Regulating Artificial Intelligence in Finance: Putting the Human in the Loop

Ross P Buckley,* Dirk A Zetzsche,† Douglas W Arner‡ and Brian W Tang§

Abstract

This article develops a framework for understanding and addressing the increasing role of artificial intelligence ('AI') in finance. It focuses on human responsibility as central to addressing the AI 'black box' problem — that is, the risk of an AI producing undesirable results that are unrecognised or unanticipated due to people's difficulties in understanding the internal workings of an AI or as a result of the AI's independent operation outside human supervision or involvement. After mapping the various use cases of AI in finance and explaining its rapid development, we highlight the range of potential issues and regulatory challenges concerning financial services AI and the tools available to address them. We argue that the most effective regulatory approaches to addressing the role of AI in finance bring humans into the loop through personal responsibility regimes, thus eliminating the black box argument as a defence to responsibility and legal liability for AI operations and decisions.

* KPMG Law and King & Wood Mallesons Chair of Disruptive Innovation, Australian Research Council Laureate Fellow, Scientia Professor, and Member, Centre for Law, Markets and Regulation, UNSW Sydney, Australia; Chair, Digital Finance Advisory Panel of the Australian Securities and Investments Commission. The views expressed herein are strictly my own. Email: ross.buckley@unsw.edu.au.

† Professor of Law, ADA Chair in Financial Law (Inclusive Finance), Faculty of Law, Economics and Finance, University of Luxembourg, Luxembourg; Director, Centre for Business and Corporate Law, Heinrich-Heine-University, Düsseldorf, Germany. Email: dirk.zetzsche@uni.lu; ORCID iD: https://orcid.org/0000-0003-4051-7699.

‡ Kerry Holdings Professor in Law and Director, Asian Institute of International Financial Law, Faculty of Law, University of Hong Kong, Hong Kong; Advisory Board Member, Centre for Finance, Technology and Entrepreneurship. Email: douglas.arner@hku.hk; ORCID iD: http://orcid.org/0000-0002-6910-7311.

§ Founding Executive Director, Law, Innovation, Technology and Entrepreneurship Lab (LITE Lab@HKU), Faculty of Law, University of Hong Kong, Hong Kong; Co-chair of the FinTech Association of Hong Kong's RegTech Committee, Founder, Asia Capital Market Institute (ACMI) and member of the IEEE Global Initiative in Autonomous and Intelligent Systems' Policy Committee. Email: brian.tang@asiacmi.com.

The authors thank KPMG Law and King & Wood Mallesons, the Hong Kong Research Grants Council Research Impact Fund, and the Qatar National Research Fund National Priorities Programme for financial support, Maria Lucia Passador for comments, and Maria Lai for her invaluable research assistance. All responsibility is the authors'.

I Introduction

Artificial intelligence ('AI') is a focus of global attention.[1] While AI has a long history of development, recent technological advances and digitisation have underpinned rapid and unprecedented evolution. Algorithmic trading, an early and leading AI use case, has, in the words of the European Central Bank, 'been growing steadily since the early 2000s and, in some markets, is already used for around 70% of total orders'.[2] In 2019, a major official survey of the United Kingdom ('UK') financial services industry revealed that machine learning — a form of AI — is used by two-thirds of respondents in the UK in a range of front-office and back-office applications, most commonly in anti-money laundering, fraud detection, customer services and marketing.[3] A similar survey in Hong Kong revealed that 89% of banks had adopted or planned to adopt AI applications, most often in cybersecurity, client-facing chatbots, remote onboarding, and biometric customer identification.[4] Central to this is the rise of datafication — manipulation of digitised data through quantitative data analytics, including AI.[5]

In most sectors, AI is expected to contribute to problem solving and development. Pricewaterhouse Coopers ('PwC'), probably optimistically, expects AI will boost global gross domestic product by 14% by 2030.[6] Accenture estimates that banks can save 20–25% across information technology ('IT') generally.[7] Cost savings, enhanced efficiency, entirely new opportunities and business models explain why financial services companies are expected to spend US$11 billion on AI in 2020, more than any other industry.[8]

A 2018 World Economic Forum ('WEF') report highlighted that AI-enabled systems in finance can deliver 'new efficiencies' and 'new kinds of value'.[9] However, a tight focus on these new capabilities risks overlooking a fundamental shift as financial institutions become 'more specialised, leaner, highly networked and dependent on the capabilities of a variety of technology players'.[10] The WEF suggests multiple stakeholder collaboration is required to counter the potential social

---

1 See generally Bonnie G Buchanan, *Artificial Intelligence in Finance* (The Alan Turing Institute, Report, April 2019).

2 European Central Bank, 'Algorithmic Trading: Trends and Existing Regulation' (Media Release, 13 February 2019). In Part IV(B) below, we discuss Joint Committee of the European Supervisory Authorities ('ESAs'), *Joint Committee Final Report on Big Data* (JC/2018/04, 15 March 2018) ('*ESAs Joint Committee Final Report on Big Data*').

3 Bank of England and Financial Conduct Authority, *Machine Learning in UK Financial Services* (Report, October 2019) 8.

4 Hong Kong Monetary Authority, 'Artificial Intelligence (AI) in Retail Banking' (Fact Sheet, November 2019).

5 UK Finance and Microsoft, *Artificial Intelligence in Financial Services* (Report, June 2019) 5.

6 PwC, *Sizing the Prize: What's the Real Value of AI for Your Business and How Can You Capitalise?* (Report, 2017) 4.

7 Accenture, *Redefine Banking with Artificial Intelligence* (Report, 2018) 9.

8 Amy Zirkle, 'The Critical Role of Artificial Intelligence in Payments Tech', *Fintech News* (online, 27 May 2019) <https://www.fintechnews.org/the-cririal-role-of-artificial-inteliigence-in-payments-tech/>.

9 World Economic Forum, *The New Physics of Financial Services: Understanding How Artificial Intelligence is Transforming the Financial Ecosystem* (Report, 15 August 2018) 18.

10 Ibid 19.

and economic risks of AI-enabled systems in finance.[11] Similarly, in 2019, the WEF addressed responsible AI use in finance, focusing on governance requirements and risks. Specifically, AI explainability, systemic risk, AI biases, and algorithmic collusion have been identified as prominent sources of risk in finance.

AI and automation raise major broader concerns, ranging from widespread job losses,[12] to 'the singularity' — when the capacities of AI surpass those of humans. These concerns have triggered many analyses of the ethical[13] and legal[14] implications of AI, yet few from the perspective we take here, of AI's impact in finance.[15]

Central to many of these concerns is the role of humans in the evolution of AI: the necessity of involving people in using, monitoring and supervising AI. This article develops a regulatory framework for understanding and addressing the increasing role of AI in finance. It focuses on human responsibility, the 'human in the loop', as central to tackling the AI 'black box' problem, that is: the risk that AI results in processes and operations unknown to and uncontrolled by human beings, producing undesirable results for which, arguably, only the AI may be responsible.

Part II maps the various use cases of AI in finance, and explains its rapid development. Part III highlights the risks the increasing reliance on AI in finance creates. Part IV summarises the regulatory challenges concerning financial services AI and the tools available to address them, highlighting the need to address the black box problem.

Part V presents our solution to the black box problem. We argue that the most effective regulatory approach is to bring humans into the loop, enhancing *internal* governance where financial supervision as *external* governance is unlikely to be effective. We thus propose to address AI-related issues by requiring three internal governance tools: (1) AI due diligence; (2) AI explainability; and (3) AI review committees. These tools would operate both directly and via the mechanism of personal responsibility embedded in an increasing range of financial regulatory systems, including in Australia, the European Union ('EU'), and the UK.

Part VI concludes, suggesting that this framework offers the potential to address black box issues in the context not only of AI in finance, but also in any regulated industry.

---

11 World Economic Forum (n 9) 51. See also UK Finance and Microsoft (n 5) 15.

12 Shelly Hagan, 'More Robots Mean 120 Million Workers Need to be Retrained', *Bloomberg* (online, 6 September 2019) <https://www.bloomberg.com/news/articles/2019-09-06/robots-displacing-jobs-means-120-million-workers-need-retraining>.

13 See generally Dirk Helbing, 'Societal, Economic, Ethical and Legal Challenges of the Digital Revolution: From Big Data to Deep Learning, Artificial Intelligence, and Manipulative Technologies' in Dirk Helbing (ed), *Towards Digital Enlightenment — Essays on the Dark and Light Sides of the Digital Revolution* (Springer, 2018) 47.

14 See generally Harry Surden, 'Machine Learning and the Law' (2014) 89(1) *Washington Law Review* 87.

15 Tom CW Lin, 'Artificial Intelligence, Finance, and the Law' (2019) 88(2) *Fordham Law Review* 531.

II AI and Finance

The term AI covers a series of technologies and approaches, ranging from 'if-then' rule-based expert systems,[16] to natural language processing, to the marriage of algorithms and statistics known as machine learning. Machine learning involves pattern recognition and inference trained by data rather than explicit human instructions. It progressively reduces the role of humans as AI systems expand from supervised learning to unsupervised deep learning neural networks.

A *Technical Preconditions for AI in Finance*

AI has existed since the 1970s. However, five key factors have empowered the rapid evolution of AI in the last decade: data, storage, communication, computing power, and analytics.

The role of data has been transformed by digitisation. Once data are available digitally, datafication — the application of analytics including AI — becomes effective. Thus, the 'digitisation of everything', which underpins the idea of the 'Fourth Industrial Revolution', is central to the rapid evolution of AI.[17] Larger volumes of data and datafication improve machine learning processes and the 'training' of AI systems.

Meanwhile, in line with Kryder's law, data storage quality and capacity have dramatically increased and costs decreased, resulting in ever-increasing volumes of digitally captured and stored data.[18] The internet, mobile phones and the internet of things make it easier to capture, store, manipulate, and analyse data. Further, many cloud-connected devices effectively provide unlimited data collection and storage capacity.

Computing power has also increased dramatically following Moore's Law: that the number of transistors on a microchip doubles every two years.[19] If realised, the emergence of quantum computing will open incredible new avenues of processing. Datafication also benefits from the rapid innovations in algorithms and analytical processes.

Ever-falling storage prices including cloud ubiquity, telecommunications linkages, ever-increasing computing power, and innovative algorithmic and analytical development underlie the explosion in datafication processes. This, in turn, fuels AI growth that looks set to continue, to the extent where discussions of the singularity are no longer the realm of science fiction.

---

16 In rule-based expert systems, knowledge is represented as a set of rules. For example, IF 'traffic light' is 'green' THEN the action is go: see Jiri Panyr, 'Information Retrieval Techniques in Rule-based Expert Systems' in Hans-Hermann Bock and Peter Ihm (eds), *Classification, Data Analysis, and Knowledge Organization* (Springer, 1991) 196.

17 Klaus Schwab, 'The Fourth Industrial Revolution', *World Economic Forum* (online, 14 January 2016) <https://www.weforum.org/agenda/2016/01/the-fourth-industrial-revolution-what-it-means-and-how-to-respond/>.

18 Chip Walter, 'Kryder's Law' (2005) 293(2) *Scientific American* 32.

19 Gordon E Moore, 'Cramming More Components onto Integrated Circuits' (1965) 38(8) *Electronics* 114.

B *AI in Financial Services*

While financial services have always integrated technical innovation,[20] the trend is more pronounced in the latest wave of financial technology ('fintech') innovation. Financial services, in particular, is a fertile field for the application of AI.

A major pillar of recent digital financial transformation is the large-scale use of data. Finance has long cultivated the extensive structured collection of many forms of data (for example, stock prices). Such data have been standardised and digitised since the 1970s, with new forms of capture and collection constantly emerging.

Furthermore, AI tends to perform best in rule-constrained environments, such as games like chess or Go, where there are finite ways of achieving specified objectives. In this environment, AI is outperforming humans with increasing rapidity. This is often the environment in finance — for example, stock market investment involves specific objectives (maximising profit), fixed parameters of action (trading rules and systems) and massive amounts of data. Adding technological possibility to the financial and human resources and incentives explains why finance is already transforming so rapidly as a result of digitisation and datafication, and is likely to continue.

