIGENCE AND THE FINANCIAL SECTOR: TRANSFORMATIONS, CHALLENGES AND REGULATORY RESPONSES", BY DIANA BONFIM[259]

Suddenly, AI seems to be everywhere. The widespread and immediate adoption of ChatGPT in November 2023 made the world understand that AI is no longer a distant science-fiction concept, but something available to all of us, all the time, at our fingertips. For sure, the financial sector has not been as surprised as most of us by the power and pervasiveness of AI. As Chapter 2 of this report shows, AI has been used in the financial sector for a long time now. While the popular large language models (LLMs) are a recent feature, the financial sector has long been harnessing the power of the data it holds through the use of AI tools, such as advanced machine learning techniques.

Chapter 2 discusses the opportunities, challenges, and regulatory responses. The **opportunities** are tremendous and perhaps still hard to fully grasp. In a nutshell, AI allows banks to make significant advancements in their two core areas of expertise: screening and monitoring.[260] The financial intermediation role of banks is anchored on their ability to screen borrowers. Direct lending is inefficient if it relies on the ability of savers to efficiently channel their funds to the most creditworthy and profitable lending opportunities. Banks have scaled up based on their expertise on identifying these opportunities. While the human knowledge accumulated by loan officers still plays a role today, machines are indisputably more efficient at processing the vast amounts of data available today. This allows banks (and nonbanks) to optimise the trade-offs between risk and return more efficiently, identifying the borrowers that can contribute to maximising profits subject to risk preferences and regulatory limits. When data are available, AI can do better than humans at screening,[261] thereby contributing to decreasing exposure to credit risk in the financial system.

The power of AI is also important for monitoring. While screening mitigates the adverse selection problem in lending, monitoring mitigates the moral hazard that lenders are exposed to. Banks need to remain permanently vigilant about potential deteriorations of credit risk in their borrower pool. Historically, a close relationship between lenders and borrowers has been important for banks to be able to identify risks early on.[262] Collateral also plays a key role in this process, allowing for a better alignment of incentives between

259 These are the views of the author and not those of Banco de Portugal or the Eurosystem.
260 Carletti (2004); Mester et al. (2007); Stiglitz and Weiss (1988).
261 Gambacorta et al. (2024a).
262 Berger and Udell (1995); Dass and Massa (2011); Degryse and Ongena (2005); Kysucky and Norden (2016); Rajan (1992).

borrowers and lenders.[263] That said, collateral itself needs to be monitored, as expected losses critically depend on the ability to recover collateral value in case of default. Hence, both for collateralised and uncollateralised exposures, AI can help lenders to process all the information available to adequately monitor the risks that they are exposed to.

The chapter shows that the benefits of AI in finance are not exclusive to the private sector. Actually, central banks have been at the forefront of AI usage to fulfill their mandates. The chapter refers to many applications which help central banks to safeguard price stability and the stability of the financial system. Statisticians have been using AI tools to collect, process, and compile statistical information. Economists engaged in macroeconomic and financial analysis for monetary and macroprudential policy purposes use machine learning techniques and LLMs to identify current and future developments in the economy and the financial system, thereby contributing to better informed policymaking. Supervisors are also able to process large amounts of data to identify risks and vulnerabilities in financial institutions. Finally, the oversight of payment systems, relying on unsurmountable amounts of data, also benefits from the use of AI tools to detect anomalies and disruptions.

Overall, the benefits of AI in the financial system are already visible and meaningful, but these are likely the first steps of more to come.

Of course, new opportunities often also mean new **risks**. The larger the leap forwards in terms of innovation, the more acute and widespread the risks will be. Chapter 2 identifies several important challenges: bias and discrimination, legal risk, cybersecurity, market concentration, the dominant role of big techs, and risks for financial stability. As discussed in the chapter, these are old problems, but AI raises new challenges.

Over time, policymakers have made efforts to mitigate bias and discrimination, adopting consumer protection and fair lending practices. It could be argued that AI could be blind to human bias and discrimination. However, just as humans are sometimes not aware of their attitudes and bias, AI can also be a bit human here. This does not arise from the intelligence itself, but rather from the algorithms and the data used to feed the models.[264]

The chapter also discusses an important concern related to legal risk. The fact that AI models are generally a 'black box', aggravated by hallucination problems, may lead to increased litigation risks. Supporting and defending past decisions may not be as easy as if there is a human trace of decision making.

Concerns over cybersecurity are also unavoidable. Cybercrime itself is more and more reliant on AI, and AI systems may also be more vulnerable to attacks. However, AI also offers powerful cyber resilience opportunities, which should be harnessed by the financial sector.

263 Berger and Udell (1990); Stiglitz and Weiss (1981).
264 Fuster et al. (2022).

The chapter includes a detailed and insightful discussion of the AI supply chain. According to Gambacorta and Shreeti (2025), the AI supply chain is structured around five components: hardware, cloud computing, training data, foundation models, and AI applications. All of these require massive investments and scale. Often there is a winner-takes-all dimension in some parts of this supply chain, as scale, scope, and network economies are pervasive. This has led to extreme market concentration in some segments of the AI supply chain, which inevitably creates risks.

The rise of big techs cannot be dissociated from this. Big techs can lever on the loop created by the close links between data, cloud computing resources, and AI models. This raises concerns over possible abuse of market power and rent extraction from consumers.

The chapter also discusses risks to financial stability. AI models rely on similar datasets and financial institutions use similar models. This can increase volatility in financial markets and systemic risk in the system overall. On another dimension, there is uncertainty over the effects of AI in the economy. Almost for sure, there will be winners and losers as in any technological revolution. While exposure to the winners will have positive spillovers on the financial system, the opposite will happen to those exposed to the losers. Efforts to identify who these might be will be crucial to mitigate future losses.

The chapter ends by discussing how to **regulate** AI. There are trade-offs between the vertices of a complex triangle: stability, efficiency, and consumer protection. The chapter lays out some principles to regulate AI considering these trade-offs. These include common principles such as societal wellbeing, transparency, accountability, fairness, privacy protection, safety, human oversight, and robustness. There are three models around the world as to how to approach this: a market-driven approach in the United States, a state-driven approach in China, and a societal protection approach in Europe. However, as emphasised in the chapter, AI cannot be regulated within borders, and international cooperation is essential to ensure a proper balance between the benefits and risks of AI in finance.

The chapter is incredibly rich and encompassing. However, given the unknown unknowns on the implications of AI for the financial system, additional questions arise. How to balance the opportunities and risks in screening? What other new risks may emerge? How should we think about the regulation of the different types of risks that banks are exposed to? What are the implications of a changing geopolitical landscape?

How to balance the opportunities and risks in screening?

The chapter presents solid evidence that AI models offer greater accuracy in assessing credit risk, based on existing literature.[265] However, greater accuracy is only possible when there are data to train and feed the models. For instance, approval and pricing decisions on consumer loans usually rely on standardised data. Traditionally, applicants

265 See, for example, Gambacorta et al. (2024a).

provide information on their financial and professional situation, which is processed through a credit scoring model. AI can scale this up in several ways, using more powerful credit risk models or exploring a broader set of data. For instance, online behaviours can be exploited to create more accurate risk profiles than those based solely on self-reported information.[266] Financing mortgages is also quite standardised and model-based. For large, well-established firms, with a wealth of data available, AI can surely lead to enhancements on loan approval and pricing.

But is the same true for segments of the credit market where asymmetric information is more acute? The debate on the role of relationship- versus transaction-based lending is not new,[267] but AI brings it to a different level. Given the potentially large gains offered by AI on loans that can rely on standardised information, how willing will banks be to finance small and medium enterprises (SMEs)? How costly will this be? The decision to finance smaller firms, which are much more informationally opaque, usually requires the collection of soft information through a close relationship between the bank and the firm. This human interaction may become more expensive in relative terms, making SMEs' access to finance even more challenging. This can have detrimental effects on employment and investment, hurting economic growth.[268] Start-ups may be even more vulnerable to this. The knowledge acquired by banks in lending to newly established firms can be very important in shaping their growth and success.[269] If it becomes even more costly for banks to take these risks (and for firms to be able to access funding), we may risk that AI actually contributes to a less innovative and dynamic economy. That said, there are encouraging results on the ability of AI to also improve screening of start-ups.[270] The use of non-traditional data and payments information can also be helpful for more informationally opaque firms.

A related concern is that the data-driven nature of AI screening decisions may make it less able to make the right decisions after major disruptions or structural shifts. For instance, the COVID-19 pandemic rendered most credit risk models useless for decisions during that period. Firms were exposed to a sudden liquidity shock and their creditworthiness could not be assessed solely based on their past financial performance. The key question was to identify which firms would be viable going forward if they were provided with liquidity to withstand temporary lockdowns and disruptions in supply chains. In a fast-changing world, past information is not necessarily the best anchor for decisions, even if the non-linearity embedded in machine learning algorithms can offer flexibility.[271] In recessions and crisis periods, the human knowledge that forms the basis of relationship lending has proven to be critical to mitigate the impact of shocks and help the economy recover.[272]

266 Berg et al. (2025).
267 Rajan (1992).
268 Bonfim et al. (2023).
269 Bonfim et al. (2025).
270 Lyonnet and Stern (2024).
271 Gambacorta et al. (2024a).
272 Bolton et al. (2006).

An interesting point is that AI can change the trend of increasing reliance on collateral for access to funding.[273] This means that firms and industries for which collateral is scarce (intangibles) can overcome difficulties in access to credit. While this is good news, there are also trade-offs. Instead of relying on collateral, lenders will rely more on data. This may help firms and industries with positive past performance but without collateralisable assets to improve their access to credit. But AI will tend to focus on past winners. The industries that were more successful in the past are more likely to have better credit scores in data-driven models. Will this lead to an economy that is less prone to innovation, where startups will find it even harder to have access to finance? Will the economy become less flexible and able to adapt to shocks, due to the reliance on algorithms for decision making?

What other new risks may emerge?

The chapter does an excellent job of discussing how old risks may take new forms due to the role of AI in finance. But there will inevitably be new problems. Let me highlight two.

First, as shown in Figure 9 of Chapter 2, banks bank on trust. This is at the core of the banking business. Trust was conquered through decades of learning from crises and shocks, which led to the fine-tuning over time of a complex regulatory and supervisory institutional framework. In contrast, humans seem less willing to trust generative AI services, notably in banking (as shown in the same figure). Will the perception that AI may be taking over many decisions and processes in the banking industry affect trust in banks?

Second, AI might heighten systemic risk and interconnectedness on a scale never seen before. Actually, when we think about these two concepts, we consider mainly financial connections, such as those that exacerbated contagion during the global financial crisis. But AI and technology in finance create a new type of systemic risk, which is more operational than financial. Financial institutions increasingly rely on a common set of models, many of them sourced from common consultants and suppliers, leading to increased 'groupthink' in financial decision making. Moreover, data have become one of the most valuable assets.[274] Banks have traditionally relied on a wealth of data, but they have not always been at the forefront of the ability to process it. Agents who concentrate the ability to hold and process data may be able to establish information monopolies, with important implications for competition and, ultimately, financial stability.

273 Rampini and Viswanathan (2013); Beck et al. (2023).
274 Farboodi and Veldkamp (2023).

How should we think about the regulation of the different types of risks that banks are exposed to?

Arguably, AI is likely to help in managing and mitigating risks in the financial sector. The benefits on credit risk are already visible, as shown extensively in the chapter. Market risk and liquidity risk may also benefit from the use of AI in measuring and adapting to risks. But what about operational risk? Data on operational risk are not as pervasive, as it refers mostly to rare events. This means that AI is less able to leverage on the power of data to make meaningful contributions. At the same time, operational risk in itself has become a more important source of risk for the financial system, with cybersecurity taking an increasingly prominent role in banks' concerns, compounded by increasing geopolitical risks and risks stemming from climate change. The systemic risks posed by the use of common models and AI tools add to this.

While AI can nevertheless help to improve the ability to measure and manage operational risk, it might be relevant to reflect on the balance of risks and regulation in the financial system. Regulation is designed to deal mostly with the traditional risks, notably on credit. The regulation on operational risk is more blunt and less targeted. A discussion on a move towards better regulation and oversight of these operational risks may be warranted.

What are the implications of a changing geopolitical landscape?

The chapter ends with a discussion of the principles for regulation of AI in the financial system, which cannot be done without strong international cooperation. This would always be a challenge, but the current geopolitical landscape makes it even more complex.

By asking ChatGPT "*As a GenAI agent, what are your views on the risks stemming from the geopolitical situation and the use of AI in the financial system?*", I obtained the following answer: "*Geopolitical tensions can amplify the risks of AI misuse in the financial system, particularly through cyberattacks, economic manipulation, or disinformation campaigns targeting markets. The increasing reliance on AI-driven trading and risk assessment could create systemic vulnerabilities if adversarial actors exploit model weaknesses or data dependencies. To mitigate these threats, strong international cooperation, robust regulatory oversight, and transparent AI governance frameworks are essential.*" It seems that we, humans and machines, all agree on the need for strong international cooperation. The decisions on how to implement this are still in the hands of humans.

Summing up, this chapter is a remarkable contribution to better understand a novel and complex problem, with many interesting and novel insights. AI brings many gains to the financial sector. But, inevitably, there are also risks, some of them unforeseeable. This means that regulation has a role to play and that it must be anchored on international cooperation. This is perhaps the greatest risk.

5.2 DISCUSSION OF CHAPTER 2, "ARTIFICIAL INTELLIGENCE AND THE FINANCIAL SECTOR: TRANSFORMATIONS, CHALLENGES, AND REGULATORY RESPONSES", BY RONIT GHOSE

Artificial intelligence (AI) is driving a powerful flywheel of innovation, with advanced systems creating better AI, which in turn accelerates breakthroughs in computing and capabilities at unprecedented pace.

AI could be the general-purpose technology (GPT) of the 2020s and 2030s, profoundly transforming finance and money.

GPTs transform entire economies, altering how we live and work. They create opportunities for growth and innovation, often enhancing quality of life. However, they also disrupt existing systems, creating short-term losers alongside long-term gains.

The steam engine commoditised production and physical movement, powering the industrial revolution. More recently, the internet revolutionised communication and information exchange. Similarly, AI may augment or even substitute for human intelligence, including analysis, decision making, and content creation.

Prior technological cycles have eliminated old jobs and firms, while creating new ones. AI will repeat this cycle – potentially at a faster pace. The challenge that AI creates is that it may create a skew of smaller numbers of winners (turbocharged professionals and AI startup entrepreneurs) and a larger number of losers (median workers, incumbent firms).

The rise of generative AI (GenAI), exemplified by ChatGPT's launch in November 2022, marked a turning point. GenAI introduced an intuitive user interface, making AI accessible to the masses and sparking widespread interest among consumers and decision makers.

I believe GenAI has revolutionary potential in financial services because the sector is information-rich. Data are its raw material. In many respects, finance is the perfect sector for the application of AI.

How will AI be used in finance?

For years, AI – particularly machine learning (ML) – has been used in finance on structured data and for quantitative tasks. Today, AI is primarily applied to risk management and pricing. GenAI will expand these use cases.

In the short to medium term, we can expect the biggest impact at incumbent financial institutions to be on internal-facing tasks and improvements in productivity rather than lots of new products. Incumbents will focus on improvements in areas such as software and coding, transaction monitoring and compliance, and so on.

A lot of bank functions, such as credit underwriting, algorithmic trading, portfolio construction, and transaction monitoring, already utilise AI/deep learning. GenAI will create new opportunities beyond productivity improvements but some of the blue-sky work – newer products and services, bots using tokenised money, and decentralised AI – will likely take time to build and be rolled out to market.

Below is an overview of the main AI/GenAI use cases in finance:

Coding and software: Large banks employ thousands of software developers – often 15–25% of their workforce. AI can streamline coding by automating repetitive tasks, optimising code, and accelerating development cycles.

Search and summarisation: The financial services sector is characterised by its data- and document-intensive nature. AI can sift through vast datasets, distill pertinent information, and deliver concise summaries that can be used as input for faster decision making and executing next actions.

Transaction monitoring, compliance, and conduct: AI-powered systems can excel in monitoring external and internal conduct. By continuously analysing transactions and behaviour, and detecting anomalies in real-time, AI mitigates risks, ensures regulatory adherence, and minimizes fraudulent activities.

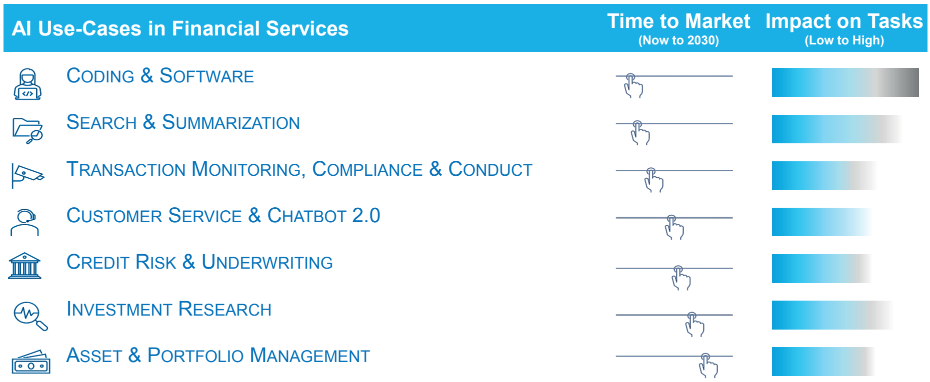
Customer service (Chatbot 2.0): AI-powered chatbots can deliver personalised, 24x7 customer support, resolving queries promptly with human-like interactions.

Credit risk and underwriting: AI can analyse diverse datasets of traditional and non-traditional data to assess credit risk and facilitate underwriting processes with greater accuracy and speed compared to traditional models.

Investment research: Fundamental research involves a lot of search and summarisation of information, datasets, and generation of text and charts. GenAI can bring time and cost efficiencies by automating information search and retrieval tasks.

Asset and portfolio management: AI can identify investment opportunities, optimise asset allocations, and personalise portfolios at scale, enabling advisors and portfolio managers to focus on high-value activities like client engagement and alpha generation.

SUMMARY OF GENERATIVE AI USE CASES IN FINANCE, ESTIMATED WIDER ADOPTION, AND POTENTIAL IMPACT ON TASKS[275]



Source: Citi GPS (2024).

The next frontier: Your own digital Jarvis

The rise of autonomous agents could usher in an era where people increasingly rely on AI bots to manage their lives. Instead of prompting LLMs with one-line instructions, we can rely on intelligent digital assistants embedded with advanced capabilities.

These bots, equipped with sophisticated algorithms and access to vast amounts of data, will negotiate with counterparties to secure the best possible deals for users. This shift will not only streamline services but also ensure decisions are made with a level of precision and foresight that humans may not have.

Consumers may no longer need to spend time gathering information, comparing different items, or executing tasks. Instead, they are likely to focus on yes/no/switch decisions, while AI handles the legwork.

Consider mortgage renewals: rather than manually searching for rates, an AI agent could track mortgage expiration dates, analyse market conditions, compare loan products, negotiate with lenders for optimal terms, and automate paperwork. Initially, users may require AI agents to seek human confirmation before executing financial decisions, but as trust grows, more tasks can be handled autonomously.[276]

In this new paradigm, the critical decision for consumers will be selecting the right bot. Choosing bot-powered advisers, much like choosing human personal finance advisers, will become a key task. But who will bots work for – big techs, trusted institutions such as banks, or startups?

275 See the Citi GPS report, *AI in Finance: Bot, Bank & Beyond* (Citi GPS, 2024).
276 See the Citi GPS report, *Agentic AI: Finance & the 'Do It For Me' Economy* (Citi GPS, 2025).

Leading banks will likely opt to provide their own AI-powered services. But big techs may have a competitive advantage in terms of being digitally native and have faster go-to-market speed. In some markets, they may also have stronger consumer brands.

Smaller firms and startups may see their growth turbocharged by the growth of agents. AI may dramatically improve their reach. But will consumers want to spend much time thinking about which agent to use? We are likely to default to known and trusted brands.

New startups and the future of work

Throughout history, general-purpose technologies – the steam engine, electricity, the internet – have reshaped economies. Will AI spark another wave of creative destruction or a retrenchment of current oligarchic capitalism?

In previous cycles, technology revolutions led to the rise of new firms. The steam engine era saw large factories and industrial businesses replace artisanal firms. Electricity further transformed mass production and urbanisation. The First and Second Industrial Revolutions ushered in a new era of banks and financial firms.

Similarly, the internet era from the 1990s onwards led to the emergence of new e-commerce, media and fintech firms, while traditional players struggled to adapt. AI could follow a similar pattern. The combination of cloud, AI, and agents could trigger a revolution in the delivery of digital services, including finance.

Just as cloud computing moved the economics of entrepreneurship from capex to opex, agentic AI moves them from employee payroll to software subscription. Digital banks and regulated fintechs will be able to leverage their tech infrastructure and licenses to grow even faster than before. Meanwhile, larger financial firms may move with caution due to cultural, regulatory, and technological constraints.

AI is also transforming knowledge work at scale. Much like the printing press democratised information, AI is decentralising access to knowledge and creativity. Wider adoption of AI will likely bring productivity gains to the finance sector by automating and augmenting current tasks.

Repetitive tasks – data entry, reporting, compliance – are ripe for automation. As AI advances, autonomous agents can enable a single professional or a small team to achieve what previously required entire departments. More can be done with less people, leading to less headcount.

New roles will emerge too, especially focused on AI development, ethics, governance and oversight. Industries centred on training, fine-tuning, and auditing AI systems will proliferate, making data curation and quality assurance critical growth areas.

Navigating AI's potential pitfalls

While AI promises advancements for finance and economic productivity, considerable challenges also exist. The dark side of AI is not just a speculative dystopia. Algorithms can reinforce existing social inequalities; GenAI can spread misinformation, as well as be used for fraud and obscure human accountability.

Bias and discrimination: AI is only as good as the data it is trained on. For example, AI-powered credit-decision models could inadvertently favour certain sections of society, potentially resulting in an increase in social and economic disparities. Addressing algorithmic bias requires attention to data collection, pre-processing, and algorithmic design to ensure fairness and inclusivity.

Lack of transparency: AI systems can be opaque, obscuring the decision-making process and underlying logic. This poses a challenge as humans cannot comprehend how AI arrived at its conclusion, leading to distrust and resistance to adoption. Explicability of AI models is a growing area of concern, especially since enterprise and societal adoption of AI is growing.

Misinformation and manipulation: AI-generated content could contribute to the spread of false information or misleading content at scale, manipulating public opinion, undermining trust in reliable sources, and leading to confusion in society. A particular risk of AI is its ability to tell stories that resonate with an individual's pre-existing beliefs and viewpoints, reinforcing echo chambers and ideological silos.

Hallucination: GenAI is prone to 'hallucinate' (i.e., to generate information not based on real data) but present this as fact. Hallucination is often caused by training data limitations and the model's probabilistic nature. When the model encounters a prompt for which it has insufficient or no training data, it creates a plausible but incorrect response.

Market concentration: AI models used by consumers, businesses, and governments may be provided by a small number of private companies, largely based in the United States or China. AI supplier dominance may echo the concentration and single point of failure risks that already exist in cloud computing.

Talent pipelines and entry-level job concerns: Many repetitive jobs are potentially at risk of automation, intensifying job polarisation and potentially economic inequality. Just as offshoring and outsourcing of white-collar jobs in the past few decades raised questions about talent pipelines and entry-level jobs, AI will reframe the conversation, but this time globally. The nature of roles in shared services hubs will change.

Amplified social engineering: The Global Anti-Scam Alliance estimated $1 trillion in losses to scams in 2023. The UK's National Crime Agency ranked scams as the number one form of crime. AI introduces new cybersecurity risks as bad actors leverage ts capabilities to launch more sophisticated and personalised cyberattacks. Advanced AI algorithms could be exploited by malicious actors to evade traditional detection methods using GenAI-driven malware that adapts its behaviour based on the target's defences. Already, half of all online traffic consists of bot activity, much of it malicious.[277] As AI advances, we could see a surge in sophisticated AI-driven scams.

AI-generated deepfake scams: AI can generate human audio, images, or text that is not real. Today, AI-generated audio has already crossed that uncanny valley in terms of being imperceptible to the human ear; AI-generated video is quickly improving. Agentic AI will lead to the mass production and distribution of deepfakes. In finance, this could be used to manipulate transactions, create synthetic identity fraud, and automate scamming. In 2025, we can expect audio- and video-based impersonations to increase due to democratisation of GenAI, enabling bad actors to impersonate known individuals such as colleagues, customers, family members, politicians, and actors. From banks' perspective, the risk and implications from threats of GenAI-based impersonations are tangible today and will accelerate this year.

5.3 DISCUSSION OF CHAPTER 3, "AI'S IMPACT ON FINANCE: RESHAPING INFORMATION AND ITS CONSEQUENCES", BY ROBIN L. LUMSDAINE

I thoroughly enjoyed reading this chapter. Starting with the premise that much of finance is information-driven, it argues that the combination of AI and data abundance are transforming the discipline through increased efficiency in processing information and greater transparency in the sharing of that information. It then systematically discusses how various financial market participants (market makers, brokers, asset managers) stand to benefit from this 'AI revolution', before warning about certain frictions that might inhibit the speed of this transformation. Among the frictions highlighted are the production of information that is of low social value, a potential increase in both information and, hence, market power asymmetry, and increased market fragility due to the black box nature of most AI models.

Evolution or revolution?

The chapter describes the increasing use of AI to process and extract information from increasingly large amounts of data ('data abundance') as an "evolution" or "revolution" interchangeably. This distinction matters, not as a semantic quibble but in terms of how we think about the ability of those affected (humans, firms, organisations, regulations) to adapt. When we think of a revolution, we think of sudden, rapid, disruptive change[278] – in an economic context it is a change that alters the future path of economic growth and

277 Imperva (2024).
278 For example, "a rapid, fundamental transformation" (Wikipedia); "a sudden, extreme, or complete change in the way people live, work, etc." (Brittannica),

development. When we think of evolution, we think of a more gradual transformation that occurs over a prolonged period, such that change is often only recognised in hindsight.[279] An evolution is a permanent change; a revolution is rare but episodic – a perceived permanent change until the next one comes along.

The implications of this change for finance are not yet clear. There is an important sentence in the second paragraph of the chapter: "*Conversely, if these technologies degrade the quality of financial information, it would warrant concern and potentially justify policy intervention*". This gives rise to a key question when thinking about the extent to which the confluence of AI and data abundance will benefit the financial sector. Does data abundance improve the quality of information? Is more always better?

When I think about the amount of information that is available now versus when I began my career, it truly is an exciting time of immense data abundance. But when thinking about processing that information in a meaningful way – for example, in order to understand the complexity of today's financial markets, or to forecast asset prices – that excitement quickly turns to information overload.