The financial resources, human resources and incentives are clear: financial intermediaries generate massive amounts of income for their stakeholders, including management, investors and employees. Accordingly, they attract some of the very best human resources. Those human and financial resources have strong reasons to continually search for advantages and opportunities for profit, and thus invest substantially in research, analytics and technology, to the extent that an entire academic field — finance — focuses on research in the area along with major teams at financial institutions and advisory firms. This makes finance unique from an AI perspective.

Due to ever-improving performance in data gathering, processing, and analytics, AI increasingly affects all operational and internal control matters of financial intermediaries, from strategy setting,[21] to compliance,[22] to risk management and beyond.[23]

---

[20] Douglas W Arner, Jànos Barberis and Ross P Buckley, 'The Evolution of FinTech: A New Post-Crisis Paradigm?' (2016) 47(4) *Georgetown Journal of International Law* 1271.

[21] John Armour and Horst Eidenmüller, 'Self-driving Corporations?' (2020) 10(1) *Harvard Business Law Review* 87, 96–7.

[22] Kenneth A Bamberger, 'Technologies of Compliance: Risk and Regulation in a Digital Age' (2010) 88(4) *Texas Law Review* 669, 690–93, 701–2.

[23] Saqib Aziz and Michael Dowling, 'Machine Learning and AI for Risk Management', in Theo Lynn, John G Mooney, Pierangelo Rosati and Mark Cummins (eds), *Disrupting Finance: FinTech and Strategy in the 21st Century* (Palgrave, 2019) 33.

C *AI Use Cases*

For that reason, algorithms and AI are frequently used on the front-end or back-end of an increasing range of processes and functions in finance.[24] AI use cases span a range of customer processes from on-boarding to instant responses to credit applications,[25] and also include operations and risk management,[26] trading and portfolio management,[27] payments and infrastructure,[28] data security and monetisation,[29] and regulatory and monetary oversight and compliance.[30]

Skyrocketing costs of compliance and sanctions have induced financial institutions to focus on back-office AI-solutions, in the form of regulatory technology ('regtech'). Regtech solutions include Amazon Alexa-like voice bots used for compliance queries,[31] and bots to review commercial loan contracts performing reportedly the equivalent of 360,000 hours of work each year by lawyers and loan officers.[32] AI is being applied to equities trade execution for maximum speed at best price,[33] post-trade allocation requests,[34] and to calculate policy payouts.[35] AI also drives the trend to seek alternative data for investment and lending decisions,[36] prompting the mantra 'all data is credit data'.[37]

---

24 Hong Kong Monetary Authority and PwC, *Reshaping Banking with Artificial Intelligence* (Report, December 2019) 33; Bank of England and Financial Conduct Authority (n 3) 4.

25 Accenture (n 7) 13, 15, 17.

26 Buchanan (n 1) 2; Financial Stability Board, *Artificial Intelligence and Machine Learning in Financial Services* (Report, November 2017) 16; Oliver Wyman and China Securities Credit Investment Company, *China Credit-tech Market Report: Technology-Driven Value Generation in Credit-Tech* (Report, 2019) 11, 13; Dirk A Zetzsche, William A Birdthistle, Douglas W Arner and Ross P Buckley, 'Digital Finance Platforms: Towards A New Regulatory Paradigm' (2020) 23(1) *University of Pennsylvania Journal of Business Law* 273, 298 ('Digital Finance Platforms').

27 Financial Stability Board (n 26) 15; Tom Baker and Benedict Dellaert, 'Regulating Robo Advice across the Financial Services Industry' (2018) 103(2) *Iowa Law Review* 713; Andrei A Kirilenko and Andrew W Lo, 'Moore's Law versus Murphy's Law: Algorithmic Trading and its Discontents' (2013) 27(2) *Journal of Economic Perspectives* 51.

28 Zirkle (n 8).

29 Buchanan (n 1) 2; Financial Stability Board (n 26) 21.

30 Financial Stability Board (n 26) 19–20; Okiriza Wibisono, Hidayah Dhini Ari, Anggraini Widjanarti, Alvin Andhika Zulen and Bruno Tissot, 'The Use of Big Data Analytics and Artificial Intelligence in Central Banking' (IFC Bulletin No 50, May 2019).

31 Olivia Oran, 'Credit Suisse has Deployed 20 Robots within Bank, Markets CEO Says', *Reuters* (online, 2 May 2017) <https://www.reuters.com/article/milken-conference-creditsuisse-idCNL1N1I31PJ>.

32 Hugh Son, 'JPMorgan Software Does in Seconds What Took Lawyers 360,000 Hours', *Bloomberg* (online, 28 February 2017) <https://www.bloomberg.com/news/articles/2017-02-28/jpmorgan-marshals-an-army-of-developers-to-automate-high-finance>.

33 Laura Noonan, 'JPMorgan Develops Robot to Execute Trades', *Financial Times* (online, 31 July 2017) <https://www.ft.com/content/16b8ffb6-7161-11e7-aca6-c6bd07df1a3c>.

34 Martin Arnold and Laura Noonan, 'Robots Enter Investment Banks' Trading Floors', *Financial Times* (online, 7 July 2017) <https://www.ft.com/content/da7e3ec2-6246-11e7-8814-0ac7eb84e5f1>.

35 Kevin Lui, 'This Japanese Company is Replacing its Staff with Artificial Intelligence', *Fortune* (online, 6 January 2017) <https://fortune.com/2017/01/06/japan-artificial-intelligence-insurance-company/>.

36 Anthony Malakian, 'AI and Alternative Data: A Burgeoning Arms Race', *WatersTechnology* (online, 20 June 2017) <https://www.waterstechnology.com/trading-tools/3389631/ai-and-alternative-data-a-burgeoning-arms-race>; Zetzsche et al (n 26).

37 Mikella Hurley and Julius Adebayo, 'Credit Scoring in the Era of Big Data' (2016) 18(1) *Yale Journal of Law and Technology* 148, 151.

AI's potential to process data, seemingly without human bias, is central to its utility. First, AI treats past data with the same precision as more recent data; in contrast, humans tend to overly prioritise more recent data. Second, correctly programmed AI treats all data objectively, while humans tend to discriminate among datapoints based on their experience, values and other non-rational judgements. AI can be unbiased in not following its own agenda or having cognitive biases.[38] Yet, the nature of AI creates other risks.

III Risks: AI in Finance

Analysed in terms of traditional financial regulatory objectives,[39] major AI-related risks arise in data, financial stability, cybersecurity, law and ethics.[40] We deal with each in turn.

A *Data Risks*

Key functions of AI include data collection, data analysis, decision-making and the execution of those decisions.[41] Not all AI technology performs all of these functions, and their use varies across industries and different areas of finance. Nonetheless, the centrality of data to the deployment of any useful AI model cannot be overstated,[42] and it is therefore necessary to analyse the risks created by data-dependent functions.

1 *AI Data Collection and Analysis*

Data collection has long been a major bottleneck in machine learning, for two reasons.[43] First, data collection is expensive. Second, large providers of data collection and analysis services may be unwilling to share data they have with other providers, which may sell the data or become future competitors of the data originator — one of the problems open banking is designed to address.[44] Data availability therefore intersects with data privacy and protection.

---

38 Sergio Gramitto Ricci, 'The Technology and Archeology of Corporate Law', *LawArXiv* (16 August 2018) 37–8 <https://doi.org/10.31228/osf.io/zcqn7>.

39 Douglas W Arner, *Financial Stability, Economic Growth and the Role of Law* (Cambridge University Press, 2007).

40 The French prudential regulatory authority, Autorité de Contrôle Prudentiel et de Résolution ('ACPR'), identified four similar risk categories, in ACPR, *Artificial Intelligence: Challenges for the Financial Sector* (Report, December 2018).

41 See Dirk Nicolas Wagner, 'Economic Patterns in a World with Artificial Intelligence' (2020) 17(1) *Evolutionary and Institutional Economics Review* 111.

42 Henri Arslanian and Fabrice Fischer, *The Future of Finance: The Impact of FinTech, AI, and Crypto on Financial Services* (Springer, 2019) 167, 177; Accenture Federal Services, *AI: All About the Data* (Report, 2020) 4 <https://www.accenture.com/us-en/insights/us-federal-government/ai-all-about-data>.

43 Yuji Roh, Geon Heo and Steven Euijong Whang, 'A Survey on Data Collection for Machine Learning: A Big Data — AI Integration Perspective' (2019) *IEEE Transactions on Knowledge and Data Engineering* 1 <https://doi.org/10.1109/TKDE.2019.2946162>.

44 Treasury (Cth), *Review into Open Banking: Giving Customers Choice, Convenience and Confidence* (Report, December 2017) <https://treasury.gov.au/sites/default/files/2019-03/Review-into-Open-Banking-_For-web-1.pdf>.

The availability of highly structured machine-readable data — a major obstacle for AI advancements elsewhere, including healthcare[45] — has increased the ease of adoption of AI in many areas of financial services.[46] However, other challenges remain.

First, data quality may be poor. An oft-repeated example is the use of training data by Enron for compliance AI.[47] From a legal perspective, protected factors come under threat if AI discriminates based on factors, proxies for these factors, or other factors altogether, that describe little more than *a* part of social and financial relations within society. For instance, an algorithm that determines creditworthiness based on consistency of telephone use (rather than complete economic and financial data) may discriminate against members of religions who tend not to use their phones one day per week.[48] Quality is a pervasive problem.[49] As both the value of high-quality information and the threats posed by information gaps continue to grow, regulators should focus on the development of widely used and well-designed data standards.[50]

Second, the data may be biased, either from data selection issues ('dashboard myopia') or data reflecting biases in society at large.[51] Prejudiced decision-makers may mask their biases by wittingly or unwittingly using biased data.[52] Biased data could likewise be selected in efforts to enhance an executive's personal bonus or reduce organisational oversight.[53] For this reason, understanding the context of the data — when, where, and how it was generated — is critical to understanding its utility and potential risks.[54]

Of course, bad data will result in bad AI analysis, the age-old adage of 'garbage in, garbage out'.[55] Similarly, inappropriately or suboptimally selected AI model architecture and parameters can distort analysis.[56] For example, in 2019 it was

---

45 Moritz Lehne, Julian Sass, Andrea Essenwanger, Josef Schepers and Sylvia Thun, 'Why Digital Medicine Depends on Interoperability' (2019) 2 *npj Digital Medicine* 79:1–5, 1 <https://www.nature.com/articles/s41746-019-0158-1>.

46 Tapestry Networks and EY, *Data Governance: Securing the Future of Financial Services* (Report, January 2018) 17.

47 Luca Enriques and Dirk A Zetzsche, 'Corporate Technologies and the Tech Nirvana Fallacy' (2020) 72(1) *Hastings Law Journal* 55, 76.

48 Dirk A Zetzsche, Ross P Buckley, Douglas W Arner and Jànos Barberis, 'From FinTech to TechFin: The Regulatory Challenges of Data-Driven Finance' (2018) 14(2) *New York University Journal of Law and Business* 393, 417–18.

49 *See* Tadhg Nagle, Thomas C Redman and David Sammon, 'Only 3% of Companies' Data Meets Basic Quality Standards', *Harvard Business Review* (online, 11 September 2017) <https://hbr.org/2017/09/only-3-of-companies-data-meets-basic-quality-standards>.

50 Richard Berner and Kathryn Judge, 'The Data Standardization Challenge' in Douglas W Arner, Emilios Avgouleas, Danny Busch and Steven L Schwarcz (eds), *Systemic Risk in the Financial Sector: Ten Years after the Great Crash* (McGill-Queen's University Press, 2019) 135, 148–9.