Signal-to-noise

The chapter starts with what it describes as "*three pillars of the big data revolution in finance*" – data abundance, techniques to convert these data to decisions, and the increase in computing power – and argues that "*these factors reduce the cost of producing information*". It goes on to describe how these increases in information lead to more precise financial market signals, and hence, improved decision making. But I might argue that it is not the cost of information production that is reduced, but the cost of information *analysis*, and that whether that analysis results in the production of additional information, or just synthesised or repackaged existing information, remains to be seen.

I found the chapter to be somewhat silent on the question of information quality, despite its discussion of potential 'market failures' or frictions that inhibit envisioned efficiency gains, resulting in potential destabilising effects. There are indications that information quality matters, such as the reference to the possibility of information with low social value. There is no question that AI has enabled the processing of vast quantities of information, beyond what humans (or even machines previously) could process.[280] But more is not necessarily better; it is not just the amount of information that matters, but the quality. Specifically, it is important to ensure that any added information increases our signal-to-noise ratio, or at least does not reduce it. Now, many would argue that the

279 For example, "A gradual process in which something changes…" (American Heritage Dictionary); "a gradual process of change and development" (Cambridge Dictionary); "a process of gradual change that takes place over many generations" (Collins Dictionary).

280 In an ad hoc poll of participants in the "AI Frontiers in Finance" seminar by Oliver Giesecke on "Deep Learning for Corporate Bonds" on 11 March 2025, conducted by seminar organiser Sascha Steffen at the start of the seminar, 55% of respondents chose "more efficient processing of large data sets" as the "biggest advantage of using deep learning techniques in bond investing".

scope for finding an important nugget of information is still higher the more information one processes. This is certainly true if the information we are looking for is not actually in the original information set we were searching. But I would argue that the benefit of more information depends on how likely it is that we will find that nugget in the first place. Finding a needle in a haystack is hard enough, without there being more hay to sift through. So, it is important to step back and consider what added value is gained. Arguably, if our signal-to-noise ratio is already high (i.e., we have a lot of signal), the marginal improvement from processing additional information is less than if we are still searching for signal. Put another way, at any given point of time, what we hope for is that we are trading on signal, not noise.

Traders trade

Now let's consider aggregating up those points in time into a typical trading day. When I was in the private sector, one of the things I observed is that traders are constantly staring at their screens – looking at prices, waiting for an opportunity to trade. I've wondered, therefore, whether efficient processing of information truly is good for the functioning of markets. Specifically, if we split the available time in a day into time containing information versus time containing no information, faster processing/incorporation of information would mean that a larger proportion of the day contains no information! So, if we accept the fact that traders trade, this means that a greater proportion of trades will be based on noise.

Past performance...

We all know the adage, "past performance is not necessarily an indicator of future returns". Existing AI models are all trained with past information. But the chapter notes that "*Brogaard and Zareei (2023) use ML algorithms to find profitable trading rules based on past returns*." How can this be?

The advent of generative techniques has increased the hope that one day AI will be able to deliver the kind of out-of-the-box thinking that a human can, leading to superior returns. But just because information is new to us does not mean it is really new. This reminds me of some of the fundamental questions that scientists have been grappling with for years – can something (be it energy, or matter) be generated out of nothing?

The answer is partially yes, if we think carefully about the information set that decisions are based on. The extent to which markets are efficient rests fundamentally on what information set is being considered. Asset price researchers might discuss this in terms of whether known factors still span the space of an information set that is increasingly growing or whether AI-generated information requires the existence of one or more new factors; market participants might similarly discuss whether AI-generated information produces alpha.

The risk of the information asymmetries that Section 3 notes, however, gives rise to some participants having 'pseudo proprietary' information. In that sense, the ability of some participants to process (but not necessarily produce additional) information still could give them an advantage, regardless of the quality of that information, particularly when coupled with the ability to make large, market-moving trades based on it.

Because most algorithms are largely trained on existing/past information, it is hard to see how new information is being created. And there is increasing evidence that machines do not extrapolate well. A footnote in the chapter seems to acknowledge this, noting that AI algorithms tend not to perform well when confronted with situations that differ a lot from the underlying data they were trained on.

Correlated information/persuasion bias

The risk that AI-generated information will be perceived as new when it is in fact just repackaged, or at most highly correlated with known information, brings to mind the seminal paper on persuasion bias by DeMarzo et al. (2003). Their theoretical model highlights the idea that people hearing the same information from multiple sources fail to account for the fact that it may have originated from the same, single source, that is, they do not adequately adjust for correlated information.[281] This result has been demonstrated empirically in numerous studies.[282]

It will be interesting to see whether AI-generated information suffers from similar bias. In principle, an AI algorithm can be programmed to correctly account for information correlation, assuming such correlation can be accurately measured. But the same might not be true for decision makers using greater amounts of such information in an increasingly data-abundant environment.

In some ways correlation is helpful. Predictability is predicated on some ability to draw on past experience. But too much correlation can lead to sticky reactions. Reinforcement learning attempts to balance that inherent tension. It is possible that an algorithm can more reliably achieve that balance than a human.

Other behavioural biases

Two other behavioural patterns are worth mentioning. The chapter cites a number of reasons that data abundance may shift focus away from long-term information and prediction towards a more short-term perspective. This potential rotation of the information curve suggests an inherent (or heightened?) present bias that may or may not lead to optimal decision making.

281 This behavioral pattern has been referred to as 'correlation neglect' in subsequent literature (e.g., Levy and Razon, 2015; Bohren, 2016; Levy et al., 2022).
282 See, for example, Brandts et al. (2015) and Enke and Zimmermann (2019).

At the beginning of this discussion, I mentioned the possibility of information overload. The chapter acknowledges an inherent trade-off that people need to make in the face of increasing amounts of information, noting that, *"[i]n making these choices, intermediaries weight their private benefit of allocating additional informational capacity to one type of information against their private cost of being less informed about another type."* Taking this limited attention capacity into account, the processing speed that AI affords may indeed help to more quickly identify a desired outcome.[283]

Perceptions versus reality

Much of my research has focused on the distinction between people's perceptions and data-based information, partly because in many circumstances, it is not necessarily the latter that drives decisions but the former.[284] Perceptions can lead to a self-fulfilling reality. In that sense, the question is not really whether AI produces new information or not, but rather whether people *believe* it does.[285]

Perception is also the result of information processing. As more and more of this task is ceded to AI, and decisions are based on AI input, we run the risk that it is AI's perceptions that are reflected in market pricing and not our own. Many of us have already encountered something like this with students using AI tools in their writing. It becomes hard to tell whose perceptions the writing reflects. The same will be the case when we consider market prices – not only will they reflect information but also the perception of that information as interpreted by AI.

A criticism that is often levelled at those who would seek to involve algorithms in decision-making is that many algorithms are a 'black box', i.e., something that is not well understood. The chapter similarly makes the point that the increasing use of AI to process information creates additional (operational) risk, due to humans' lack of ability to fully comprehend the algorithm's decision-making process. Given our human nature to distrust things we don't fully understand (a learned behaviour from early childhood),[286] AI then has the potential to create a vicious and never-ending cycle, whereby we build increasingly complex models to process increasing amounts of information, perhaps generating more information that then requires more complexity to try to understand. But others will point to real-world examples where we are willing to accept things we don't understand, in part because they have been validated or demonstrated to be reliable. Many of us use modes of transportation that we don't fully understand. So how different is our ability to comprehend an algorithm's decision-making process from our ability to fully comprehend that of a human?

283 For discussion of limited attention, see, for example, Hirshleifer and Teoh (2003) and Gabaix et al. (2006).
284 Bassett and Lumsdaine (2001); Lumsdaine and Potter van Loon (2018).
285 See also Greenwich Associates (2017).
286 Reimann et al. (2017) find that while trust is a heritable trait, distrust is learned.

Preference for humans

We seem to have a clear preference for humans, particularly when it comes to decisions. The chapter emphasises a "*clear division of labour: machines assess the likelihood of various outcomes and humans use this assessment to make decisions*" but notes that this "*division of tasks is becoming blurred*", particularly with developments in reinforcement learning.[287] Yet there are numerous reasons to think that AI might actually outperform, as the chapter points out. For one thing, AI might display fewer cognitive biases than a human. In addition, citing evidence that the combination of human and machine-learning (ML) forecasts is better than ML alone, the chapter draws the inference that "humans possess unique abilities in forming forecasts". But there is a large literature documenting that combining statistically generated forecasts results in improved accuracy over a single forecast, so it is not clear that the gains are specific to the inclusion of a human forecast.[288]

Our preference for humans is illustrated in the fact that most of our current regulations are based on human responses (e.g., the demonstration of communication or intent in determining liability). Is this not in itself a behavioural bias? Why should we prefer human decision makers? Humans have a level of accountability, whereas algorithms do not. Current regulatory efforts are attempting to navigate this discrepancy, for example by placing accountability on the providers of AI models (the EU AI Act) and emphasising the need for appropriate controls and governance, with sufficient human oversight – the 'human-in-the-loop' requirement.

Humans are the ultimate black box

It's worth emphasising that humans are the ultimate black box. And yet we are comfortable with their input. So why might this be? One key reason is trust. There is a large literature documenting that the way people develop and maintain trust in humans and machines differs.[289] Additionally, errors are more easily forgiven when made by a human while trust is more rapidly abandoned when the error is algorithmic.[290] Section 3.2 talks about the risk of AI amplifying information asymmetries: trust might be a key mechanism through which it will be amplified.

In a world of data abundance, it is important to ensure that the information on which decisions are based is accurate and reliable, regardless of whether that information is generated by a human or by AI. It remains to be seen whether behavioural biases commonly observed in humans such as excessive reliance on past experience, persuasion

287 See, for example, Dou et al. (2023).
288 See Timmermann (2006) and Wang et al. (2023) for surveys on forecast combination.
289 Hoff and Bashir (2015); Cabbidu et al. (2022); Chugarova and Luhan (2022).
290 Dietvorst et al. (2015, 2018).

bias, correlation neglect, limited attention, and a preference for humans also will be present in the AI setting. The benefits of AI for finance, as well as the risks, will depend critically on the quality of the information that is generated, and whether a high signal-to-noise ratio can be maintained.

5.4 DISCUSSION OF CHAPTER 4, "CORPORATE FINANCE AND GOVERNANCE WITH ARTIFICIAL INTELLIGENCE: OLD AND NEW", BY SEAN CAO

This chapter explores the frontier where artificial intelligence intersects with corporate governance. As a review article, it steps away from empirical execution and instead presents high-level conceptual insights meant to provoke thought and guide future research. My discussion seeks to amplify key points, drawing on examples, industry observations, and related academic work to spur meaningful engagement.

Based on my experience reviewing research, AI-related studies can generally be categorised into three types. Type 1 is methodological, using AI as a tool to answer traditional questions. For instance, researchers might use machine learning to predict restatements or returns, which are conventional topics enhanced by novel methods. Type 2 involves fundamentally new questions that emerge specifically due to the existence and influence of AI. These questions inherently incorporate AI, such as: How do humans collaborate with AI? How do corporate executives allocate tasks between AI agents and humans? Type 3 focuses on AI policy issues and examines how AI is reshaping corporate governance, an area increasingly influenced by complex societal challenges. The current chapter primarily falls within Type 2, with some connection to Type 3.

New boundaries of information asymmetry: Public information asymmetry

The first highlight in the chapter pertains to a shift in the definition and boundaries of information asymmetry. Traditionally, this concept referred to the knowledge gap between insiders and outsiders. Consider the case of Tesla: insiders possessed private production data that external investors lacked. However, with the advent of AI, this gap is shrinking. A notable example from the chapter recounts how, in 2017, external analysts used satellite imagery to estimate Tesla's Model 3 production. This form of alternative data, processed by AI, allowed outsiders to approximate previously exclusive insider knowledge. In this case, AI reduces traditional information asymmetry.

Yet, while AI narrows one gap, it widens another. The chapter introduces the novel concept of 'public information asymmetry', referring to disparities between external parties who all technically have equal access to public data. The gap now lies not in access to information, but in the power to process information. An analyst equipped with AI resources can extract insights from satellite images or complex financial disclosures that a retail investor cannot. Thus, a new information divide emerges among outsiders.

The asymmetry is no longer due to information exclusivity but stems from unequal abilities to interpret shared data. Analysts and institutions with access to advanced models can act on information that retail investors cannot practically process, despite both having theoretical access. This concept redefines fairness in financial markets. Previously, the concern was access; now, it's capability.

Several papers delve into this idea. One is Cao et al. (2024), which presents empirical evidence that AI-enabled analysts can better exploit alternative data, leading to advantageous trading outcomes. Another is Cao et al. (2025), which studies visual data from CEO presentations. The team extracted images from executive slide decks, which include photos of production sites, properties, or construction plans, and examined their informational value.

The researchers classified these images using a large vision model, distinguishing between those depicting past operations, future plans, forecasts, or general visuals. They then categorised investors into three groups. First, AI-enhanced institutional investors possess in-house talent and resources, allowing them to trade effectively on image-based signals. Second, non-AI institutional investors, while not using such images, maintain competitive returns due to their general sophistication. Third, retail investors, who lack access to AI capabilities, are effectively excluded from this informational advantage.

Despite equal access to the CEO presentations, only certain investors can process and act on these data. This is the essence of public information asymmetry, which is a gap based not on access but on processing power. It is a striking shift with serious implications for equity in capital markets.

A proposed solution is improving AI accessibility. While many focus on developing better models, fewer address how accessible these tools are. Take Google's BERT and ChatGPT: both are powerful, but ChatGPT's web interface makes it vastly more accessible. BERT, though released earlier, remains relatively unknown outside technical circles due to its programming complexity. Accessibility matters. Tools that are easy to use can democratise information processing and mitigate public information asymmetry.

AI accountability

The second key point discussed concerns AI accountability, particularly within the context of investment firms that are actively integrating AI into their operational workflows. To clarify what AI accountability means in practice, we can examine an observed trend in mid-sized investment companies that reveals both the increasing capabilities of AI and the persistent need for human oversight.

A typical investment workflow in such firms comprises four main stages: stock selection, fundamental analysis, risk management, and trade execution. In the initial stage, many firms have already implemented AI systems to identify potential investment opportunities. These AI-driven tools are used to generate a shortlist of stocks based on quantitative models and proprietary algorithms.

The second stage involves human analysts who act as gatekeepers. Although AI can generate a list of recommended stocks, not all selections are viable. Human analysts conduct a negative screening to identify red flags that AI might overlook. They rely on fundamental analysis to validate AI-generated picks and ensure that risky or inappropriate stocks are excluded. This process is critical because overlooking certain qualitative or contextual risks could compromise the integrity of the investment portfolio.

The third stage focuses on risk management. Here, another team evaluates how the inclusion of a new stock affects the overall volatility and risk exposure of the portfolio. Their objective is to maintain a balanced risk profile and avoid any extreme fluctuations that could result from poor stock selection. The final stage addresses microstructure concerns, where firms aim to minimise trading costs and optimise execution strategies.

AI is currently most prominent in the first stage, but its role is expanding into the second and third stages. Firms are beginning to use large language models trained on decades of analysts' notes. These notes contain detailed information about how to identify red flags and evaluate stocks. In other words, they document the reasoning processes and decision-making patterns of experienced professionals. By learning from these insights, AI systems are evolving into what might be termed AI fundamentalists, capable of not only generating stock lists but also performing the critical gatekeeping functions once reserved for humans.

A similar transformation is occurring in risk management. Firms are training AI systems on historical risk assessment strategies, allowing them to develop AI-based risk control mechanisms. In both cases, the deployment of AI leads to significant operational efficiency. For instance, teams that previously required 20 analysts and 10 risk managers might now function with just one or two individuals overseeing AI systems. This shift offers both labour cost savings and opportunities for business expansion without the need to scale human resources proportionally.

However, a question arises: Why does the position continue to be held by a real person after AI automation? One possible reason is that the individual still plays a functional role, potentially making better decisions than AI when acting as a gatekeeper. Another reason concerns accountability. Although AI can replicate many aspects of analytical reasoning and even outperform in some cases, it cannot assume legal or ethical responsibility.

Despite the growing capabilities of AI, firms consistently retain at least one human in the loop. Directors, CEOs, and managers must personally assume responsibility for decisions made under their supervision, even when those decisions are supported or executed by AI systems. This reflects a broader issue in corporate governance, where the delegation of decision making to AI must be carefully balanced with the need for human responsibility and oversight. The idea that "AI can do everything for you but cannot go to jail for you" is highlighted in the chapter to illustrate the continuing need for human responsibility.

The misalignment problem in AI agency

The third highlight is the misalignment problem in AI agency. This refers to a disconnect between what AI systems optimise and what humans actually want.

For AI systems to function effectively, they require clearly defined objectives, often operationalised as 'ground truth' labels in supervised learning contexts. This requirement is readily met in well-structured tasks such as stock return prediction, where the target variable (future returns) is explicit and quantifiable, enabling models to optimise by minimising prediction error in a rigorous manner.

However, many corporate governance decisions are ambiguous. Should a firm prioritise shareholder value, ESG compliance, or ethical standards? What weight should be given to each? Humans themselves often lack consensus on these objectives, making it nearly impossible to train AI effectively.

For AI to support decision making in corporate governance, we must first articulate clear goals. Without that, models will optimise for the wrong metrics, creating outcomes that deviate from broader societal or long-term interests.

This is a fundamental challenge. In engineering, problems are well-defined. In social sciences, ambiguity prevails. Unlike rocket launches, human decision making involves multiple, conflicting goals.

This complexity is our opportunity. Social science scholars must define the ground truths AI should optimise. We must not assume AI outputs are inherently correct. They reflect the objectives we feed them. Poorly defined goals yield poorly aligned results.

Many existing finance and accounting studies assume AI-generated outputs are the gold standard. But if the model was trained on ambiguous or inconsistent data, its predictions may reinforce flawed practices.

We must rethink training data in corporate governance. What decisions are correct? Historical board decisions may be inconsistent or politically driven. Without clarity, we risk teaching AI the wrong lessons.

In future research, scholars might explore how experimental designs or simulated governance environments could help generate cleaner training samples. Alternatively, ensemble models that incorporate multiple stakeholder perspectives may offer a more balanced approach to optimisation.

Other thoughts: Who should address AI and governance?

This brings us to a final reflection: Should AI and governance challenges be addressed by computer scientists or finance scholars? AI development fundamentally depends on three interrelated pillars. The first is algorithm design, as most foundational models and innovations, such as transformers, are available through open-source platforms. The second is computational power, which has become increasingly accessible through cloud-based services offering scalable infrastructure. The third and arguably most critical pillar is domain knowledge, which entails a deep understanding of institutional data, organisational contexts, and regulatory frameworks. It is in this third area where accounting and finance scholars can make the most distinctive contributions.

This is where accounting and finance scholars must lead. Consider domain-specific data like CEO presentations or analyst reports. CS scholars, without familiarity with this context, may miss key nuances.

Even within our field, this requires deep expertise. For example, identifying images in CEO slides that signal future growth demands knowledge of financial disclosures, not just technical skills.

AI is increasingly accessible. Pre-trained models like GPT and open-source alternatives like Lema allow scholars to work with AI without advanced programming skills. The barrier to entry is lower than ever. What matters now is asking meaningful questions and understanding institutional details.

Educating the next generation is also crucial. A free textbook co-authored by Wei Jiang and myself aims to address this, emphasising institutional knowledge, data literacy, and practical applications of large models in financial contexts.

Looking ahead, collaboration between finance, law, and computer science can help develop better frameworks for AI adoption in governance.

Conclusion

AI is transforming corporate governance in three key ways. It redefines information asymmetry by creating disparities in data processing, challenges accountability since it cannot assume legal responsibility, and raises concerns about misalignment when human goals are ambiguous.

These issues extend beyond technical design and require expertise in institutional contexts and human judgement. Finance and accounting scholars are well equipped to lead this work, drawing on their deep understanding of financial data, organisational practices, and regulatory environments.

Rather than treating AI as a tool for old questions, scholars must examine how it reshapes the questions themselves. This chapter calls for deeper engagement to guide AI's integration with rigor and insight.

5.5 DISCUSSION OF CHAPTER 4, "CORPORATE FINANCE AND GOVERNANCE WITH ARTIFICIAL INTELLIGENCE: OLD AND NEW", BY LUCA ENRIQUES[291]

This commentary examines the tensions and insights presented in Chapter 4 of the report. The chapter offers a comprehensive analysis of how artificial intelligence technologies are reshaping corporate governance and market dynamics, yet several nuances warrant further examination. First, my comments focus on what I find to be a neglected angle in the chapter, namely, AI as a tool to optimise managerial opportunism. Second, this commentary casts doubt on the suggestion that global harmonisation will be the best way to address the policy challenges of AI in finance with specific regard to corporate governance and information asymmetries in public markets. Third, it provides a more optimistic view of the impact on market efficiency of the new information asymmetries that the chapter identifies.

The agency problem revisited: AI as a tool for managerial opportunism

The chapter effectively outlines how AI systems introduce novel agency problems when acting as oracles, agents, or sovereigns, focusing predominantly on the misalignment between AI's programmed objectives and principals' intended outcomes. However, it overlooks a critical dimension of the agency problem: the potential for AI to become an instrument for managerial opportunism.

While the chapter correctly notes that AI systems themselves do not exhibit moral hazard in the traditional sense of pursuing personal perks, it fails to adequately address how human agents (particularly managers) might leverage these technologies to optimize their own objectives rather than those of shareholders. The fundamental insight missing from the chapter is that unless human and shareholder interests are perfectly aligned, AI – being a tool controlled by managers – is more likely to be deployed in service of managerial interests.[292]

291 This text has been drafted with the help of Claude.ai based on the author's slides for the IESE workshop where the report was discussed.
292 Enriques and Zetzsche (2020).

The choice of input data, the framing of objectives, and the implementation of AI systems within corporate structures all remain under managerial control. This creates a meta-agency problem where the very design and deployment of AI tools become subject to the same agency conflicts the chapter analyses. For instance, managers might selectively feed data into LLMs or customise algorithmic parameters in ways that validate their preferred strategies or obscure underperformance, while maintaining the appearance of data-driven objectivity.

The chapter's discussion of incentive compatibility in AI algorithm design and regulatory frameworks is valuable but addresses a secondary concern. Before considering how to ensure AI systems act in accordance with their programmed objectives, we must examine the incentives shaping those who determine what those objectives should be. The chapter would benefit from exploring this layer of complexity, if only to rule out its relevance, on the basis that existing analogic governance tools keep this meta-agency problem under control.

Regulatory approaches: Global harmonisation versus experimental federalism

The chapter makes a compelling case for "*consistent, global and unified governance frameworks*" to address the cross-border nature of AI systems and reduce regulatory arbitrage. While this approach has merit, especially for coordinating responses to systemic risks, it may underestimate both the practical challenges of achieving global regulatory consensus and the potential benefits of regulatory diversity.

First, the chapter's call for uniform criteria to distinguish legitimate actions from manipulative behaviours in market contexts overlooks existing legal mechanisms that already mitigate cross-border regulatory arbitrage. The 'effects doctrine' in market abuse regulation, for instance, allows jurisdictions to punish manipulative conduct affecting their markets regardless of where it originates. While AI's superior capabilities may complicate enforcement, they do not necessarily create new risks of regulatory arbitrage in this specific domain.

Second, the chapter's emphasis on global harmonisation fails to acknowledge the value of regulatory experimentation at national and/or regional levels. Diverse regulatory approaches can function as laboratories for policy innovation, potentially leading to more effective long-term regulatory outcomes compared to a coordinated, top-down approach. By allowing different jurisdictions to test varied regulatory responses to AI-driven market behaviours, we can gain valuable empirical evidence about which approaches most effectively balance innovation and risk mitigation.

The chapter's citation of the *International Scientific Report on the Safety of Advanced AI*[293] as evidence of progress toward global frameworks is encouraging. However, it would be strengthened by a more nuanced discussion of when harmonisation is necessary versus when regulatory diversity might better serve market development and innovation. A more balanced approach might advocate for minimum global standards in areas with clear externalities or systemic risks, while encouraging experimentation in domains where the consequences remain localised or where optimal regulatory approaches are uncertain.

The evolution of information asymmetries: Market adaptation and efficiency

The chapter's analysis of how AI and big data transform information asymmetries in public markets correctly highlights that big data provide insiders with new sources of non-public material information while alternative data simultaneously reduce information asymmetries between insiders and outsiders. However, the chapter's framing of these developments primarily as threats to market liquidity may be overly pessimistic.

Information asymmetries are not merely market imperfections to be eliminated but fundamental drivers of trading activity and informational efficiency. Traders constantly seek unpriced information, and their actions incorporate this information into prices, benefiting market participation and liquidity over time. The chapter's concern that new forms of information asymmetry might harm liquidity fails to fully appreciate this dynamic perspective.

While AI-driven analysis of alternative data might initially create disparities among market participants, these disparities incentivise innovation and adaptation. The chapter briefly mentions, but does not fully explore, how markets might evolve in response to AI-enabled trading strategies, similar to how they adapted to high-frequency trading through the development of dark pools and smart order routing systems.[294] This adaptive capacity suggests that initial asymmetries might be mitigated over time as more participants develop necessary capabilities or as market mechanisms evolve.

The redefinition of inside information and market integrity

One of the chapter's most valuable contributions is its exploration of how AI redefines the boundaries between inside and outside information. Traditional frameworks that distinguish between public and non-public information become increasingly inadequate when AI can extract non-obvious insights from public data or when alternative data sources blur the line between legitimate research and unfair information advantages.

293 Bengio et al. (2025).
294 O'Hara (2015).

The chapter effectively identifies how AI amplifies the processing of public information, leading to what it terms "*public information asymmetry*", where technological capabilities rather than access to inside information determine trading advantages. This shift has profound implications for market surveillance and regulation, which the chapter begins to address through its discussion of redefining equal access in an AI-driven market and establishing principles for the fair use of alternative data.

However, the chapter could more explicitly recognise that not all information asymmetries warrant regulatory intervention. The key concern should not be with information asymmetry itself but rather with insider-dominated informed trading. That is because when insider trading is permitted, professionally informed traders stand to systematically lose to them, which may decrease market efficiency because insiders may be worse at interpreting the price effects of events at the macro level, may refrain from trading on bad news, given the negative impact this may have on their job security, and might also be irrationally optimistic in interpreting new information.[295] A more granular typology of information advantages, distinguishing between those that primarily improve market efficiency and those that primarily enable the extraction of rents or undermine market integrity, would be useful.

Feedback effects and reflexivity in AI-driven markets

The chapter touches upon, but does not fully develop, the implications of feedback effects where market perception (now AI-driven) influences managerial decisions. This reflexive relationship between market signals and corporate behaviour takes on new dimensions in an AI-dominated environment, where algorithms on both sides might interact in unpredictable ways.