51 Lin (n 15) 536–7.

52 Solon Barocas and Andrew D Selbst, 'Big Data's Disparate Impact' (2016) 104(3) *California Law Review* 671, 692.

53 Enriques and Zetzsche (n 47) 75–7.

54 Lin (n 15) 536.

55 Alberto Artasanchez and Prateek Joshi, *Artificial Intelligence with Python* (Packt Publishing, 2nd ed, 2020) 46.

56 Deloitte, *AI and Risk Management: Innovating with Confidence* (Report, 2018) 8.

revealed that two parameters in Deutsche Bank's anti-financial crime systems (in effect since 2010) were defined incorrectly, potentially allowing suspicious transactions to avoid detection.[57] Such deficiencies may expose financial services organisations to competitive harm, legal liability, or reputational damage.[58]

Finally, the process of cleaning data is typically very demanding in terms of human resources.[59]

2 *AI Decision-Making and the Execution of Decisions*

AI systems may perform similar calculations simultaneously, and one AI's decisions may influence the tasks performed by another. 'Herding' results when actors make use of similar models to interpret signals from the market.[60] Algorithms trading simultaneously in millisecond trading windows have caused extreme volatility events, referred to as 'flash crashes', when unexpected situations arise.[61] This has resulted in worldwide regulatory efforts that address algorithmic trading.[62]

Similar problems can arise with robo-advisors, where one AI may front-run another AI's recommendation. While risk management tools such as price limits and stop loss-commands (themselves algorithms) can mitigate some of the risks, these tools are costly and do not address all risks generated by multiple AI performing similar tasks.

The alternative to uncoordinated behaviour is more frightening: tacit collusion. If several self-learning algorithms discover that cooperation in capital markets is more profitable than competition, they could cooperate and manipulate information and prices to their own advantage. There is evidence for self-learning AI colluding in price setting.[63] Multiple AI colluding in financial markets pricing is likely.

---

57 Olaf Storbeck, 'Deutsche Bank Glitch Blocked Reporting of Suspicious Transactions', *Financial Times* (online, 22 May 2019) <https://www.ft.com/content/d537f416-7c71-11e9-81d2-f785092ab560>.

58 Fernanda Torre, Robin Teigland and Liselotte Engstam, 'AI Leadership and the Future of Corporate Governance: Changing Demands for Board Competence' in Anthony Larsson and Robin Teigland (eds) *The Digital Transformation of Labor: Automation, the Gig Economy and Welfare* (Routledge, 2020) 116, 127.

59 Thomas C Redman, 'If Your Data Is Bad, Your Machine Learning Tools Are Useless', *Harvard Business Review* (online, 2 April 2018) <https://hbr.org/2018/04/if-your-data-is-bad-your-machine-learning-tools-are-useless>.

60 World Economic Forum, *Navigating Uncharted Waters: A Roadmap to Responsible Innovation with AI in Financial Services* (Report, 23 October 2019) 62.

61 Buchanan (n 1) 6.

62 Kirilenko and Lo (n 27) 53–5.

63 Ariel Ezrachi and Maurice E Stucke, 'Artificial Intelligence and Collusion: When Computers Inhibit Competition' [2017] (5) *University of Illinois Law Review* 1775.

B *Financial Stability Risks*

In 2017, the Financial Stability Board analysed and summarised a broad range of possible financial stability implications of AI and machine learning.[64] The Board noted their substantial promise, contingent upon proper risk management. Its report stressed that oligopolistic or monopolistic players may surface as a result of additional third-party dependencies caused by 'network effects and scalability of new technologies'.[65] Some of these new market participants are currently unregulated and unsupervised.[66] These third-party dependencies and interconnections could have systemic effects.[67] Further, the lack of interpretability or 'auditability' of AI and machine learning methods has the potential to contribute to macroeconomic risk unless regulators find ways to supervise the AI.[68] This is particularly challenging because of the opacity of models generated by AI or machine learning,[69] and AI-related expertise beyond those developing the AI is limited, in the private sector and among regulators.[70]

C *Cybersecurity*

AI could be used to attack, manipulate, or otherwise harm an economy and threaten national security either directly through its financial system and/or by effecting the wider economy.[71] Algorithms could be manipulated to undermine economies to create unrest, or to send wrong signals to trading units to seek to trigger a systemic crisis.[72] The cybersecurity dimension is more serious as many financial services firms rely on a small group of technology providers, creating a new form of risk we term 'techrisk'.[73] That many AI-enabled systems have not been tested in financial crisis scenarios further amplifies this risk.[74]

Important ways to address cybersecurity include:

(a) investing in cybersecurity resources, including in-house expertise and training of employees;

---

64 Financial Stability Board (n 26). The Board is 'an international body that monitors and makes recommendations about the global financial system': 'About the FSB', *Financial Stability Board (FSB)* (Web Page) <https://www.fsb.org/about/>.

65 Ibid 33–4.

66 European Banking Federation ('EBF'), *EBF Position Paper on AI in the Banking Industry* (Report EBF_037419, 1 July 2019) 26.

67 Lin (n 15) 544.

68 Financial Stability Board (n 26) 33.

69 Ibid 33–4.

70 Ibid.

71 Lin (n 15) 538–9.

72 Ross P Buckley, Douglas W Arner, Dirk A Zetzsche and Eriks Selga, 'TechRisk' [2020] (1) *Singapore Journal of Legal Studies* 35, 43-4.

73 Douglas W Arner, Ross P Buckley, and Dirk A Zetzsche, 'Fintech, Regtech and Systemic Risk: The Rise of Global Technology Risk' in Douglas W Arner, Emilios Avgouleas, Danny Busch and Steven L Schwarcz (eds), *Systemic Risk in the Financial Sector: Ten Years after the Great Crash* (McGill-Queen's University Press, 2019) 69.

74 Buchanan (n 1) 29.

(b) having protocols in place to cooperate swiftly with other financial intermediaries, to ensure fast detection of, and responses to, these attacks, with or without involvement of regulators; and

(c) building national and international systems for sharing information as well as contingency and defence planning.[75]

D *Legal Risks*

One acronym often used to describe AI and machine learning governance considerations is 'FAT', meaning 'fairness, accountability and transparency'.[76]

In relation to accountability for the use of AI, many scholars and practitioners start with an analysis of how existing liability regimes, such as product liability, tort and vicarious liability, may be used to address the legal risks and liability.[77] The foundational concepts of those regimes, like causation and damages, are not easily applied to AI and its corporate and individual creators.[78]

Since the 2008 Global Financial Crisis, the legal and regulatory compliance of many financial institutions around the world has been found wanting. Boston Consulting Group reported that as of 2019, financial institutions, including 50 of the largest European and North American banks, had paid US$381 billion in cumulative financial penalties since the GFC.[79]

Regtech is increasingly seen as a way to address legal and regulatory requirements, and many solution providers are using AI and machine learning in areas such as on-boarding, anti-money laundering and fraud detection. Yet, some of

---

75 Buckley et al (n 72) 61–2.

76 FAT-ML conferences have been held by Princeton University since 2014: FAT / ML, *Fairness, Accountability, and Transparency in Machine Learning* (Web Page) <https://www.fatml.org/>. However, some other acronyms have emerged, such as the FEAT principles (adding 'ethics') of the Monetary Authority of Singapore ('MAS'): see Monetary Authority of Singapore, 'MAS Introduces New FEAT Principles to Promote Responsible Use of AI and Data Analytics' (Media Release, 12 November 2018): <https://www.mas.gov.sg/news/media-releases/2018/mas-introduces-new-feat-principles-to-promote-responsible-use-of-ai-and-data-analytics>.

77 See Greg Swanson, 'Non-Autonomous Artificial Intelligence Programs and Products Liability: How New AI Products Challenge Existing Liability Models and Pose New Financial Burdens' (2019) 42(3) *Seattle University Law Review* 1201; Iria Giuffrida, 'Liability for AI Decision-Making: Some Legal and Ethical Considerations' (2019) 88(2) *Fordham Law Review* 439; Andrea Bertolini, *Artificial Intelligence and Civil Liability* (European Parliament, Policy Department for Citizens' Rights and Constitutional Affairs, Directorate-General for Internal Policies, PE 621.926, July 2020) <https://www.europarl.europa.eu/RegData/etudes/STUD/2020/621926/IPOL_STU(2020)621926_EN.pdf>; Brian W Tang, 'Forging a Responsibility and Liability Framework in the AI Era for RegTech' in Jànos Barberis, Douglas W Arner and Ross P Buckley (eds), *The REGTECH Book: The Financial Technology Handbook for Investors, Entrepreneurs and Visionaries in Regulation* (Wiley, 2019) 235.

78 See, eg, European Commission Report from the Expert Group on Liability and New Technologies – New Technologies Formation, *Liability for Artificial Intelligence and Other Emerging Digital Technologies* (2019) <https://ec.europa.eu/transparency/regexpert/index.cfm?do=groupDetail.groupMeetingDoc&docid=36608>.

79 Boston Consulting Group, *Global Risk 2020: It's Time for Banks to Self-Disrupt* (April 2020) 20 <https://image-src.bcg.com/Images/BCG-Global-Risk-2020-It%E2%80%99s-Time-for-Banks-to-Self-Disrupt-Apr-2020_tcm9-243862.pdf>.

the largest recent bank fines have arisen from legal risks related to the use of technology. For example, in 2020 the Federal Court of Australia confirmed the largest civil penalty in Australian history; namely, Westpac's fine of A$1.3 billion by the Australian Transaction Reports and Analysis Centre. The penalty was largely attributed to systemic risk management failures including those related to several programming 'glitches' leading to 23 million breaches of financial crime laws over five years for not reporting suspicious bank transfers (such as customers paying for child exploitation abroad), and failure to retain back-up files.[80] The result of such large-scale regtech failures has been increasing attention from regulators in terms of the adequacy of technology systems in financial institutions, with a leading example being a 2020 action against Citigroup that resulted in a US$400 million fine by the United States ('US') Office of the Comptroller of the Currency and a direction to Citigroup to improve its internal technological systems.[81]

As more financial institutions, fintechs and crypto-asset service providers incorporate AI into their systems, including their regtech infrastructure, the legal risks for such regulated entities may well increase.

E *Ethics and Financial Services*

Ethics in finance are crucial. Ethical issues became prominent following the 2008 Global Financial Crisis and have received continued attention due to subsequent scandals, including those relating to the London InterBank Offered Rate ('LIBOR'),[82] foreign exchange[83] and most recently the entire Australian financial system.[84] Some financial services ethical questions will likely be addressed by future regulatory or self-regulatory efforts that fall into three areas: (1) AI as non-ethical actor; (2) AI's influence on humans; and (3) artificial stupidity and maleficence.

---

80 *Chief Executive Officer of the Australian Transaction Reports and Analysis Centre v Westpac Banking Corporation* [2020] FCA 1538. See also Brian Monroe, 'After Months of Court Battles, Westpac Settles with Austrac, Agrees to pay $1.3 billion for Millions of AML Failings, Ties to Child Exploitation Network', *Association of Financial Crime Specialists* (Web Page, 6 October 2020) <https://www.acfcs.org/after-months-of-court-battles-westpac-settles-with-austrac-agrees-to-pay-1-3-billion-for-millions-of-aml-failings-ties-to-child-exploitation-network/>; Charlotte Grieve, 'The Westpac Scandal: How Did It Happen?', *The Sydney Morning Herald* (online, 9 December 2019) <https://www.smh.com.au/business/banking-and-finance/the-westpac-scandal-how-did-it-happen-20191206-p53ho2.html>; James Eyers, 'How a Software Bug Triggered Westpac's Woes', *Australian Financial Review* (online, 21 November 2019) <https://www.afr.com/companies/financial-services/how-a-software-bug-triggered-westpac-s-woes-20191121-p53csx>.

81 Jesse Hamilton and Jennifer Surnane, 'Citigroup Pays $400 Million Penalty, Must Get U.S. Approval on Deals', *Bloomberg* (online, 4 October 2020) <https://www.bloomberg.com/news/articles/2020-10-07/citigroup-pays-400-million-must-get-u-s-approval-on-deals?sref=lP8hg7Cm>.