For instance, if corporate managers know that AI systems are analysing their disclosures for sentiment and specific keywords, they might strategically modify their language in ways that game these systems. Similarly, if trading algorithms respond predictably to certain news patterns, this might incentivise the strategic timing or framing of corporate announcements. The chapter acknowledges these dynamics but could more thoroughly explore how they might reshape both market information environments and corporate disclosure strategies.

Conclusion: Towards a dynamic understanding of AI in finance

The chapter provides a comprehensive analysis of how AI transforms corporate governance and market information environments. However, its most significant limitation is its insufficient attention to the dynamic, evolutionary nature of these transformations. Markets and institutions adapt to technological disruptions, often in ways that mitigate initial concerns while introducing new challenges.

295 Goshen and Parchomovsky (2005).

A more dynamic perspective would recognise that the initial asymmetries created by AI and alternative data might evolve as market participants develop counter-strategies, as regulations adapt, and as technologies themselves become more widely accessible. The case of high-frequency trading's evolution from a disruptive force to a more integrated part of market making offers a valuable precedent that might inform our expectations about AI's long-term impact.

References

Abdul Jabbar, M., O. Wibisono, A. Widjanarti and A. Zulen (2022), "Machine learning for measuring central bank policy credibility and communication from news", *IFC Bulletin* 59.

Abis, S. (2022), "Man vs. Machine: Quantitative and Discretionary Equity Management", available at SSRN.

Abis, S. and L. Veldkamp (2024), "The changing economics of knowledge production", *The Review of Financial Studies* 37(1): 89–118.

Accornero, M. and G. Boscariol (2021), "Machine learning for anomaly detection in data sets with categorical variables and skewed distributions", *IFC Bulletin* 57.

Acemoglu, D., D. Autor, J. Hazell and P. Restrepo (2022), "Artificial Intelligence and Jobs: Evidence from Online Vacancies", *Journal of Labor Economics* 40: 293-340.

AFM – Dutch Authority for the Financial Markets (2023), "Machine learning in algorithmic trading", 3 March.

Aghion, P. and P. Bolton (1992), "An incomplete contracts approach to financial contracting", *The Review of Economic Studies* 59(3): 473–494.

Agrawal A., J.S. Gans and A. Goldfarb (2017), "Prediction machines, judgement and complexity", in A. Agrawal, J. Gans and A. Goldfarb (eds), *The Economics of Artificial Intelligence*, University of Chicago Press.

Agrawal, A., J. Gans, and A. Goldfarb (2018), *Prediction machines: The Simple Economics of Artificial Intelligence*, Harvard Business Press.

Ahnert, T., S. Doerr, N. Pierri and Y. Timmer (2024), "Information technology in banking and entrepreneurship", *Management Science*, forthcoming.

Akerlof, G.A. (1970), "The market for 'lemons': Quality uncertainty and the market mechanism", *Quarterly Journal of Economics* 84: 488–500.

Ala-Pietila, P., Y. Bonnet, U. Bergmann et al. (2020), "The assessment list for trustworthy artificial intelligence (ALTAI)", European Commission.

Aldasoro, I., O. Armantier, S. Doerr, L. Gambacorta and T. Oliviero (2024a), "Survey evidence on gen AI and households: job prospects amid trust concerns", *BIS Bulletin* 86.

Aldasoro, I., S. Doerr, L. Gambacorta, S. Notra, T. Oliviero and D. Whyte (2024b), "Generative artificial intelligence and cybersecurity in central banking", CEPR Discussion Paper 19244.

Aldasoro, I., S. Doerr, L. Gambacorta, and D. Rees (2024c), "The impact of artificial intelligence on output and inflation", CEPR Discussion Paper 19604.

Aldasoro, I., L. Gambacorta, A. Korinek, V. Shreeti and M. Stein (2024d), "Intelligent financial system: how AI is transforming finance", CEPR Discussion Paper 19181.

Aldasoro, I., S. Doerr, L. Gambacorta, G. Gelos and D. Rees (2024e), "Artificial intelligence, labour markets and inflation", SUERF Policy Brief 923.

Altman, S. (2024), "The intelligence age", 23 September (https://ia.samaltman.com).

Altman, S (2025), "Reflections", 6 January (https://blog.samaltman.com/reflections).

Altmeyer P., A.M. Demetriou, A. Bartlett and C.C.S. Liem (2024), "Stop Making Unscientific AGI Performance Claims", paper presented at 41st ICML, Vienna.

Amodei, D. (2024), "Machines of loving grace", October (https://darioamodei.com/machines-of-loving-grace).

Anthropic (2024), "Computer Use", 22 October (https://www.anthropic.com/news/3-5-models-and-computer-use).

Aquilina, M., E. Budish and P. O'Neill (2022), "Quantifying the High-Frequency Trading 'Arms Race'", *The Quarterly Journal of Economics* 137: 493–564.

Aramonte, S., W. Huang and A. Schrimpf (2021), "DeFi risks and the decentralization illusion", *BIS Quarterly Review,* December.

Araujo, D., G. Bruno, J. Marcucci, R. Schmidt and B. Tissot (2022), "Machine learning applications in central banking: an overview", *IFC Bulletin* 57.

Araujo, D., G. Bruno, J. Marcucci, R. Schmidt and B. Tissot (2023), "Data science in central banking: applications and tools", *IFC Bulletin* 59.

Araujo, D., S. Doerr, L. Gambacorta, and B. Tissot (2024), "Artificial intelligence in central banking," *BIS Bulletin* 2024.

Armantier O., S. Doerr, J. Frost, A. Fuster and K. Shue (2021), "Whom do consumers trust with their data? US survey evidence", *BIS Bulletin* 42.

Armas, A., M. Luna, and C. Montoro (2024), "The Changing Nature of the Financial System: Implications for Resilience and Long-term Growth in EMEs – The Peruvian Experience", BIS Papers No. 148, Bank for International Settlements.

Armour, J, R. Parnham and M. Sato, M. (2020), "Augmented lawyering", working paper.

Asker, J., C. Fershtman and A. Pakes (2024), "The impact of artificial intelligence design on pricing", *Journal of Economics and Management Strategy* 33: 276–304.

Assad, S., R. Clark, D. Ershov and L. Xu, (2024), "Algorithmic Pricing and Competition: Empirical Evidence from the German Retail Gasoline Market," *Journal of Political Economy* 132: 723-771.

Avdjiev, S., B. Bogdanova, P. Bolton, W. Jiang and A. Kartasheva (2020), "CoCo issuance and bank fragility", *Journal of Financial Economics* 138(3): 593–613.

Bai, J., T. Phillipon and A. Savov (2016), "Have financial markets become more informative", *Journal of Financial Economics* 122: 625-654.

Baik, B., A.G. Kim, D.S. Kim and Y. Sangwon (2024), "Vocal Delivery Quality in Earnings Conference Calls", *Journal of Accounting and Economics* (forthcoming).

Balch, T., D. Byrd, W. Jiang and B Yang (2025), "Learning to misbehave without being taught: Market manipulation by AI with reinforcement learning", working paper, Emory University.

Baldauf, M. and J. Mollner (2020), "High Frequency Trading and Market Performance", *Journal of Finance* 75: 1495-1526.

Barberis, N. and L. Jin (2023), "Model-free and Model-based Learning as Joint Drivers of Investor Behavior", NBER Working Paper 31081.

Barbopoulos, L., R. Dai, T. Putnins and A. Saunders (2023), "Market efficiency when machines access information", available at SSRN.

Baron, M., J. Brogaard, B. Hagströmer and A. Kirilenko (2019), "Risk and Return in High-Frequency Trading", *Journal of Financial and Quantitative Analysis* 54: 993-1024.

Bartov, E., L. Faurel and P.S. Mohanram (2018), "Can Twitter help predict firm-level earnings and stock returns?", *The Accounting Review* 93: 25–57

Bassett, W.F. and R.L. Lumsdaine (2001), "Probability Limits: Are Subjective Assessments Adequately Accurate?", *Journal of Human Resources* 36(2): 327-363.

Beason, T. and S. Wahal (2020), "The Anatomy of Trading Algorithms", available at SSRN.

Beaumont, P., H. Tang and E. Vansteenberghe (2024), "Collateral effects: The role of fintech in small business lending", *Débats économiques et financiers* 42.

Beaver Finance (2022), "DeFi security lecture 7 — Price oracle manipulation", *Medium*, 13 January.

Beck, T., L. Gambacorta, Y. Huang, Z. Li and H. Qiu (2022), "Big techs, QR code payments and financial inclusion", BIS Working Paper 1011.

Beck, T., T. Peltonen, E. Perotti, A. Sánchez Serrano and J. Suarez (2023), *Corporate credit and leverage in the EU: recent evolution, main drivers and financial stability implications*, European Systemic Risk Board Report of the Advisory Scientific Committee 14.

Beerman, K., J. Prenio and R. Zamil (2021), "Suptech tools for prudential supervision and their use during the pandemic", *FSI Insights on Policy Implementation* 37.

Begenau, J., M. Farboodi and L. Veldkamp (2018), “Big data in finance and the growth of large firms”, *Journal of Monetary Economics* 97: 71–87.

Bell S., B. Gadanecz, L. Gambacorta, F. Perez-Cruz and V. Shreeti (2025), “Artificial intelligence and human capital: challenges for central banks”, *BIS Bulletin* 100.

Benamar H., T. Foucault and C. Vega, (2021), “Demand for information, macroeconomic uncertainty and the response of U.S. treasury securities to news”, *Review of Financial Studies* 34: 3403-3455.

Bengio, Y., S. Mindermann, D. Privitera et al. (2025), “International AI safety report”, arXiv preprint, arXiv:2501.17805.

Berg, T., V. Burg, A. Gombović and M. Puri (2020), “On the rise of fintechs – credit scoring using digital footprints”, *Review of Financial Studies* 33: 2845-97.

Berg, T., V. Burg, J. Keil and M. Puri (2025), "The Economics of 'Buy Now, Pay Later': A Merchant’s Perspective", *Journal of Financial Economics,* forthcoming.

Berger, A.N. and G.F. Udell (1990), "Collateral, Loan Quality, and Bank Risk", *Journal of Monetary Economics* 25: 21–42.

Berger, A.N. and G.F. Udell (1995), "Relationship lending and lines of credit in small firm finance", *Journal of Business* 68(3): 351–381.

Biais, B., T. Foucault, T. and S. Moinas (2015), “Equilibrium Fast Trading”, *Journal of Financial Economics* 116: 292-313.

Biglaiser G., J. Cremer and A. Mantovani (2024), “The economics of the cloud”, TSE Working Paper 1520.

BIS – Bank for International Settlements (2019), *Big techs in finance: Opportunities and risks,* BIS Annual Economic Report, June.

BIS (2023), *Blueprint for the future monetary system: improving the old, enabling the new*, BIS Annual Economic Report, June.

BIS (2024), *Artificial intelligence and the economy: implications for central banks*, BIS Annual Economic Report, June.

BIS Innovation Hub (2023), “Project Aurora: the power of data, technology and collaboration to combat money laundering across institutions and borders”, May.

Björkegren, D., J.E. Blumenstock and S. Knight (2020), “Manipulation-proof machine learning”, arXiv preprint arXiv:2004.03865.

Blankenburg J. and O. Röhe (2024), “TIA Platform: Analysis of implementation of structural reforms”, Central Banking Technology Leadership Forum, 4-5 December.

Blankespoor, E., B.P. Miller and H.D. White (2014) “Initial evidence on the market impact of the XBRL mandate”, *Review of Accounting Studies* 19: 1468–1503.

Blankespoor, E., E. deHaan and I. Marinovic (2020), “Disclosure processing costs, investors’ information choice, and equity market outcomes: A review”, *Journal of Accounting and Economics* 70: 1–46.

Bogousslavsky, N., V. Fos and D. Murayev (2024), “Informed Trading Intensity”, *Journal of Finance* 79: 903-948.

Bohren, J.A. (2016), “Informational Herding with Model Misspecification”, *Journal of Economic Theory* 163: 222-247.

Boissay, F., T. Ehlers, L. Gambacorta and H.S. Shin (2021), “Big techs in finance: on the new nexus between data privacy and competition”, in R. Rau, R. Wardrop and L. Zingales (eds), *The Palgrave Handbook of International Finance,* Springer International Publishing.

Bolandnazar, M., R.J. Jackson, W. Jiang and J. Mitts (2020), “Trading against the random expiration of private information: A natural experiment”, *Journal of Finance* 75: 5–44.

Bolton, P., X. Freixas, L. Gambacorta and P.E. Mistrulli (2016), "Relationship and transaction lending in a crisis", *Review of Financial Studies* 29(10): 2643–2676.

Bond, P., A. Edmans and I. Goldstein (2012), “The real effects of financial markets”, *Annual Review of Economics* 1: 339-36.

Bonelli, M. (2024), “Data-Driven Investors”, working paper, London Business School.

Bonelli, M. and T. Foucault (2024), “Does big data devalue traditional expertise? Evidence from active funds?”, working paper, HEC Paris.

Bonfim, D., C. Custódio and C. Raposo (2023), "Supporting Small Firms through Recessions and Recoveries", *Journal of Financial Economics* 147: 658–688.

Bonfim, D., A. Matyunina, R. De Haas and S. Ongena (2025), "Bank Specialization in Lending to New Firms", mimeo.

Bostrom, N. (2015), *Superintelligence: Paths, Dangers, Strategies*, Oxford University Press.

Botti, N., F. Haberkorn and S. Khan (2025), “Efficacy of AI RAG Tools for Complex Information Extraction and Data Annotation Tasks on Bank Documents”, 4th IFC and Bank of Italy Workshop on Data Science in Central Banking, 18-20 February.

Bourveau, T., J. Brendel and J. Schoenfeld (2023), “Decentralized finance (DeFi) assurance: Early evidence”, *Review of Accounting Studies*, forthcoming.

Brandts, J., A.E. Giritligil and R.A. Weber (2015), “An Experimental Study of Persuasion Bias and Social Influence”, *European Economic Review* 80: 214-229.

Brav, A., W. Jiang, F. Partnoy and R. Thomas (2008), “Hedge fund activism, corporate governance, and firm performance”, *Journal of Finance* 63: 1729–1775.

Breckenfelder, J. (2024), “Competition among high frequency traders and market quality”, *Journal of Economic Dynamics and Control* 166: 2-22.

Brogaard, J. and C. Garriott (2019), “High frequency trading competition”, *Journal of Financial and Quantitative Analysis* 54: 1469-1497.

Brogaard, J. and A. Zareei (2023), “Machine Learning and the Stock Market”, *Journal of Financial and Quantitative Analysis* 58: 1431-1472.

Brogaard, J., T. Hendershott and R. Riordan (2017), “High frequency trading and the 2008 short sale ban”, *Journal of Financial Economics* 124: 22-42.

Brown, Z.Y. and A. MacKay (2023), “Competition in Pricing Algorithms”, *American Economic Journal: Microeconomics* 15: 109–56.

Browning, J. and Y. LeCun (2022), “AI and the limits of language”, *Noema*, 23 August.

Brynjolfsson, E. and K. McElheran (2019), “Data in action: data-driven decision making and predictive analytics in US manufacturing”, Rotman School of Management Working Paper 3422397.

Brynjolfsson, E., D. Li and L. Raymond (2023), “Generative AI at work”, NBER Working Paper 31161.

Bryzgalova, S., M. Pelger and J. Zhu (2020), “Forest through the trees: Building cross-sections of stock returns”, available at SSRN.

Buchak, G., G. Matvos, T. Piskorski and A. Seru (2018), “Fintech, regulatory arbitrage, and the rise of shadow banks”, *Journal of Financial Economics* 130(3): 453-483.

Buchak, G., G. Matvos, T. Piskorvski and A. Seru (2022), “Why is intermediating houses so difficult? Evidence from iBuyers”, NBER Working Paper 28252.

Buckmann, M., G. Potjagailo and P. Schnattinger (2023), “Dissecting UK service inflation via a neural network Phillips curve”, Bank of England, Bank Underground, 10 July.

Budish, E., P. Cramton, and J. Shim (2015), “The High-Frequency Trading Arms Race: Frequent Batch Auctions as a Market Design Response,” *The Quarterly Journal of Economics* 130: 1547–1621.

Bushman, R.M. and A.J. Smith (2001), “Financial accounting information and corporate governance”, *Journal of Accounting and Economics* 32: 237–333.

Byrd, D. (2022), “Learning not to spoof”, in *Proceedings of the Third ACM International Conference on AI in Finance*, pp. 139–147.

Cabiddu, F., L. Moi, G. Patriotta, and D. Allen (2022), "Why Do Users Trust Algorithms? A Review and Conceptualization of Trust and Trust Over Time", *European Management Journal* 40: 685-706.

Cagala, T., J. Hees, D. Herurkar, M. Meier, N.-T. Nguyen, T. Sattarov, K. Troutman and P. Weber (2021), "Unsupervised outlier detection in official statistics", *IFC Bulletin* 57.

Calvano, E., G. Calzolari, V. Denicolo and S. Pastorello (2020a), "Artificial intelligence, algorithmic pricing, and collusion", *American Economic Review* 110(10): 3267–3297.

Calvano, E., G. Calzolari, V. Denicolò, J. Harrington and S. Pastorello (2020b), "Protecting consumers from collusive prices due to AI", *Science* 370: 1040-1042.

Campello, M., L.W. Cong and L. Zhou (2024), "Alphamanager: A data-driven-robust-control approach to corporate finance", working paper, University of Florida.

Cao, K. and H. You (2024), "Fundamental Analysis via Machine Learning", *Financial Analysts Journal* 80: 74-98.

Cao, J., J. Chen, J. Hull, Z. Poulos (2020), "Deep Hedging of Derivatives Using Reinforcement Learning", working paper, University of Toronto.

Cao, S., W. Jiang, B. Yang and A.L. Zhang (2023), "How to talk when a machine is listening: Corporate disclosure in the age of AI", *The Review of Financial Studies* 36(9): 3603–3642.

Cao, S., W. Jiang, J. Wang and B. Yang (2024), "From man vs. machine to man + machine: The art and AI of stock analyses", *Journal of Financial Economics* 160.

Cao, S., Y. Cheng, M. Wang, Y. Xia and B. Yang (2025), "Visual information and AI divide: Evidence from corporate executive presentations", available at SSRN.

Carletti, E. (2004), "The structure of bank relationships, endogenous monitoring, and loan rates", *Journal of Financial Intermediation* 13(1): 58–86.

Carletti, E., A. Fatas, S. Claessens and X. Vives (2020), *The bank business model in the post-Covid 19 world*, The Future of Banking 2, CEPR Press.

Cartea, A., P. Chang, and G. Garcia-Arenas (2023), "Spoofing and Manipulating Order Books with Learning Algorithms", working paper, Oxford-Man Institute of Quantitative Finance.

Cespa, G. and T. Foucault (2014), "Illiquidity Contagion and Liquidity Crashes", *The Review of Financial Studies* 27: 1615–1660.

Cespa, G. and X. Vives (2025), "Market opacity and fragility: Why liquidity evaporates when it is most needed", working paper.

Chabakauri, G., V. Fos and W. Jiang (2025), "Trading ahead of barbarians' arrival at the gate: Insider trading on non-inside information", *The Review of Finance*, forthcoming.

Chainalysis (2023), "Oracle manipulation attacks are rising, creating a unique concern for DeFi", blog, 7 March.

Chan, A., C. Ezell, M. Kaufmann et al. (2024), "Visibility into AI Agents", arXiv preprint arXiv:2401.13138.

Charpentier, A., E. Romuald and C. Remlinger (2020), "Reinforcement learning in economics and finance", working paper, ArXiv.

Chen, Q., I. Goldstein and W. Jiang (2007), "Price informativeness and investment sensitivity to stock price", *The Review of Financial Studies* 20(3): 619–650.

Chen, H., P. De, Y. Hu and B.-H. Hwang (2014), "Wisdom of crowds: The value of stock opinions transmitted through social media", *The Review of Financial Studies* 27(5): 1367–1403.

Chen, Y., B. Kelly and D. Xiu (2022b), "Expected Returns and Large Language Models", available at SSRN.

Chen, L., M. Pelger and J. Zhu (2023a), "Deep learning in asset pricing", *Management Science* 70: 714-750.

Chen, M.A., S.S. Hu, J. Wang and Q. Wu (2023b), "Can blockchain technology help overcome contractual incompleteness? Evidence from state laws", *Management Science* 69(11): 6540–6567.

Chi, F., B.-H. Hwang and Y. Zheng (2023), "The use and usefulness of big data in finance: Evidence from financial analysts", *Management Science,* forthcoming.

China Technology Ministry (2019), *Governance Principles for a New Generation of Artificial Intelligence: Develop Responsible Artificial Intelligence.*

Chinco, A., A. Clark-Joseph and M. Ye (2019), "Sparse Signals in the Cross-section of Returns", *Journal of Finance* 74: 449 – 492.

Christian, B. (2020), *The Alignment Problem: Machine Learning and Human Values,* W.W. Norton & Company.

Chugunova, M. and W.J. Luhan (2024), "Ruled by Robots: Preference for Algorithmic Decision Makers and Perceptions of their Choices", *Public Choice* 202: 1-24.

Citi GPS (2024), *AI in Finance: Bot, Bank & Beyond*, June.

Citi GPS (2025), *Agentic AI: Finance & the 'Do It For Me' Economy*, January.

Claessens, S. (2009), "Competition in the Financial Sector: Overview of Competition Policies", *The World Bank Research Observer* 24(1): 83–118.

CMA – Competition and Markets Authority (2024), "AI foundation models: update paper", 11 April.

Coase, R.H. (1937), "The nature of the firm", in Williamson and S.G. Winter (eds) (1993), *The Nature of the Firm: Origins, Evolution, and Development*.

Cohen, L., C. Malloy and Q. Nguyen (2020), "Lazy Prices", *The Journal of Finance* 75: 1371-1415.

Coleman, B., K. Merkley, and J. Pacelli (2022), "Human Versus Machine: A Comparison of Robo-Analyst and Traditional Research Analyst Investment Recommendations", *The Accounting Review* 97: 221–244.

Colliard, J.E, T. Foucault and S. Lovo (2023), "Algorithmic Pricing and liquidity in Securities Markets", available at SSRN.

Cong, L.W. and Z. He (2019), "Blockchain disruption and smart contracts", *The Review of Financial Studies* 32(5): 1754–1797.

Cong, L.W., Z. He and J. Li (2021), "Decentralized mining in centralized pools", *The Review of Financial Studies* 34(3): 1191–1235.

Cong, L., Ke, T., J. Wang and Y. Zhang (2022), "AlphaPortfolio: Direct Construction Through Deep Reinforcement Learning and Interpretable AI", available at SSRN.

Cong, L.W., D. Rabetti, C.C. Wang and Y. Yan (2023), "Centralized governance in decentralized organizations", available at SSRN.

Costello, A., A. Down and M. Metha (2020), "Machine+ man: a field experiment on the role of discretion in augmenting AI-based lending models", *Journal of Accounting and Economics* 70: 1013-1060.

Coughlan, J. and A.G. Orlov (2023), "High-frequency trading and market quality: Evidence from account-level futures data", *Journal of Futures Markets* 43(8): 1126–1160.

Crisanto, J.C., C.B. Leuterio, J. Prenio and J. Yong (2024), "Regulating AI in the financial sector: Recent developments and main challenges", *FSI Insights* 63.

D'Acunto, F., N. Prabhala and A. Rossi (2019), "The promises and pitfalls of robo-advising", *Review of Financial Studies* 32: 1984-2020.

da Silva, T. and D. Thesmar (2024), "Noise in Expectations: Evidence from Analyst Forecasts", *The Review of Financial Studies* 37: 1494–1537.

De Fiore, F., G. Cornelli, L. Gambacorta, and C. Manea (2024), "Big tech, financial intermediation and the macroeconomy", Boston FED Discussion Paper, 12 November.

De Roure, C., L. Pelizzon and P. Tasca (2016), "How does P2P lending fit into the consumer credit market?", Deutsche Bundesbank Discussion Paper 30.

Danielsson, J., R. Macrae, A. Uthemann (2022), "Artificial intelligence and systemic risks", *Journal of Banking and Finance* 140: 106290.

Dass, N. and M. Massa (2011), "The impact of a strong bank-firm relationship on the borrowing firm", *Review of Financial Studies* 24(4): 1204–1260.

DeepMind (2024), "Project Mariner" (https://deepmind.google/technologies/project-mariner/).

Degryse, H. and S. Ongena (2005), "Distance, lending relationships, and competition", *Journal of Finance* 60: 231–266.

DeMarzo, P.M., D. Vayanos, and J. Zwiebel (2003), "Persuasion Bias, Social Influence, and Unidimensional Opinions", *The Quarterly Journal of Economics* 118(3): 909-968.

Denes, J., A. Lestrade and L. Richardet (2021), "Using Twitter data to measure inflation perception", *IFC Bulletin* 57.

Dessaint, O., T. Foucault and L. Frésard (2024), "Does alternative data improve financial forecasting? The horizon effect", *Journal of Finance* 79: 2237-2287.

Dessaint, O., T. Foucault and L. Frésard (2025), "The Horizon of Investors' Information and Corporate Investment", available at SSRN.

Diamond, D.W. and R.E. Verrecchia (1991), "Disclosure, liquidity, and the cost of capital", *Journal of Finance* 46: 1325–1359.

Dietvorst, B.J., J.P. Simmons, and C. Massey (2015), "Algorithm Aversion: People Erroneously Avoid Algorithms After Seeing Them Err", *Journal of Experimental Psychology: General* 144(1): 114-126.

Dietvorst, B.J., J.P. Simmons, and C. Massey (2018), "Overcoming Algorithm Aversion: People Will Use Imperfect Algorithms if They Can (Even Slightly) Modify Them", *Management Science* 64(3): 1155-1170.

Doerr, S. (2020), "Housing booms, reallocation and productivity", BIS Working Paper 904.

Dong, G.N. (2024), "Can AI Replace Stock Analysts? Evidence from Deep Learning Financial Statements", available at SSRN.

Dou, W.W., I. Goldstein and Y. Ji (2023), "AI-powered trading, algorithmic collusion, and price efficiency", available at SSRN.

Du, S., K. Guo, F. Haberkorn, A. Kessler, I. Kitschelt, S.J. Lee, A. Monken, D. Saez, K. Shipman and S. Thakur (2024), "Do anecdotes matter? Exploring the Beige Book through textual analysis from 1970 to 2023", *IFC Bulletin*, forthcoming.

Duffie, D., T. Foucault, L. Veldkamp and X. Vives (2022), *Technology and Finance*, The Future of Banking 4, CEPR Press.

Dugast, J. and T. Foucault (2018), "Data abundance and asset price informativeness", *Journal of Financial Economics* 130(2): 367–391.