82 See HM Treasury, *The Wheatley Review of LIBOR* (Final Report, September 2012) <https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/191762/wheatley_review_libor_finalreport_280912.pdf>.

83 Byron Kaye, 'FX Collusion Scandal Reaches Australia, Class Action Launched', *Reuters* (online, 27 May 2019) <https://www.reuters.com/article/us-australia-banks-idUSKCN1SX06V>.

84 *Royal Commission into Misconduct in the Banking, Superannuation and Financial Services Industry* (Final Report, February 2019) ('*Banking Royal Commission Final Report*').

1 *AI as Non-Ethical Actor*

Algorithms per se neither 'feel' nor have 'values'. Training machines in values is difficult, since humans often lack insights into the human psyche: that is, people often cannot tell why they act as they do.[85] While some ethical concerns, such as banning interest under Shariah law, can be codified in ways that suit algorithms, drivers for most human actions are more subtle and contextual, and subject to change with circumstances.

AI's lack of ethical foundations could seriously harm portfolio values of financial assets if, as is likely, the AI misprices reputational risk. For instance, Microsoft's AI bot, Tay, designed and deployed to engage in casual online conversation with users, learned from the interactions and exhibited severely anti-Semitic and misogynistic behaviour within 16 hours due to user mischief.[86] A broader deployment would have been even more devastating. Volkswagen's severe ethical shortcomings in using technology to evade regulatory requirements were all too human, but software controlling engine performance in test situations could foreseeably be programmed by AI in the future,[87] including in a way that optimises cost savings over regulatory compliance.

This risk is intensified by access to vast data about individual human users. The more data an AI has about a certain person, the greater the risk the AI may nudge the person into buying an unsuitable financial product or profile the person for credit determinations. The advent and rise of unsupervised learning, generative adversarial networks that generate their own data and powerful autoregressive language models, such as Generative Pre-trained Transformer 3, increase the potential impact of AI that operates with limited human intervention. While unethical conduct can be mitigated by more diverse and broadly trained technical teams programming the AI, the core issue remains that the code itself is a non-ethical actor that does not necessarily constantly review, revise and reflect on its performance as we hope humans do.[88] AI needs human monitoring and guidance for ethical decision-making: humans in the loop are essential so that a human or group of humans is responsible for the actions of AI.

2 *AI's Influence on Humans*

AI can enhance or diminish human capacity. AI as augmented intelligence could turn an unskilled person into a skilled investor, via recommendations. The same applies to human decision-making errors revealed in behavioural finance: for example, AI could be programmed to address certain human biases (such as

---

85 Enriques and Zetzsche (n 47) 78.

86 Elle Hunt, 'Tay, Microsoft's AI Chatbot, Gets a Crash Course in Racism from Twitter', *The Guardian* (online, 24 March 2016) <https://www.theguardian.com/technology/2016/mar/24/tay-microsofts-ai-chatbot-gets-a-crash-course-in-racism-from-twitter>.

87 Capgemini Research Institute, *Accelerating Automotive's AI Transformation: How Driving AI Enterprise-wide Can Turbo-charge Organizational Value* (Report, March 2019) 17–18.

88 Lin (n 15) 537–8.

confirmation bias, optimism bias and negativity bias) when making investment decisions.[89]

Conversely, AI could decrease human capacity. As the need to develop advanced maths and other sophisticated data analytical capacities decreases with appropriate programmes being widely available, humans' data analytic capacities may atrophy. This is supported by the WEF's suggestion that increasing reliance on AI in the future could lead to the erosion of 'human financial talent' as humans lose the skills required to challenge AI-enabled systems or to respond well to crises.[90] Accordingly, while coaching AI could be used to enhance financial and technological literacy of staff and investors, resulting in better resource allocation, exploitative AI could ask or nudge clients to invest in overpriced financial products that benefit only the product originator.

Research into how humans respond to computer-generated incentives is ongoing and hints at serious risks.[91] Humans respond to certain communications with an enhanced degree of trust. As AI becomes more pervasive in its interaction with users, disclosed or otherwise, it may implicitly generate an increasing level of trust. AI developers thus bear a high level of responsibility and there is a strong need for ethical restrictions through rules and internal controls of financial institutions and their management.

3 *Artificial Stupidity and Maleficence*

Protection against AI mistakes and unethical behaviour is a major concern. Errors and unethical behaviour can arise from poor or criminally motivated programming, or from inadequate datasets, or correlations with other events resulting in harmful unforeseen consequences. Another important example could arise if certain conduct results in liability for which consumers sue far more than institutional clients, as an algorithm could decide to avoid consumer relationships, thereby financially excluding them.

F *Risk Typology: Framework of Analysis*

The risks of AI in finance fall into three major categories: (1) information asymmetry; (2) data dependency; and (3) interdependency.[92]

First, AI enhances information asymmetry about the functions and limits of certain algorithms. Third-party vendors often understand the algorithms far better

---

89 Chau Duong, Gioia Pescetto and Daniel Santamaria, 'How Value-Glamour Investors Use Financial Information: UK Evidence of Investors' Confirmation Bias' (2014) 20(6) *The European Journal of Finance* 524.

90 World Economic Forum (n 60) 69–71.

91 See Andrea Ferrario, Michele Loi and Eleonora Viganò, 'In AI We Trust Incrementally: A Multilayer Model of Trust to Analyze Human-Artificial Intelligence Interactions' (2020) 33(3) *Philosophy & Technology* 523; Omri Gillath, Ting Ai, Michael S Branicky, Shawn Keshmiri, Robert B Davison and Ryan Spaulding, 'Attachment and Trust in Artificial Intelligence' (2021) 115 *Computers in Human Behavior* 106607 <https://doi.org/10.1016/j.chb.2020.106607>.

92 Enriques and Zetzsche (n 47) 75–90.

than the financial institutions that buy and use them and their supervisors. However, for proprietary and competitive reasons, technology vendors traditionally fail to fully explain how their creations work. Increased transparency through explainability and interpretability needs to be demanded by users, financial institutions and regulators alike.

Second, AI enhances data dependency as data sources are critical for its operation. The effects and potentially discriminatory impact of AI may change with a different data pool.

Third, AI enhances interdependency. AI can interact with other AI with unexpected consequences, enhancing or diminishing its operations in finance.[93]

The law will need to address the risks of AI by preventive regulation and corrective liability allocation. Given the rapid developments in AI, drafting and enforcing these rules is a serious challenge. Rather than the much-discussed private law dimension and liability allocation,[94] we focus on regulatory tools in Parts IV–V below.

IV Regulating AI in Finance: Challenges for External Governance

As we have pointed out above (Part I), the use of AI in finance has become a focus of regulatory attention. We summarise general frameworks proposed by regulators (including data protection and privacy), before turning to financial regulators' approaches to AI. We then argue that traditional regulatory approaches to financial supervision, such as external governance frameworks, are not likely to be effective in this context and, instead, external governance must require internal governance, in particular personal responsibility.

A *General AI Frameworks*

General frameworks addressing degrees of human responsibility in developing and dealing with AI, are evolving worldwide.

1 *AI Principles*

The first development in what was to be a remarkable flurry of activity in this space was the UK House of Lords Select Committee on Artificial Intelligence defining five general principles of AI development and treatment in late 2017.[95]

The most influential of all the subsequent initiatives occurred in May 2019, with the adoption of the Organisation for Economic Co-operation and Development ('OECD') AI Recommendation and its five principles:

---

93 Lin (n 15) 542.

94 Mark A Lemley and Bryan Casey, 'Remedies for Robots' (2019) 86(5) *University of Chicago Law Review* 1311, 1313. See also above n 77.

95 Select Committee on Artificial Intelligence, *AI in the UK: Ready, Willing and Able?* (House of Lords Paper No 100, Session 2017–19) 125 [417].

(i) AI must be beneficial for people and the planet;
(ii) AI system design must comply with general legal principles such as the rule of law;
(iii) AI systems must be transparent and explainable;
(iv) AI systems should be robust, secure and safe; and
(v) all AI actors (including system developers) must be accountable for compliance with these principles.[96]

Drawing on the OECD AI Recommendation, the G20 endorsed the G20 AI Principles in June 2019.[97] In September 2019, endorsing the OECD AI Recommendation, the US Chamber of Commerce released 'Principles on Artificial Intelligence', calling for US businesses to abide by these standards.[98]

In November 2019, the Australian Government Department for Industry, Innovation and Science announced the *AI Ethics Framework*, based on eight key principles: human, social and environmental wellbeing; human-centred values; fairness; privacy protection; reliability and safety; transparency and explainability; contestability; and accountability.[99]

In China, the Beijing Academy of Artificial Intelligence released its AI Principles in May 2019[100] and the Ministry of Science and Technology National New Generation Artificial Intelligence Governance Expert Committee published its Governance Principles for a New Generation of AI in June 2019.[101]

Numerous parallel AI ethics initiatives were also generated by the private sector and by many researchers.[102]

2 *Data Protection and Privacy*

Data protection and privacy commissioners have increasingly viewed the governance of AI as within their purview. For instance, the 40th International Conference of Data Protection and Privacy Commissioners in 2018 endorsed six guiding principles as core values to preserve human rights in the development of AI:

---

96 Organisation for Economic Co-operation and Development ('OECD'), *Recommendation of the Council on Artificial Intelligence* (OECD/LEGAL/0449, 22 May 2019) paras 1.1–1.5 <https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449> ('OECD AI Recommendation').

97 G20, 'G20 AI Principles' in annex to *G20 Ministerial Statement on Trade and Digital Economy* (8–9 June 2019) <https://www.g20-insights.org/wp-content/uploads/2019/07/G20-Japan-AI-Principles.pdf>.

98 US Chamber of Commerce, 'US Chamber of Commerce Principles on Artificial Intelligence' (Press Release, 23 September 2019).

99 Department of Industry, Science, Energy and Resources (Cth), 'AI Ethics Principles' (Web Page) <https://www.industry.gov.au/data-and-publications/building-australias-artificial-intelligence-capability/ai-ethics-framework/ai-ethics-principles>.

100 Beijing Academy of Artificial Intelligence, 'Beijing AI Principles' (Web Page, 28 May 2019) <https://www.baai.ac.cn/news/beijing-ai-principles-en.html>.

101 Ministry of Science and Technology National New Generation Artificial Intelligence Governance Expert Committee, *Governance Principles for a New Generation of Artificial Intelligence: Develop Responsible Artificial Intelligence* (Report, 17 June 2019).

102 See, eg IEEE, *Ethically Aligned Design: A Vision for Prioritizing Human Well-being with Autonomous and Intelligent Systems, Version II* (Report, 2018); Future of Life Institute, 'Asilomar AI Principles' (Web Page, 2017) <https://futureoflife.org/ai-principles/?cn-reloaded=1>.

(1) Fairness;
(2) Continued attention, vigilance, and accountability;
(3) AI system transparency and intelligibility;
(4) AI system responsible development and design by applying the principles of privacy by default and privacy by design;
(5) Empowerment of individuals; and
(6) Reduction and mitigation of unlawful biases/discrimination arising from AI data use.[103]

The Conference called for AI common governance principles and a permanent working group on Ethics and Data Protection in AI.[104]

Article 22 of the *European General Data Protection Regulation* ('*GDPR*')[105] also requires ethical AI performance.[106] Entitled 'Automated individual decision-making, including profiling', art 22(1) states that a data subject has 'the right to not be subject to a decision based solely on automated processing, including profiling, which produces legal effects'. Caveats apply under art 22(2) if the decision 'is necessary for the entering into, or performance of, a contract between the data subject and the data controller', or in other specific cases. It has been argued that the data subject has the right to insist on human intervention in purely AI-driven decisions, and to contest the decision,[107] although any 'right to explanation' under the *GDPR* has been found wanting.[108] At the same time, the UK Information Commissioner's Office has issued guidance on AI and data protection, and has conducted a public consultation on an auditing framework for AI.[109]

B *Financial Regulation and AI*

Globally, regulators have started considering how AI impacts financial services and to issue regulatory guidance.