Dugast, J. and T. Foucault (2025), “Equilibrium Data Mining and Data Abundance”, *Journal of Finance* 80: 211-258.

ECB – European Central Bank (2023), “Suptech: thriving in the digital age”, *Supervision Newsletter*, 15 November.

Easley, D., M. López de Prado, M. O’Hara, and Z. Zhang (2021), “Microstructure in the Machine Age”, *The Review of Financial Studies* 34: 3316–3363.

Edmans, A., S. Jayaraman and J. Schneemeier (2017), “The source of information in prices and investment-price sensitivity”, *Journal of Financial Economics* 126: 74-96.

Eisfeldt, A.L. and G. Schubert (2024), “AI and finance”, NBER Working Paper 33076.

Enke, B. and F. Zimmermann (2019), “Correlation Neglect in Belief Formation”, *Review of Economic Studies* 86: 313-332.

Enriques, L. and D. Zetzsche (2020), “Corporate technologies and the tech Nirvana fallacy”, 72 *Hastings Law Journal*: 55-98.

ESMA – European Securities and Markets Authority (2020) “MiFID II/MiFIR review report on Algorithmic Trading”, Consultation Paper.

European Union (2024), “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”.

Ewertz, J., C. Knickrehm, M. Nienhaus and D. Reichmann (2024), “Listen Closely: Measuring Vocal Tone in Corporate Disclosures”, available at SSRN.

Ezrachi, A. and M. Stuche (2017), “Artificial intelligence and collusion: When computers inhibit competition”, *University of Illinois Law Review* 5: 1776-1808.

Farboodi, M. and L. Veldkamp (2020), “Long run growth of financial data technology”, *American Economic Review* 110: 2485-2523.

Farboodi, M. and L. Veldkamp (2023), "Data and Markets", *Annual Review of Economics* 15: 23–40.

Farboodi, M., R. Mihet, T. Philippon and L. Veldkamp (2019), “Big data and firm dynamics”, *AEA Papers and Proceedings* 109: 38-42.

Farboodi, M., A. Matray, L. Veldkamp and V. Venkateswaran, (2022), “Where Has All the Data Gone?”, *The Review of Financial Studies* 35: 3101–3138.

Farboodi, M., D. Singal, L. Veldkamp and V. Venkateswaran (2024), “Valuing Financial Data”, *The Review of Financial Studies*, forthcoming.

Feyen, E., J. Frost, L. Gambacorta, H. Natarajan and M. Saal (2021), “A policy triangle for Big Techs in finance”, VoxEU.org, October.

Foucault, T., R. Kozhan, and W. Tham (2017), "Toxic Arbitrage" *Review of Financial Studies* 30: 1053-1094.

Foucault T., M. Pagano, and A. Röell (2024), *Market Liquidity: Theory, Evidence and Policy,* Oxford University Press.

Frazzini, A., R. Israel, T. Moskowitz (2018), "Trading Costs", available at SSRN.

Froot, K., N. Kang, G. Ozik and R. Sadka (2017), "What do measures of real-time corporate sales say about earnings surprises and post-announcement returns?", *Journal of Financial Economics* 125(1): 143–162.

Frost, J., L. Gambacorta, Y. Huang, H. S. Shin and P. Zbinden (2019), "BigTech and the changing structure of financial intermediation", *Economic Policy* 34(100): 761–799.

Fudenberg, D. and J. Tirole (1991), *Game Theory,* MIT Press.

Fuster, A., M. Plosser, P. Schnabl and J. Vickery (2019), "The role of technology in mortgage lending", *The Review of Financial Studies* 32(5): 1854–99.

Fuster, A., P. Goldsmith-Pinkham, T. Ramadorai and A. Walther (2022), "Predictably Unequal? The Effects of Machine Learning on Credit Markets", *The Journal of Finance* 77(1): 5-47.

Gabaix, X., D. Laibson, G. Moloche, and S. Weinberg (2006), "Costly Information Acquisition: Experimental Analysis of a Boundedly Rational Model", *American Economic Review* 96(4): 1043-1068.

Gambacorta, L. and V. Shreeti (2025), "The AI supply chain", BIS Paper 154.

Gambacorta, L., Y. Huang, Z. Li, H. Qiu, and S. Chen (2023), "Data vs. Collateral.", *Review of Finance* 27(2): 369–398.

Gambacorta, L., Y. Huang, H. Qiu and J. Wang (2024a), "How do machine learning and non-traditional data affect credit scoring? New evidence from a Chinese fintech firm", *Journal of Financial Stability* 73: 101284.

Gambacorta, L., B. Kwon, T. Park, P. Patelli and S. Zhu (2024b), "CB-LMs: language models for central banking", BIS Working Papers 1215.

Gambacorta, L., H. Qiu, D. Rees and S. Shan (2024c), "Generative AI and labour productivity: a field experiment on coding", BIS Working Paper 1208.

Gambacorta, L., F. Sabatini and S. Schiaffi (2025), "Artificial intelligence and relationship lending", Bank of Italy Working Paper 1476.

Ganong, P. and P. Noel (2023), "Why do borrowers default on mortgages?", *The Quarterly Journal of Economics* 138(2): 1001–1065.

Gans, J. (2024), "Market power in Artificial Intelligence", mimeo.

Gao, M. and J. Huang (2020), "Informing the market: The effect of modern information technologies on information production", *Review of Financial Studies* 33: 1367-1411.

Gartner (2024), "Gartner says worldwide IaaS public cloud services revenue grew 16.2% in 2023", 22 July.

Georges, C. and J. Pereira (2021), "Market stability with machine learning agents", *Journal of Economic Dynamics and Control* 122: 104032.

Ghosh, P., B. Vallee and Y. Zeng (2024), "Fintech lending and cashless payments", *The Journal of Finance*, forthcoming.

Glode, V., R.C. Green and R. Lowery (2012), "Financial Expertise as an Arms Race", *The Journal of Finance* 67: 1723-1759.

Goldman Sachs (2018), "Liquidity, Volatility, Fragility", *Top of Mind* 68, 12 June.

Goldstein, I. (2023), "Information in financial markets and its real effects", *Review of Finance* 27(1):1–32.

Goldstein, I. and L. Yang (2015), "Information Diversity and Complementarities in Trading and Information Acquisition", *The Journal of Finance* 70: 1723-1765.

Gopinath, G. (2024), "Crisis Amplifier? How to Prevent AI from Worsening the Next Economic Downturn, speech at the AI for Good Global Summit, 30 May.

Goshen, Z. and G. Parchomovsky (2005), "The essential role of securities regulation", *Duke Law Journal* 55: 711-782.

Gray, N. and D. Jones (2025), "Transforming survey analysis: tools for central banks", 4th IFC and Bank of Italy Workshop on Data Science in Central Banking, 18-20 February.

Green, C., R. Huang, Q. Wen and D. Zhiu (2019), "Crowdsourced employer reviews and stock returns", *Journal of Financial Economics* 134: 236-251.

Greenwich Associates (2017), "Alternative data for alpha" (https://www.greenwich.com/equities/alternative-data-alpha).

Grennan, J. and R. Michaely (2021), "FinTechs and the Market for Financial Analysis", *Journal of Financial and Quantitative Analysis* 56: 1877-1907.

Grossman, S.J. and O.D. Hart (1986), "The costs and benefits of ownership: A theory of vertical and lateral integration", *Journal of Political Economy* 94(4): 691–719.

Gu, S., B. Kelly and D. Xiu (2020), "Empirical Asset Pricing via Machine Learning", *The Review of Financial Studies* 33: 2223–2273

Hacker, P. (2023), "Manipulation by algorithms. Exploring the triangle of unfair commercial practice, data protection, and privacy law", *European Law Journal* 29: 142–175.

Hafsi, H. and V. Vittori (2024), "Optimal Execution with reinforcement learning", https://arxiv.org/abs/2411.06389.

Hambly, B.M., R. Xu and H. Yang (2023), "Recent advances in reinforcement learning in finance", working paper, University of Oxford.

Han, Z. (2024), "Does AI adoption in M&A teams improve deal performance?", working paper, Emory University.

Hart, O. (1995), *Firms, Contracts, and Financial Structure*, Clarendon Press.

Hart, O. and J. Moore (1988), "Incomplete contracts and renegotiation", *Econometrica* 56(4): 755–785.

Harvey, C.R. (2016), "Cryptofinance", available at SSRN.

Harvey, C.R., A. Ramachandran and J. Santoro (2021), *DeFi and the Future of Finance,* John Wiley & Sons.

Hau, H., Y. Huang, C. Lin, H. Shan, Z. Sheng and L. Wei (2024), "FinTech credit and entrepreneurial growth", *The Journal of Finance* 79(5): 3309-60.

Healy, P.M. and K.G. Palepu (2001), "Information asymmetry, corporate disclosure, and the capital markets: A review of the empirical disclosure literature", *Journal of Accounting and Economics* 31(1-3): 405–440.

Hendershott, T., C.M. Jones and A.J. Menkveld (2011), "Does algorithmic trading improve liquidity?", *Journal of Finance* 66: 1–33.

Hennessy, C.A. and C.A. Goodhart (2023), "Goodhart's law and machine learning: A structural perspective", *International Economic Review* 64(3): 1075–1086.

Hirshleifer, D. and S.H. Teoh (2003), "Limited Attention, Information Disclosure, and Financial Reporting", *Journal of Accounting and Economics* 36: 337-386.

Hoff, K.A. and M. Bashir (2015), "Trust in Automation: Integrating Empirical Evidence on Factors That Influence Trust", *Human Factors* 57(3): 407-434.

Huang, J. (2018), "The customer knows best: The investment value of consumer opinions", *Journal of Financial Economics* 128: 164-182.

Huang Y., L. Zhang, Z. Li, H. Qiu, T. Sun and X. Wang (2020), "Fintech credit risk assessment for SMEs: Evidence from China", IMF Working Paper 193.

Hué S., C. Hurlin, C. Perignon and S. Saurin (2023), "Measuring the Driving Forces of Predictive Performance: Application to Credit Scoring", arXiv.org, 2212.05866.

Hurlin, C., C. Perignon and S. Saurin (2024), "The fairness of credit scoring models", *Management Science,* forthcoming.

IIF-EY – Institute of International Finance and EY (2025), *IIF-EY Annual Survey Report on AI/ML Use in Financial Services: Public Summary,* January.

IMF – International Monetary Fund (2024), *Steadying the course: Uncertainty, artificial intelligence, and financial stability,* Global Financial Stability Report, October 2024.

Imperva (2024), *Bad Bot Report,* April.

IoT Analytics Research (2023), *Generative AI Market Report 2023–2030.*

Jagtiani, J. and C. Lemieux (2019), "The roles of alternative data and machine learning in fintech lending: evidence from the Lending Club consumer platform", *Financial Management* 48(4): 1009-1029.

Jame, R., R. Johnston, S. Markov and M. Wolfe (2016), "The value of crowdsourced earnings forecasts", *Journal of Accounting Research* 54: 1077–1109.

Jansen, M., H.Q. Nguyen and A. Shams (2024), "Rise of the machines: the impact of automated underwriting", *Management Science,* forthcoming.

Jensen, M. and W.H. Meckling (1976), 'Theory of the firm: Managerial behavior, agency costs and ownership structure", *Journal of Financial Economics* 3(4): 305–360.

Jiang, J., K. Sirnivasan and Y. Huang (2025a), "The impact of GenAI on content creation – evidence from music videos", available at SSRN.

Jiang, W., J. Park, R. Xiao and S. Zhang (2025b), "AI and the extended workday: Productivity, contracting efficiency, and distribution of rents", working paper, Emory University.

Jin, S. (2023), "More than words: Can tone of consumer product reviews help predict firms' fundamentals?", *Journal of Business Finance and Accounting* 50: 1910–1942.

John, K., L. Kogan and F. Saleh (2023), "Smart contracts and decentralized finance", *Annual Review of Financial Economics* 15: 523–542.

Johnson, J., A. Rhodes and M. Wildenbeest (2023), "Platforms design when sellers use pricing algorithms", *Econometrica* 91: 1841-1879.

JP Morgan (2019), *Alternative Data Handbook.*

Kamenetsky Yadan, S. (2021), "Anomaly detection methods and tools for big data", *IFC Bulletin* 57.

Katona Z, M.O. Painter, P.N. Patatoukas, and J. Zeng (2025), "On the Capital Market Consequences of Big Data: Evidence from Outer Space", *Journal of Financial and Quantitative Analysis* 60: 551-579.

Kelly, B. and D. Xiu (2023), "Financial Machine Learning", *Foundations and Trends® in Finance* 13: 205-363.

Khandani, A., A. Kim and A. Lo (2010), "Consumer credit-risk models via machine-learning algorithms", *Journal of Banking and Finance* 4(11): 2767–87.

Kim, A., M. Muhn and V. Nikolaev (2024), "Financial Statement Analysis with Large Language Models", Fama-Miller Working Paper, Chicago Booth.

Klein, T. (2021), "Autonomous algorithmic collusion: Q-learning under sequential pricing", *Rand Journal of Economics* 52: 538-599.

Korinek A. and A. Balwit (2024), "Aligned with Whom? Direct and Social Goals for AI Systems," in J. Bullock et al. (eds), *Oxford Handbook of AI Governance*, Oxford University Press.

Korinek, A. and J. Vipra (2024), "Concentrating intelligence: scaling and market structure in artificial intelligence", *Economic Policy* 40(121): 225-256.

Kozak, S., S. Nagel and S. Santosh (2020), "Shrinking the cross-section," *Journal of Financial Economics* 135: 271-292.

Kysucky, V. and L. Norden (2016), "The benefits of relationship lending in a cross-country context: A meta-analysis", *Management Science* 62(1): 90–110.

Lehar, A. and C.A. Parlour (2023), "Battle of the bots: Flash loans, miner extractable value and efficient settlement", available at SSRN.

Leland, H.E. and D.H. Pyle (1977), "Informational asymmetries, financial structure, and financial intermediation", *Journal of Finance* 32(2): 371–387.

Lessmann, S., B. Baesens, H. Seow and L. Thomas (2015), "Benchmarking state-of-the-art classification algorithms for credit scoring: An update of research," *European Journal of Operational Research* 247: 124-136.

Levi, S.D. and A.B. Lipton (2018), "An introduction to smart contracts and their potential and inherent limitations", Harvard Law School Forum on Corporate Governance, 26 May.

Levy, G. and R. Razin (2015) "Correlation Neglect, Voting Behavior, and Information Aggregation", *American Economic Review* 105(4): 1634–1645.

Levy, G., I.M. de Barreda and R. Razin (2022), "Persuasion with Correlation Neglect: A Full Manipulation Result", *American Economic Review: Insights* 4(1): 123–138

Lewis, M. (2014), *Flash boys: a Wall Street revolt,* WW Norton & Company.

Liu, Y., and H. Wang (2024), "Who on Earth Is Using Generative AI?", Policy Research Working Paper 10870, World Bank.

Liu, L., G. Lu and W. Xiong (2022): "The big tech lending model", NBER Working Paper 30160.

Liu, J., I. Makarov and A. Schoar (2023), "Anatomy of a run: The Terra Luna crash. Technical report", NBER Working Paper 31160.

Lopez-Lira, A. and Y. Tang (2023), "Can ChatGPT forecast stock price movements? Return predictability and large language models", available at SSRN.

Loughran, T. and B. McDonald (2011), "When is a liability not a liability? Textual analysis, dictionaries, and 10-Ks", *Journal of Finance* 66(1): 35–65.

Lumsdaine, R.L. and R.J.D. Potter van Loon (2018), "Do Survey Probabilities Match Financial Market Beliefs?", *Journal of Behavioral Finance* 19(2): 209-220.

Lyonnet, V. and L. Stern (2024), "Machine learning about venture capital choices," available at SSRN.

MacKay, A. and S. Weinstein (2022), "Dynamic pricing algorithms, consumer harm, and regulatory response," Working Paper 22-050, Harvard Business School.

Makarov, I. and A. Schoar (2022), "Cryptocurrencies and decentralized finance (DeFi)", NBER Working Paper 30006.

Matz, S.C., J.D. Teeny, S.S. Vaid et al. (2024), "The potential of generative AI for personalized persuasion at scale," *Scientific Reports* 14: 4692.

Mercer (2024), *AI Integration in Investment Management.*

Mester, L.J., L.I. Nakamura and M. Renault (2007), "Transactions accounts and loan monitoring", *Review of Financial Studies* 20(3): 529–556.

Morris, M.R., J. Sohl-Dickstein, N. Fiedel, T. Warkentin, A. Dafoe, A. Faust, C. Farabet, and S. Legg (2024), "Levels of AGI: Operationalizing Progress on the Path to AGI", in *Proceedings of the 41st International Conference on Machine Learning.*

Müller, K. and E. Verner (2024), "Credit allocation and macroeconomic fluctuations", *The Review of Economic Studies* 91(6): 3645–76.

Murray, S., Y. Xia and H. Xiao (2024), "Charting by Machines", *Journal of Financial Economics* 153: 1037-1091.

Myers, S.C. and N.S. Majluf (1984), "Corporate financing and investment decisions when firms have information that investors do not have", *Journal of Financial Economics* 13(2): 187–221.

Nagel, S. (2021), *Machine Learning in Asset Pricing,* Princeton University Press.

Narang, N. (2013), *Inside the Black Box,* Wiley.

NIST – National Institute of Standards and Technology (2023), "Artificial Intelligence Risk Management Framework (AI RMF 1.0)".

Nordhaus, W. (2021), "Are We Approaching an Economic Singularity? Information Technology and the Future of Economic Growth", *American Economic Journal: Macroeconomics* 13: 299–332.

Noy, S. and W. Zhang (2023), "Experimental evidence on the productivity effects of generative artificial intelligence", *Science* 381(6654): 187-192.

OECD – Organisation for Economic Co-operation and Development (2017), *Algorithms and collusion: Competition policy in the digital age,* OECD Publishing.

OECD (2021), *Artificial Intelligence, Machine Learning and Big Data in Finance: Opportunities, Challenges, and Implications for Policy Makers*, OECD Publishing.

OECD (2024), "Miracle or myth? Assessing the macroeconomic productivity gains from Artificial Intelligence", OECD Artificial Intelligence Papers, 29, November.

Ofcom (2023), *Cloud services market study,* Final report.

O'Hara, M. (2014), "High-frequency trading and its impact on markets", *Financial Analysts Journal* 70(3): 18-27.

OpenAI (2025), "Operator" (https://openai.com/index/introducing-operator/).

Ornelas, J. and A. Pecora (2022), "Does fintech lending lower financing costs? Evidence from an emerging market", Banco Central Do Brazil Working Paper 571.

Pagano, M. and T. Jappelli (1993), "Information Sharing in Credit Markets", *Journal of Finance* 48(5): 1693–18.

Pagnotta, E.S. and T. Philippon (2018), "Competing on Speed", *Econometrica* 86: 1067-1115.

Parton, M., Angelone, M., Metta, C., et al. (2022), "Artificial intelligence and renegotiation of commercial lease contracts affected by pandemic-related contingencies from covid-19. the project a.i.a.co", arXiv:2210.09515.

Peng, S., W. Swiatek, A. Gao, P. Cullivan, and H. Chang, H (2024), "AI Revolution on Chat Bot: Evidence from a Randomized Controlled Experiment", arXiv preprint, arXiv:2401.10956.

Peress, J. (2014), "Learning from stock prices and economic growth", *Review of Financial Studies* 27: 2998-3059.

Philippon, T. (2015), "Has the US finance industry become less efficient? On the theory and measurement of financial intermediation", *American Economic Review* 105: 1408-1438.

Phillipon, T. (2020), "On Fintech and Financial Inclusion", BIS Working Paper 841.

Rajan, R.G. (1992), "Insiders and outsiders: The choice between informed and arm's-length debt", *Journal of Finance* 47(4): 1367–1400.

Rampini, A.A. and S. Viswanathan (2013), "Collateral and capital structure", *Journal of Financial Economics* 109(2): 466–492.

Reimann, M., O. Schilke, and K.S. Cook (2017), "Trust is Heritable, Whereas Distrust is Not", *Proceedings of the National Academy of Sciences* 114(27): 7007-7012.

Retterath, A. (2020), "Human Versus Computer: Benchmarking Venture Capitalists and Machine Learning Algorithms for Investment Screening", available at SSRN.

Ross, S. (1973), "The economic theory of agency: The principal's problem", *American Economic Review* 63: 134–139.

Ross, S. (1977), "The determination of financial structure: The incentive-signalling approach", *Bell Journal of Economics* 8: 23–40.

Rossi, A. and S. Utkus (2024), "The diversification and welfare effects of robo-advising", *Journal of Financial Economics* 157: 1038-1069.

Rubio, J., P. Barucca, G. Gage, J. Arroyo and R. Morales-Resendiz (2021), "Classifying payment patterns with artificial neural networks: an autoencoder approach", *IFC Bulletin* 57.

Russell, S. (2020), *Human Compatible: Artificial Intelligence and the Problem of Control*, Penguin Books.

Sabetti, L. and R. Heijmans (2020), "Shallow or deep? Detecting anomalous flows in the Canadian automated clearing and settlement system using an autoencoder", De Nederlandsche Bank Working Paper 681.

Scheurer, J., M. Balesni, and M. Hobbhahn (2023), "Technical report: Large language models can strategically deceive their users when put under pressure", arXiv preprint, arXiv:2311.07590.

SEC – US Securities and Exchange Commission (2023), "Conflicts of Interest Associated with the Use of Predictive Data Analytics by Broker-Dealers and Investment Adviser", Release No. 34-97990.

Sheng, J., Z. Sun, B. Yang and A. Zhang (2024), "Generative AI and asset management", working paper, University of California, Irvine.

Shiller, R.J. (1988), "Portfolio insurance and other investor fashions as factors in the 1987 stock market crash", *NBER Macroeconomics Annual* 3: 287–297.

Shkilko, A. and K. Sokolov (2020), "Every Cloud Has a Silver Lining: Fast Trading, Microwave Connectivity, and Trading Costs", *The Journal of Finance* 75: 2899-2927.

Stiglitz, J. (1994), “The role of the state in financial markets”, in *Proceedings of the World Bank Annual Conference on Development Economics.*

Stiglitz, J.E. and A. Weiss (1981), "Credit rationing in markets with imperfect information", *American Economic Review* 71: 393–410.

Stiglitz, JE. and A. Weiss (1988), "Banks as social accountants and screening devices for the allocation of credit", NBER Working Paper 2710.

Suleyman, M. and M. Bhaskar (2023), *The Coming Wave: Technology, Power, and the Twenty-first Century's Greatest Dilemma*, Crown.

Sundaresan, S. and Z. Wang (2015), “On the design of contingent capital with a market trigger”, *Journal of Finance* 70(2): 881–920.

Sutton, R., and A. Barto (2019), *Reinforcement learning*, MIT Press.

Szabo, N. (1994), “Smart contracts”, working paper, University of Amsterdam.

Tan, J., C.K. Shum and M. Akmal Amri (2021), “Supervisory letter writing app: expediting letter drafting and tone consistency”, *IFC Bulletin* 57.

Tang, V.W. (2018), “Wisdom of crowds: Cross-sectional variation in the informativeness of third-party-generated product information on Twitter”, *Journal of Accounting Research* 56: 989–1034.

Tang, H. (2019), “Peer-to-peer lenders versus banks: substitutes or complements?”, *Review of Financial Studies* 32(5): 1900-1938.

Timmermann, A. (2006), “Forecast Combinations”, in G. Elliott, C.W.J. Granger and A. Timmermann (eds), *Handbook of Economic Forecasting*, Vol. I, Elsevier.

US Department of the Treasury (2016), “Opportunities and challenges in online marketplace lending”, May.

van Binsbergen, J., X. Han and Lopez-Lira (2023), “Man versus Machine Learning: The Term Structure of Earnings Expectations and Conditional Biases”, *The Review of Financial Studies* 36: 2361–2396.

Verrecchia, R.E. (1982), “Information acquisition in a noisy rational expectations economy”, *Econometrica* 50: 1415–1430.

Vives, X. (2019), “Digital disruption in Banking”, *The Annual Review of Financial Economics* 11: 243-272.

Vives, X. and Z. Ye (2025a), “Information technology and lender competition”, *Journal of Financial Economics* 163, 103957

Vives, X. and Z. Ye (2025b), “Fintech entry, lending market competition, and welfare”, *Journal of Financial Economics*, forthcoming.

Wah, E. (2016), "How Prevalent and Profitable are Latency Arbitrage Opportunities on U.S. Stock Exchanges?", Available at SSRN.

Wang, X., R.J. Hyndman, F. Li, and Y. Kang (2023), "Forecast Combinations: An Over 50-Year Review", *International Journal of Forecasting* 39: 1518-1547.

Watkins, C. and P. Dayan (1992), "Q-learning", *Machine Learning* 8: 279-292.

Weller, B. (2018), "Does Algorithmic Trading Reduce Information Acquisition?", *The Review of Financial Studies* 31: 2184–2226

Woodward, M. (2018), "Bumping up the competition: The influence of IEX's speed bump in US financial markets", available at SSRN.

Yampratoom, E. (2024), "Experiences from implementing business use cases with generative AI", Central Banking Technology Leadership Forum, 4-5 December.

Yermack, D. (2017), "Corporate governance and blockchains", *Review of Finance* 21(1): 7–31.

Yi, H.C., D. Choi and Y. Kim (2022), "Dynamic factor model and deep learning algorithm for GDP nowcasting", *Bank of Korea Economic Analysis* 28(2).

Yu, H., X. Hao, L. Wu, Y. Zhao and Y. Wang (2023), "Eye in outer space: satellite imageries of container ports can predict world stock return", *Humanities and Social Sciences Communications* 10: 1–16.

Zhang, Y. (2024), "Do Mutual Funds Benefit from the Adoption of AI Technology?", HKUST Business School Research Paper No. 2024-165.

Zhu, C. (2019), "Big data as a governance mechanism", *The Review of Financial Studies* 32(5): 2021–2061.

CEPR PRESS

The seventh report in The Future of Banking series, part of the Banking Initiative at IESE Business School, examines the fundamental transformations induced by artificial intelligence and the policy challenges it raises. It focuses on three main themes: the use of AI in financial intermediation, central banking and policy, and regulatory challenges; the implications of data abundance and algorithmic trading for financial markets; and the effects of AI on corporate finance, contracting, and governance. Across these domains, the report emphasises that while AI has the potential to improve efficiency, inclusion, and resilience, it also poses new vulnerabilities that call for adaptive regulatory responses.

ISBN: 978-1-912179-93-0

187 boulevard Saint-Germain | 75007 PARIS | FRANCE
2 Coldbath Square | LONDON EC1R 5HL | UK
TEL: +44 (0)20 7183 8801 | FAX: +44 (0)20 7183 8820
EMAIL: CEPR@CEPR.ORG
WWW.CEPR.ORG