In 2016, the European Supervisory Authorities ('ESAs') (European Banking Authority, European Securities and Markets Authority and European Insurance and

---

[103] 40th International Conference of Data Protection and Privacy Commissioners, *Declaration on Ethics and Data Protection in Artificial Intelligence* (23 October 2018) 3–6 <http://globalprivacyassembly.org/wp-content/uploads/2019/04/20180922_ICDPPC-40th_AI-Declaration_ADOPTED.pdf>.

[104] Ibid 5–6.

[105] *Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation)* [2016] OJ L 119/1 ('*GDPR*').

[106] Jimmie Franklin, 'GDPR Has Kept AI Ethical, Despite Concerns', *International Financial Law Review* (online, 2 October 2019) <https://www.iflr.com/Article/3896942/GDPR-has-kept-AI-ethical-despite-concerns.html>.

[107] Paul Voigt and Axel von dem Bussche, *The EU General Data Protection Regulation (GDPR): A Practical Guide* (Springer, 2017) 180–2.

[108] Sandra Wachter, Brent Mittelstadt and Luciano Floridi, 'Why a Right to Explanation of Automated Decision-Making Does Not Exist in the *General Data Protection Regulation*' (2017) 7(2) *International Data Privacy Law* 76.

[109] Information Commissioner's Office (UK), *Guidance on AI and Data Protection* (30 July 2020); Information Commissioner's Office (UK), *Guidance on the AI Auditing Framework: Draft Guidance for Consultation* (14 February 2020).

Occupational Pensions Authority) published a discussion paper on Big Data risks for the financial sector, which included discussion on AI.[110]

The 2018 *ESAs Joint Committee Final Report on Big Data* found that Big Data risks are best addressed by existing legislation on data protection, cybersecurity and consumer protection, even though such legislation may not have been written specifically to address Big Data risks.[111] This legislation includes: the *GDPR*;[112] the second *Payment Services Directive* ('*PSD2*');[113] the second *Markets in Financial Instruments Directive* ('*MiFID II*');[114] and the *Insurance Distribution Directive* ('*IDD*').[115]

The ESAs' organisational and prudential requirements involve sound internal control mechanisms, market activity monitoring, record-keeping, and management of conflicts of interest.[116] The requirements emphasise business principles such as: acting honestly, fairly and professionally; refraining from misleading conduct; ensuring products suit the needs of clients; and establishing fair claims/complaints handling processes.[117] Further, to ensure fair and transparent consumer treatment, the ESAs encourage Big Data good practices, such as regularly monitored robust processes, transparent consumer compensation mechanisms, and compliance with the *GDPR*.[118]

Other financial regulators are likewise increasingly engaging with AI. These include (in chronological order):

---

110 Joint Committee of the European Supervisory Authorities, *Discussion Paper on the Use of Big Data by Financial Institutions* (Discussion Paper No JC/2016/86, 19 December 2016).

111 *ESAs Joint Committee Final Report on Big Data* (n 2) 23.

112 *GDPR* (n 105).

113 *Directive (EU) 2015/2366 of the European Parliament and of the Council of 25 November 2015 on Payment Services in the Internal Market* [2015] OJ L 337/35 ('*PSD2*').

114 *Directive 2014/65/EU of the European Parliament and of the Council of 15 May 2014 on Markets in Financial Instruments* [2020] OJ L 173/349 ('*MiFID II*').

115 *Directive (EU) 2016/97 of the European Parliament and of the Council of 20 January 2016 on Insurance Distribution* [2020] OJ L 26/19 ('*IDD*').

116 *MiFID II* (n 114) arts 17, 23; *Directive 2013/36/EU of the European Parliament and of the Council of 26 June 2013 on Access to the Activity of Credit Institutions and the Prudential Supervision of Credit Institutions and Investment Firms* [2018] OJ L 176/338, art 79 ('*CRD IV*'); *IDD* (n 115) arts 17, 27–8; *Directive 2014/17/EU of the European Parliament and of the Council of 4 February 2014 on Credit Agreements for Consumers relating to Residential Immovable Property* [2018] OJ L 60/34 art 7 ('*MCD*').

117 *MiFID II* (n 114) arts 16, 24; *IDD* (n 115) arts 14, 17(1), 25; *MCD* (n 116) art 7(1); *AIFMD* art 12; *Directive 2009/65/EC of the European Parliament and of the Council of 13 July 2009 on the Coordination of Laws, Regulations and Administrative Provisions relating to Undertakings for Collective Investment in Transferable Securities [2020] OJ L 302/32*, arts 13–14; *PSD2* arts 19(6), 101; *Delegated Regulation (EU) 2017/565 of 25 April 2016 Supplementing Directive 2014/65/EU of the European Parliament and of the Council as regards Organisational Requirements and Operating Conditions for Investment Firms and Defined Terms for the Purposes of that Directive* [2019] OJ L 87/1, art 26.

118 *ESAs Joint Committee Final Report on Big Data* (n 2) 24.

- The Monetary Authority of Singapore introduced the new Fairness, Ethics, Accountability and Transparency ('FEAT') Principles to promote responsible use of AI and data analytics in November 2018.[119]
- De Nederlandsche Bank issued a discussion paper on principles for responsible use of AI, namely soundness, accountability, fairness, ethics, skills and transparency (or 'SAFEST') in July 2019.[120]
- The Bank of England and the Financial Conduct Authority ('FCA') published a survey entitled *Machine Learning in UK Financial Services* in October 2019.[121]
- The Hong Kong Monetary Authority issued its twelve 'High-level Principles on Artificial Intelligence' in November 2019.[122]

Singapore's FEAT Principles were updated in February 2019 and again in January 2020 to reflect Singapore's Personal Data Protection Commission's Proposed AI Governance Framework, which has two guiding principles: (i) that organisations must ensure that decision-making using AI is explainable, transparent and fair, and (ii) that AI solutions should be human-centric.[123] This Framework provides guidance in the following areas:

(1) Internal governance structures and measures;
(2) Appropriate AI decision-making models, including determining acceptable risk appetite and circumstances for human-in-the-loop, human-over-the-loop and human-out-of-the-loop approaches;
(3) Operations management, including good data accountability practices and minimising inherent bias; and
(4) Customer relationship management, including disclosure, transparency, and explainability.[124]

In November 2019, the Monetary Authority of Singapore announced the creation of the Veritas framework to promote the responsible adoption of AI and data analytics by financial institutions using open source tools as a verifiable way for financial institutions to incorporate the FEAT Principles.[125] The first phase of Veritas initiative, involving an expanded consortium membership of 25, focused on the development of fairness metrics in customer marketing and credit risk scoring.[126]

---

119 Monetary Authority of Singapore, *Principles to Promote Fairness, Ethics, Accountability and Transparency (FEAT) in the Use of Artificial Intelligence and Data Analytics in Singapore's Financial Sector* (Report, November 2018).

120 De Nederlandsche Bank, *General Principles for Use of Artificial Intelligence in Finance* (Discussion Paper, 25 July 2019).

121 Bank of England and Financial Conduct Authority (n 3).

122 Hong Kong Monetary Authority, 'High-Level Principles on Artificial Intelligence' (Media Release, 1 November 2019).

123 Singapore Personal Data Protection Commission, *A Proposed Artificial Intelligence Governance Model* (Report, January 2019); Singapore Info-Communications Media Development Authority and Personal Data Protection Commission, *Model Artificial Intelligence Governance Framework: Second Edition* (January 2020).

124 Singapore Info-Communications Media Development Authority and Personal Data Protection Commission (n 123) 20–63.

125 Monetary Authority of Singapore, 'MAS Partners Financial Industry to Create Framework for Responsible Use of AI' (Media Release, 13 November 2019).

126 Monetary Authority of Singapore, '"Fairness Metrics" to Aid Responsible AI Adoption in Financial Services' (Media Release, 28 May 2020).

The initiative is now in the second phase, which will develop the ethics, accountability and transparency assessment methodology for the two phase one use cases, plus insurance industry use cases.[127]

Similarly, the Hong Kong Monetary Authority encouraged authorised institutions in May 2019[128] to adopt and implement the Hong Kong Privacy Commissioner Ethical Accountability Framework,[129] and 2018 Data Stewardship Accountability, Data Impact Assessments and Oversight Models.[130] This was followed by the Hong Kong Monetary Authority's November 2019 High-Level Principles on AI.[131] Specifically, the Principles reinforce that banks should:

- possess sufficient expertise;
- ensure explainability of AI applications;
- use data of good quality;
- conduct rigorous model validation;
- ensure auditability of AI applications;
- implement effective management oversight of third-party vendors;
- be ethical, fair and transparent;
- conduct periodic reviews and ongoing monitoring;
- comply with data protection requirements;
- implement effective cybersecurity measures; and
- implement risk mitigation and contingency plans.

The Hong Kong Monetary Authority's Banking Conduct Department also issued Guiding Principles on Consumer Protection in respect of Use of Big Data Analytics and AI by Authorised Institutions.[132] Reinforcing a risk-based approach to Big Data analytics and AI, the principles focus on governance/accountability, fairness, transparency/disclosure, and data privacy and protection.

C *The Inadequacy of External Governance*

Financial supervisory authorities find it increasingly difficult to tackle AI-related risks through traditional means of financial supervision, that is: external governance.

We draw on five examples to support our thesis on the inadequacy of external governance regimes in addressing the risks of AI in finance: (1) the authorisation of AI; (2) the outsourcing of rules and e-personhood; (3) the role of AI with regard to key functions; (4) the qualifications of core personnel; and (5) sanctioning rules.

---

127 Monetary Authority of Singapore, 'Veritas Initiative Addresses Implementation Challenges in the Responsible Use of Artificial Intelligence and Data Analytics' (Media Release, 6 January 2021).

128 Hong Kong Monetary Authority, 'Use of Personal Data in Fintech Development' (Media Release, 3 May 2019).

129 Information Accountability Foundation and Hong Kong Privacy Commissioner for Personal Data, *Ethical Accountability Framework for Hong Kong China* (Report, October 2018).

130 Information Accountability Foundation and Hong Kong Privacy Commissioner for Personal Data, *Data Stewardship Accountability, Data Impact Assessments and Oversight Models: Detailed Support for an Ethical Accountability Framework* (Report, October 2018).

131 Hong Kong Monetary Authority (n 122).

132 Hong Kong Monetary Authority, 'Consumer Protection in Respect of Use of Big Data Analytics and Artificial Intelligence by Authorized Institutions' (Media Release, 5 November 2019).

1 *The Authorisation of AI*

Enhanced AI use influences the conditions for authorisation. If a business model seeking authorisation relies on AI, the business and operations plan must detail the functioning of the AI, the client protection features, the regulatory capital assigned to financial and operational risks for the AI-performed services, and the back-up structure in case the AI fails. Regulatory frameworks across the globe already require IT contingency plans and multiple data storage and cybersecurity strategies. These regulatory approaches are unlikely to change fundamentally, but will become even more important in practice.

One potential response to AI-based threats is a licensing requirement for AI used by financial institutions.[133] A mandatory AI insurance scheme is another.

Financial services authorities worldwide are increasingly seeking to upskill and introduce technology to perform meaningful reviews of AI.[134] To our knowledge, software to monitor a self-learning AI's conduct does not yet exist. Moreover, outcome-based testing depends on the data pools available for testing; if the test pools differ from the real-use case data pools, the results of testing may be of little value.

AI authorisation may also have several undesirable side-effects. The most important is that it may limit innovation given authorisation is costly and slow. Rules would also struggle to cope with the (often, almost daily) minor amendments and improvements to AI programmes. Re-authorisation of the code is expensive, meaning minor improvements, or a series of minor improvements together representing a major step, of existing AI may be uneconomic. Finally, for unsupervised self-learning AI, the authorised code will not be performing in practice, as by definition such self-learning AI develops while performing its services. Thus, authorisations will always be outdated.[135] Regulatory sandboxes provide a risk-controlled environment where regulatory restrictions are relaxed to foster innovation. While in some settings, sandboxes may support innovation and effective regulation,[136] assessment of performance under sandbox conditions remains a relatively poor substitute for performance under real world conditions.[137]

2 *Regulatory Outsourcing of Rules and e-Personhood*

In regulatory rulebooks worldwide, crucial supplier frameworks apply for AI owned and operated by, or outsourced to, a separate services provider. The crucial supplier is subject to additional monitoring by the outsourcing financial institution. However, financial services AI may increasingly be owned and operated in-house by the

---

133 Andrew Tutt, 'An FDA for Algorithms' (2017) 69(1) *Administrative Law Review* 83, 111.

134 See Dirk Broeders and Jermy Prenio, Financial Stability Institute ('*FSI*'), *Innovative Technology in Financial Supervision (Suptech) — The Experience of Early Users* (FSI Insights on Policy Implementation No 9, July 2018); Toronto Centre, *FinTech, RegTech and SupTech: What They Mean for Financial Supervision* (Report, August 2017).

135 Enriques and Zetzsche (n 47).

136 World Economic Forum (n 60).

137 Enriques and Zetzsche (n 47).

financial intermediary's own staff. This raises questions around the adequacy of the AI legal framework.

One option for regulating in-house AI is the granting of limited legal personality to the algorithm itself, similar to a partial licence, paired with a self-executing 'kill switch' linked to minimum requirements as to the capital available for potential liability claims. If the capital is depleted, for example due to liabilities or regulatory sanctions, the algorithm will stop operating. The argument against such limited e-personhood are similar to those against authorising AI: the calculation of capital requires a clear delineation of risks created by the AI. If the limits of AI functions are vague, as with self-learning algorithms, regulatory capital will most likely be set too low or too high. Further, authorities have cheaper ways to restrict AI use, without a financial institution AI's own regulatory capital. These include reporting requirements for AI deployment as well as losses and damages resulting from such deployment, and responding to such reporting by issuing orders limiting, or prohibiting such AI applications as deemed appropriate.

3 *AI as a Key Function Holder?*

Can an AI serve as an executive or board member of the financial institution?[138] Here, legality and practicality differ.

In some jurisdictions, executive functions can be assigned to legal entities, or the law is silent on the entity status of executives. In those jurisdictions, it may be lawful to appoint an AI as a board member, if necessary, by embedding the AI in a special purpose vehicle (that is, a parent company subsidiary with a very limited business objective) as its sole activity. In other jurisdictions, these functions must be fulfilled by people.

Regarding practicality, an AI may function as a board member for certain routine tasks (for example, securitisation vehicles in a corporate group), and for procedural monitoring, but a human board majority may be required to ensure continuing operations when challenges exceed the AI's programmed limits.

Notwithstanding this, rules allowing AI to assume functions within a financial institution must respect the existing limits of AI, especially for compliance monitoring. AI alone is poorly adapted to handle compliance matters because it lacks ethical screening abilities, and because rules are incomplete on purpose. The law is full of vague terms such as 'fair', 'adequate', 'just', and 'reasonable'. These terms allow adjustment to an ever-changing world. Financial services are heavily regulated by rules that do not always operate in yes/no terms because their meaning depends on context. For this reason, an 'AI as compliance officer' could well lead to inaccurate monitoring, widespread misreporting, and mispricing of risks.[139]

---

[138] Deep Knowledge Ventures, 'Deep Knowledge Venture's [sic] Appoints Intelligent Investment Analysis Software VITAL as Board Member — Hong Kong Venture Capital Fund Appoints Machine Intelligence as Board Member', *GlobeNewswire* (online, 13 May 2014) <https://www.globenewswire.com/news-release/2014/05/13/635881/10081467/en/Deep-Knowledge-Venture-s-Appoints-Intelligent-Investment-Analysis-Software-VITAL-as-Board-Member.html>.

[139] Enriques and Zetzsche (n 47) 74–5.

4 *The Fit-and-Proper Test for Core Personnel*

AI will likely influence regulatory practice in the fit-and-proper test for key function holders (that is, senior management or executives) and the board of directors, in two ways. First, some existing requirements may be redundant or need modification when AI is used. For instance, if AI is making decisions, a human executive's credentials may not require review.

Second, new requirements will reflect the greater reliance on AI, and some office holders may have new qualifications. EU authorities require executives of a financial intermediary to have at least three years of executive experience prior to appointment. This experience should demonstrate good standing, diligent handling of client matters and cooperation with the financial supervisory authority. However, AI experts may have accumulated their AI experience outside the financial sector, for example within a major e-commerce or software firm. Financial supervisors will need to modify some of their experience requirements as many have for licensing requirements for fintechs.

5 *Sanctioning AI*

Financial regulation typically imposes sanctions on an institution for its overall conduct and/or that of individual staff. To do so, regulators usually must prove negligence or ill intent of the institution and/or staff. When harm occurs, deficiencies in risk management systems may attract sanctions. With AI, these cases will be increasingly hard to make. Where AI fails and supervisors are incapable of establishing an AI's processes and limits with certainty, determining the culpability standard and burden of proof to be applied while retaining incentives to innovate will be very challenging. Potential sanctions may exercise little steering effect, even if sanctions are possible under the broad 'failure of risk management' rationale.[140]

This brings us to the question of sanctioning AI. Withholding compensation, naming and shaming, and financial penalties have little meaning for AI. Similarly, director disqualification — the equivalent of a death penalty in the world of corporate management — as well as civil and criminal liability, have a limited steering effect for AI in its current form, other than perhaps being imposed upon currently unregulated outsourced technology companies or their regulated client financial institutions.

Hence, any sanctioning system needs reconsidered incentives for AI creation and deployment. AI-adapted regulation could possibly:

(i) require blame-free remediation in which organisations are able to learn from failures and make improvements;

(ii) encourage collaboration to promote early detection and the avoidance of unexpected AI failures; and/or

[140] Bamberger (n 22) 676.

(iii) employ fit-for-purpose explainability with frameworks to decide 'if' explainability is a requirement on a risk- and impact-based assessment in any particular circumstance (thereby assisting organisations to prioritise their AI's objectives) and 'how' explainability should be achieved.[141]

V Putting the Human in the Loop in Finance

While regulators expect financial institutions to deploy AI responsibly and develop and use new tools to safeguard the financial system, we have shown that, given the severe information asymmetry, data dependency and interdependency that arise with AI, external governance is not well-suited to ensuring the responsible use of AI in finance.

Given these black box challenges in AI for regulatory and supervisory authorities, measures focusing on personal responsibility requirements that put the human in the loop, should be central to regulating AI-enabled systems in finance.

Two approaches are gaining increasing currency. The first involves using technology (including AI) to monitor staff behaviour and identify issues before they arise (a form of regtech). As we have argued elsewhere, regtech is a logical consequence of fintech; fintech cannot work well without properly designed and implemented regtech.[142]

The second approach is central to putting the human in the loop and will thus be expanded further here. An increasing range of regulatory systems strengthen the personal responsibility of designated senior managers — so-called 'senior manager', 'manager-in-charge', 'key function holders' or 'personal responsibility' systems. These frameworks seek to produce cultural change and an ethical environment in financial institutions through the personal responsibility of directors, management and, increasingly, individual managers.

We argue in this section that regulators should utilise and strengthen these external governance requirements in order to require human-in-the-loop systems for internal AI governance. External governance of AI risks and challenges should primarily be by mandating the quality and intensity of financial institutions' internal governance. AI-adjusted personal responsibility frameworks are vital. In this section, to provide context we first lay out the fundamentals of personal responsibility frameworks in financial regulation. Then we analyse how these frameworks can be utilised for addressing AI-related black box issues.

Such personal responsibility frameworks should be supplemented to include explicit AI responsibility, including a non-waivable AI due diligence and explainability standard.

---

[141] Accenture (n 7) 18; UK Finance and Microsoft (n 5) 10–13; World Economic Forum (n 60) 21.

[142] Douglas W Arner, Jànos Barberis and Ross P Buckley, 'FinTech, RegTech and the Reconceptualisation of Financial Regulation' (2017) 37(3) *Northwestern Journal of Law and Business* 371.

A *Personal Responsibility Frameworks in Finance*

Over the last ten years, most major financial jurisdictions have imposed, or are in the process of imposing, director and manager responsibility frameworks for financial regulation. Australia, along with the UK and Hong Kong, have implemented manager responsibility regimes. Singapore has proposed a regime with adoption currently delayed due to the COVID-19 epidemic.[143] The EU has developed a framework for internal governance and, to address information, communications and telecommunications risks in general, is going to adopt a Digital Operational Resilience Act ('DORA Proposal').[144] A similar manager responsibility approach has been proposed by the US Federal Reserve for 'Systemically Important Financial Institutions', but not yet adopted.[145]

1 *Australia: Banking Executive Accountability Regime*

The Australian Prudential Regulation Authority ('APRA') administers the Banking Executive Accountability Regime ('BEAR'),[146] which came into effect on 1 July 2018 for large banks and 1 July 2019 for smaller banks (collectively, 'authorised deposit-taking institutions').[147] Authorised deposit-taking institutions must provide individual accountability statements to APRA that clearly outline individual responsibilities and provide an accountability map showing accountability allocation across an institution (based on size, risk profile, and complexity). 'Individual accountable persons' are accountable for their actual or effective responsibilities for the management or control of a significant or substantial part, or aspect of, an authorised deposit-taking institution's operations or an authorised deposit-taking institution group. Specifically, individual accountable persons have obligations to: act 'with honesty and integrity, and with due skill, care, and diligence'; 'deal with APRA in an open, constructive and cooperative way'; and take reasonable steps in conducting their responsibilities to prevent matters arising that would adversely affect the authorised deposit-taking institution's prudential standing or reputation.[148]

---

143 In April 2020, the Monetary Authority of Singapore announced that 'it will adjust selected regulatory requirements and supervisory programmes to enable financial institutions (FIs) to focus on dealing with issues related to the COVID-19 pandemic and supporting their customers during this difficult period': Monetary Authority of Singapore, 'MAS Takes Regulatory and Supervisory Measures to Help FIs Focus on Supporting Customers' (Media Release, 7 April 2020) [1]. The Authority noted that this includes deferring the Guidelines on Individual Accountability and Conduct and that 'FIs will be provided sufficient time for transition to the new dates when announced': at [15].

144 See *Proposal for a Regulation of the European Parliament and of the Council on Digital Operational Resilience for the Financial Sector and Amending Regulations (EC) No 1060/2009, (EU) No 648/2012, (EU) No 600/2014 and (EU) No 909/2014*, COM/2020/595 final, 24 September 2020 ('DORA Proposal').

145 See Federal Reserve System, 'Proposed Supervisory Guidance', 83(8) *Federal Register* 1351 (11 January 2018, Docket No OP-1594). The US Federal Reserve's proposal cover large banks, bank-like institutions, and non-bank Systemically Important Financial Institutions.

146 Australian Prudential Regulation Authority ('APRA'), *Information Paper: Implementation of the Banking Executive Accountability Regime (BEAR)* (11 December 2020) 4.

147 *Banking Act 1959* (Cth) pt IIAA.

148 Ibid s 37C.

In response to recommendations of the Royal Commission into Misconduct in the Banking, Superannuation, and Financial Services Industry,[149] the Australian Government has proposed expanding BEAR into the Financial Accountability Regime ('FAR'),[150] with legislation proposed to be introduced in late 2020 to cover securities firms after public consultation. The underlying structure of FAR resembles BEAR, with several key differences:

- the Australian Securities and Investments Commission would join APRA in co-regulating FAR obligations;
- FAR expands regulatory scope to all APRA-regulated entities, not just authorised deposit-taking institutions; and
- regulators would have the power to define 'accountable person' (which, under BEAR, is defined in legislation) and to exempt entities from FAR obligations (which power, under BEAR, is with the Minister).[151]

Additionally, FAR imposes new obligations on accountable persons and introduces civil penalties for those in breach,[152] thereby strengthening the focus on individual accountability.

2 *United Kingdom: Senior Managers and Certification Regime*

The UK's 'Senior Managers and Certification Regime' ('SMCR') evolved from the EU framework[153] and has been influential internationally. Regime compliance is subject to firms and individuals being authorised by the UK Prudential Regulation Authority ('PRA') and the FCA. Authorised firms must ensure that individuals who perform PRA-designated senior management functions are approved.[154] Authorisation will not be granted unless the PRA and FCA are satisfied that the person meets the requirements of the *Financial Services and Markets Act 2000* (UK).[155]

The SMCR as established in 2016 applied to all individuals performing a 'Senior Management Function' at banks, building societies, credit unions, and PRA-designated investment firms. The Regime was expanded in 2018 to cover insurance firms, and again from December 2019, for FCA-regulated financial institutions, to apply to asset managers and designated activities of investment firms.[156]

---

149 *Banking Royal Commission Final Report* (n 84).

150 Treasury (Cth), *Implementing Royal Commission Recommendations 3.9, 4.12, 6.6, 6.7 and 6.8: Financial Accountability Regime* (Proposal Paper, 22 January 2020) ('*Proposal Paper*'); Treasurer (Cth), 'Update on the Implementation of the Banking, Superannuation and Financial Services Royal Commission' (Media Release, 8 May 2020) <https://ministers.treasury.gov.au/ministers/josh-frydenberg-2018/media-releases/update-implementation-banking-superannuation-and#:~:text=The%20Morrison%20Government%20has%20today,significant%20impacts%20of%20the%20coronavirus>.

151 Treasury (Cth), *Proposal Paper* (n 150) 14.

152 Ibid 6, 9, 11, 13.

153 See below Part V(A)(5).

154 *Financial Services and Markets Act 2000* (UK) s 59.

155 'Senior Managers Regime: Approvals', *Bank of England* (Web Page, 26 January 2021) <https://www.bankofengland.co.uk/prudential-regulation/authorisations/senior-managers-regime-approvals>.

156 Barnabas Reynolds, Thomas Donegan, Simon Dodd and John Adams, Shearman & Sterling, 'The UK's Expanded Senior Managers and Certification Regime: Key Issues and Action Plan for Brokers, Advisors and Asset Managers', *Perspectives* (Blog Post, 8 July 2019) <https://www.shearman.com/

The SMCR is structured around: (1) a senior managers regime for individuals who require regulatory approval; (2) a certification regime for regulated firms to assess the fitness and propriety of employees carrying out a 'significant harm' function;[157] and (3) conduct rules that apply to most bank employees.[158]

Senior managers are required to have a clear and succinct statement of responsibilities. These include regulator-prescribed responsibilities. Conduct rules for senior managers specify a 'Duty of Responsibility' that ensures the firm's business is controlled effectively and they comply with the regulatory framework.[159] Senior managers must take reasonable steps to ensure that responsibility is delegated to an appropriate person, and that the delegated responsibility is effectively discharged.[160] A senior manager must disclose any information of which the PRA or FCA would reasonably expect notice.[161] The FCA has stated the SMCR is not intended to subvert collective responsibility or collective decision-making.[162]

Conduct rules encourage a healthy culture whereby all staff must act with integrity, due skill, care and diligence, openly cooperate with the PRA and FCA, pay due regard to the interests of customers and treat them fairly, and observe standards of market conduct. Firms are accountable for employee conduct and are required to notify the regulator of any breach of the conduct rules.[163]

The scope of the current SMCR is slightly wider than the original 2016 Regime. Senior managers are responsible for the firm's policies and procedures for countering financial crime risks: such as money laundering, sanctions, fraud, tax evasion and cybercrime; compliance with the client assets sourcebook where a firm has authority to hold client's money or assets; and, for asset management firms, value for money assessments, independent director representation, and acting in investors' best interests.[164] Furthermore, such responsibilities also apply to the board[165] (or to an ad-hoc tech committee)[166] where upskilling is often needed.

---

perspectives/2019/07/the-uks-expanded-senior-managers-and-certification-regime-key-issues-and-action-plan>.

157 Financial Conduct Authority, *The Senior Managers and Certification Regime: Guide for FCA Solo-Regulated Firms* (July 2019) 10 <https://www.fca.org.uk/publication/policy/guide-for-fca-solo-regulated-firms.pdf> ('*Senior Managers and Certification Regime Guide*').

158 Linklaters, 'SMCR for Deposit Takers and PRA-Designated Investment Firms' (Web Page) <https://www.linklaters.com/en/insights/publications/smcr/smcr/smcr-for-deposit-takers-and-pra-designated-investment-firms>.

159 *Senior Managers and Certification Regime Guide* (n 157) 14.

160 Ibid.

161 KPMG, *Individual Accountability: Global Regulatory Developments in Financial Services* (Report, July 2018) 4–5.

162 Allen & Overy, *The UK Senior Managers and Certification Regime: Themes, Trends and Challenges from the First Three Years* (Report, March 2019) 17.

163 Patricia Volhard et al, Debevoise & Plimpton, 'The UK's Senior Managers and Certification Regime' (Report, 18 February 2019).

164 Ibid.

165 Magnus Falk, Financial Conduct Authority, 'Artificial Intelligence in the Boardroom', *Insight* (online, 1 August 2019) <https://www.fca.org.uk/insight/artificial-intelligence-boardroom>.

166 Although this would be the proper venue to discuss it, it seems that such committees (which are not even used by the so-called 'GAFAM' big tech five: Google, Apple, Facebook, Amazon and Microsoft) do perform a much more strategic function within the board and do not specifically address AI-related risks. See Maria Lillà Montagnani and Maria Lucia Passador, 'AI Governance

3 *Hong Kong: Securities Firm Managers-in-Charge Regime*

For Hong Kong securities firms, since 2016 senior management are defined as directors, 'responsible officers' of a corporation, and 'Managers-in-Charge'.[167] Licensed corporations must appoint a Manager-in-Charge as primarily responsible for: each core function; overall management oversight; key business lines; operational control and review; risk management; finance and accounting; IT; compliance; and anti-money laundering and combatting the financing of terrorism. For each core function, there should be at least one Manager-in-Charge responsible, although one can manage several core functions (depending on the size and scale of the corporation's operations).

General Principle 9 of the *Code of Conduct for Persons Licensed or Registered with the Securities and Futures Commission* states that senior management shall 'bear primary responsibility for ensuring the maintenance of appropriate standards of conduct and adherence to proper procedures'. A person's actual and apparent authority shall be considered to determine responsibility and its degree.[168] The Board must approve and adopt a formal document clearly setting out roles, responsibilities, accountability, and the reporting lines of senior management.[169]

Paragraph 14.1 of the Code of Conduct specifies that senior management should properly manage the risks associated with a firm's business, including performing periodic evaluation of its risk processes, internal control procedures and risk policies; and understanding the extent of their own authority and responsibilities.[170] Senior management are ultimately responsible for the adequacy and effectiveness of the firm's internal control systems.[171] Managers-in-Charge should be aware of other codes and guidelines that impose responsibilities pursuant to the *Securities and Futures Ordinance* (Kong Kong) (cap 571).[172]

4 *Singapore: Senior Manager Guidelines*

In September 2020, the Monetary Authority of Singapore issued *Guidelines on Individual Accountability and Conduct* that will be effective from 10 September 2021.[173] Senior managers will be responsible for the day-to-day operations of financial institutions in Singapore.[174] The Guidelines make senior managers

---

and Tech Committees: An Empirical Analysis in Europe and North America' (Bocconi Legal Studies Research Paper No 3728946, 9 January 2021) 29–39 <https://ssrn.com/abstract=3728946>.

167 Securities and Futures Commission, 'Circular to Licensed Corporations Regarding Measures for Augmenting the Accountability of Senior Management' (Web Page, 16 December 2016) [5] <https://www.sfc.hk/edistributionWeb/gateway/EN/circular/doc?refNo=16EC68>.

168 Ibid [1], [5], [7]–[9].

169 Ibid [28].

170 Ibid [14](b).

171 Ibid [14](c).

172 Ibid [14], [19].

173 Monetary Authority of Singapore, *Guidelines on Individual Accountability and Conduct* (10 September 2020).

174 Ibid 6 (definition of 'senior managers').

responsible for the management and conduct of 'core management functions', for the actions of their staff, and the conduct of the business.[175] Financial institutions should apply core-management-function definitions that reflect the actual responsibilities of a particular senior manager.[176] Responsibility is described as 'principles-based' and thus there is no list of mandatory responsibilities.[177] The Monetary Authority of Singapore states that the level of responsibility should reflect the senior manager's roles vis-à-vis the financial institution's Singaporean operations.[178] Senior managers are responsible regardless of their title or whether they are based overseas.[179]

5 *European Union*

The EU joint internal governance guidelines were published in 2017 by the European Banking Authority and European Securities and Markets Authority to build upon the *Commission Delegated Regulation (EU) No 604/2014* criteria that identifies categories of staff whose professional activities have a material impact on a financial institution's risk profile.[180] The joint internal governance guidelines aim to satisfy the *CRD IV* and *MiFID II* requirements and are made pursuant to *Directive 2013/36/EU* and *Directive 2014/65/EU*.[181]

The European Banking Authority and European Securities and Markets Authority internal governance guidelines, and European Insurance and Occupational Pensions Authority's guidelines on systems of governance,[182] apply to a variety of financial institutions under EU law. These guidelines govern the conduct of the management body and key function holders. 'Key function holders' refers to persons with significant influence over the direction of the institution who are not part of the management body.[183] The management body and key function holders must possess good repute, independence, honesty, integrity, knowledge, skills, and experience. Members of the management body must have sufficient time to perform their

---

175 Ibid 7 [1.2(i)]. For a definition of core management functions in relation to senior management, see ibid 20–23 (Annex B).

176 Ibid 7 [1.2(ii)].

177 Ibid [4.1]–[4.2].

178 Ibid [1.3].

179 Ibid [1.2].

180 See European Banking Authority ('EBA'), *Final Report: Guidelines on Internal Governance under Directive 2013/36/EU* (Report No EBA/GL/2017/11, 26 September 2017) ('*EBA Guidelines EBA/GL/2017/11*'); European Banking Authority and European Securities and Markets Authority ('ESMA'), *Guidelines on the Assessment of the Suitability of Members of the Management Body and Key Function Holders* (Report No ESMA71-99-598 EBA/GL/2017/12, 21 March 2018) ('*EBA and ESMA Guidelines ESMA71-99-598 EBA/GL/2017/12*'); European Banking Authority and European Securities and Markets Authority, *Final Report: Joint ESMA and EBA Guidelines on the Assessment of the Suitability of Members of the Management Body and Key Function Holders under Directive 2013/36/EU and Directive 2014/65/EU* (Report No EBA/GL/2017/12, 26 September 2017) ('*ESMA and EBA Joint Guidelines EBA/GL/2017/12*').

181 *EBA Guidelines EBA/GL/2017/11* (n 180) 5–7. See also above nn 114, 116.

182 European Insurance and Occupational Pensions Authority, *Guidelines on System of Governance* (EIOPA-BoS-14/253 EN, January 2014) <https://www.eiopa.europa.eu/content/guidelines-system-governance_en>.

183 Ibid [1.21].

functions including understanding the business of the institution, its main risks, and the implications of the business and risk strategy.[184]

Responsibilities of the management body (in particular, the Chief Executive Officer and other key executives) include setting, approving, and overseeing implementation of the overall business strategy and the key legal and regulatory policies, overall risk strategy, internal governance and control, risk capital, liquidity targets, remuneration policy, key function holders' assessment policy, internal committees functionality, risk culture, corporate culture, conflict of interest policy, and the integrity of accounting and financial reporting systems.[185] The management body is also accountable for the implementation of the governance arrangements that ensure effective and prudential management of the institution, and promote market integrity and client interests.[186]

Key function holders such as heads of internal control functions including risk management, compliance and audit functions have a crucial role in ensuring that the institution adheres to its risk strategy, complies with legal and regulatory requirements, and has robust governance arrangements.[187] A sound and consistent risk culture is a critical element of risk management. Key function holders should know and understand the extent of risk appetite and risk capacity for their role and contribute to internal communications regarding the institution's core values and staff expectations. They should promote an environment of open communication, welcoming challenges in decision-making, encouraging a broad range of views and the testing of current practices, stimulating a constructive critical attitude, and promoting an environment of open, constructive engagement throughout the entire organisation.[188] The proportionality principle applies to all governance arrangements, consistent with the institution's risk profile and business model.[189]

The European Commission's DORA Proposal aims at addressing the digital operational resilience needs of all EU-regulated financial entities and fine-tunes the aforementioned principles with a view to ICT risks in general.[190] While not mentioning AI in particular, the DORA Proposal's definition of ICT risk is all-encompassing and includes AI-related malfunctions of any kind.[191] While details of the Proposal exceed the scope of this article, its most important principle in the context of this article states that '[t]he management body of the financial entity shall define, approve, oversee and be accountable for the implementation of all arrangements related to the ICT risk management framework'.[192]

---

184 *EBA and ESMA Guidelines ESMA71-99-598 EBA/GL/2017/12* (n 180) 3 [6], 5–6 [15], 11 [26], 13 [37], 14 [39], 14 [41].

185 *ESMA and EBA Joint Guidelines EBA/GL/2017/12* (n 180) 28–42 [41]–[93] (Title III).

186 *EBA and ESMA Guidelines ESMA71-99-598 EBA/GL/2017/12* (n 180) 6, 11 [26], 13 [37], 14 [41], 31 [110].

187 Ibid 11 [33].

188 *EBA Guidelines EBA/GL/2017/11* (n 180) 34 [98].

189 *EBA and ESMA Guidelines ESMA71-99-598 EBA/GL/2017/12* (n 180) 9 [20].

190 DORA Proposal (n 144) Explanatory Memorandum.

191 Ibid art 3(4).

192 Ibid art 4(2).

6 *International Organization of Securities Commissions Consultation*

In June 2020, the Board of the International Organization of Securities Commissions ('IOSCO') published a consultation report relating to guidance on the use of AI and machine learning by market intermediaries and asset managers.[193] The very first measure of the *IOSCO AI Consultation Report* is that '[r]egulators should consider requiring firms to have designated senior management responsible for the oversight of the development, testing, deployment, monitoring and controls of AI and machine learning.'[194] If implemented at the national level, this guidance will help instil a personal responsibility framework for securities regulators across the world precisely along the lines of that for which we argue here for all financial institutions.

B *Addressing the Knowledge Gap*

The trend in financial services regulation is clear: ever-increasing personal responsibility for senior management and other individuals responsible for regulated activities within financial institutions. We argue here that such frameworks are instrumental in addressing AI-related risks.

Personal responsibility frameworks can underpin a system of addressing issues arising from AI in finance, in particular the three challenges of AI (information asymmetry, data dependency and interdependency).[195] Manager responsibility frameworks should be expanded to specifically incorporate responsibility for AI in regulated activities, thus mandating a human in the loop, especially for due diligence, fairness and explainability requirements. In many cases, this approach could be augmented by additional AI review committees. These can be highly effective in addressing black box issues and in providing a framework to address the four core financial risks relating to data, cybersecurity, systemic risk, and ethics.

1 *AI Due Diligence*

The first tool reinforcing and supporting manager responsibility is mandatory AI due diligence. Due diligence should include a full stocktaking of all characteristics of the AI. At a minimum, this must include the AI explainability standard further described in the next section. AI due diligence should be a requirement prior to AI procurement, adoption and deployment, while AI explainability is the standard to meet throughout the use of any AI to internal and external stakeholders.

---

193 Board of the International Organization of Securities Commissions ('IOSCO'), *The Use of Artificial Intelligence and Machine Learning by Market Intermediaries and Asset Managers: Consultation Report* (Report CR02/2020, June 2020) <https://www.iosco.org/library/pubdocs/pdf/IOSCOPD658.pdf> ('*IOSCO AI Consultation Report*').

194 Ibid 2, 18.

195 See above Part III(F).

To reflect data dependency, one part of AI due diligence is mapping the data sets used by the AI, including an analysis of dataset bias, data gaps and data quality.[196]

AI due diligence is key to individual responsibility systems: individuals need to conduct sufficient due diligence in the exercise of their responsibilities to avoid liability for any failures, whether from internal governance systems, employees, third parties, or ICT systems.

2 *AI Explainability*

Explainability requirements are necessary minimum standards for humans in the loop — that is, demanding that AI functions, limits and risks can be explained *to someone*. Debates exist relating to the level of granularity required and to whom such explanations should be made (for example, a programmer/statistician, user, or regulator),[197] and the term 'interpretability' is sometimes used in the context of more technical explainability.

From a regulatory approach, this 'someone' could be an appropriate senior manager and/or a member of the executive board responsible for the AI (relying on the manager's incentive to avoid sanctions) or an external institution, in particular regulators, supervisors and courts. From a consumer rights perspective, this 'someone' could be the ultimate user of the technology (as has been alluded to under the *GDPR*).[198]

Based on this analysis, first, we encourage financial regulators to introduce explainability requirements for responsible managers, including documentation and governance requirements, with a clarification of the standards depending on a risk and impact assessment and to whom the explanation is required. Second, supervisory authorities should review compliance with explainability requirements. Manager responsibility systems will thus be buttressed by explainability systems, which in turn result from personal responsibility and accountability to regulators. As with their other decisions, individual senior managers must be able to explain and take responsibility for their own direct or indirect decisions about technology, the actions of their employees and contractors and, critically, the decisions of their AI systems at least to their regulators.

3 *AI Review Committees*

In addition to due diligence and explainability requirements to address the information asymmetry concerning AI's functions and limits, financial regulators

---

196 Brian W Tang, 'The Chiron Imperative — A Framework of Six Human-in-the-Loop Paradigms to Create Wise and Just AI-Human Centaurs' in Susanne Chishti, Sophia Adams Bhatti, Akber Datoo, Drago Indjic (eds), *The LEGALTECH Book: The Legal Technology Handbook for Investors, Entrepreneurs and Fintech Visionaries* (Wiley, 2020) 38.

197 Aleksandra Mojsilovic, 'Introducing AI Explainability 360', *IBM Research Blog* (Blog Post, 8 August 2019) <https://www.ibm.com/blogs/research/2019/08/ai-explainability-360/>; Tang (n 77) 243.

198 *GDPR* (n 105).

should create independent AI review committees to provide cross-disciplinary and impartial expertise. This is an important practice emerging in some non-financial companies.[199] Some of these committees have been quite impactful, such as in Axon's management and board accepting the recommendation of its AI and Policing Ethics Board to impose a moratorium on the use of facial recognition in Axon's body cameras.[200] The impact of other committees has been less,[201] or remains to be seen.[202] Regardless, these committees are designed to augment decision-making and should not detract from the ultimate responsibility vested in management and the board regarding AI governance.

C *Personal Responsibility in Financial Regulation: Challenges in Building Human-in-the-Loop Systems*

Several concerns relating to the personal responsibility model require consideration. These include: (1) inability to fully control AI using internal governance; (2) unwillingness to curtail highly profitable AI; (3) tacit collusion between AI systems; (4) over-deterrence of innovation; and (5) the differing attitudes to AI and technology in financial services.

1 *Inability to Fully Control AI Internally*

If AI cannot be controlled by external monitors, such as financial supervisors, it may be argued that AI cannot be monitored and controlled effectively by senior management not directly involved in AI data gathering, coding and operations.

Existing methods of internal control include: internal reporting; defining risk limits in terms of risk budgets; assigning budgets for code development and data pool acquisition; and setting adequate incentives through balanced compensation models. Personal responsibility/liability systems place the responsibility for regulated conduct areas upon specific individual senior managers. Thus, a senior manager who is directly responsible for regulatory breaches arising in their area of responsibility will have strong incentives to innovate and strengthen the existing governance tools to monitor and better understand their functional area, staff, third-party contractors and suppliers, and IT systems. A culture of due diligence and explainability should then evolve to address the black box problem. Where it does not, the individual and board will nonetheless remain responsible for any harm caused.

---

199 Brian W Tang, 'Independent AI Ethics Committees and ESG Corporate Reporting on AI as Emerging Corporate and AI Governance Trends' in Susanne Chishti, Ivana Bartoletti, Anne Leslie and Shân M Millie (eds), *The AI Book: The Artificial Intelligence Handbook for Investors, Entrepreneurs and FinTech Visionaries* (Wiley, 2020) 180, 183.

200 Rick Smith, 'The Future of Face Matching at Axon and AI Ethics Board Report', *Axon* (Web Page, 27 June 2019) <https://www.axon.com/news/ai-ethics-board-report>.

201 Parmy Olson, 'Google Quietly Disbanded Another AI Review Board Following Disagreements', *Wall Street Journal* (online, 16 April 2019) <https://www.wsj.com/articles/google-quietly-disbanded-another-ai-review-board-following-disagreements-11555250401>.

202 Brent Harris, 'Establishing Structure and Governance for an Independent Oversight Board', *Facebook Newsroom* (Web Page, 17 September 2019) <https://about.fb.com/news/2019/09/oversight-board-structure>.

Naturally, the manager responsibility model requires those involved in AI development, procurement and deployment to be included within the net of responsibility. As argued in relation to techrisk, an individual should be designated as responsible for IT and technology systems.[203]

One concern often raised against a manager responsibility concept is where self-learning AI taps into unexpected or malicious data input, and produces unexpected correlations or unacceptable outcomes. However, as in the case of Microsoft's Tay, this can be countered by the proverbial mandatory 'AI off switch' depending on the risk and impact assessment and an appropriate contingency or business continuity plan. Such an 'AI kill switch' was expressly mentioned in an explanatory note to Measure 2 of the *IOSCO AI Consultation Report*, although not in the text of Measure 2 itself.[204] The extent of this will certainly depend on the AI application, but the fact AI can impose risks on clients, the financial institution and the financial system is all the more reason to rigorously analyse and scrutinise AI use in finance.

2 *Unwillingness to Curtail Highly Profitable AI*

A common issue in financial institution governance is the unwillingness to curtail profitable, yet complex, conduct. We draw analogies with the 2008 Global Financial Crisis: even though senior managers found difficulty in understanding the true risk of tranched and structured finance as well as its allocation, they had little incentive to stop complex and opaque, but highly profitable, business models — especially as they benefited from the higher profitability through enhanced pay and reputation. This argument is especially relevant in light of the recent growth of less regulated tech companies that offer new financial services and products.

This manifestation of agency risk is perennial in corporate and financial governance. While our proposal does not change management's incentives from the standpoint of profitmaking, the implementation of personal responsibility impacts directly through individual responsibility for failures, thus incentivising individual and managerial due diligence and efforts to ensure sustainability. AI review committees add another level of oversight and input, and another avenue through which explainability can be sought (in addition to managers with individual responsibility for AI activities and overall board responsibility).

3 *Tacit Collusion between AI Systems*

The profitability of tacit collusion among AI systems poses particular challenges. Accordingly, competition authorities are increasingly focused on this issue.[205]

The WEF has suggested this be mitigated by:

---

[203] Buckley et al (n 72) 61.

[204] *IOSCO AI Consultation Report* (n 193) 19–20.

[205] Bundeskartellamt and Autorité de la Concurrence, *Algorithms and Competition* (Report, November 2019); UK Competition and Markets Authority, *Pricing Algorithms: Economic Working Paper on the Use of Algorithms to Facilitate Collusion and Personalised Pricing* (October 2018).

(i) restricting AI-enabled systems communication with their environments to 'explicitly justifiable business purposes';[206]

(ii) ensuring their AI-enabled systems' decisions are explainable by 'valid, legal business reasons';[207] and

(iii) requiring humans to oversee decisions made by AI-enabled systems.[208]

These are good suggestions, but may not always be sufficient to fully mitigate this substantial risk, in particular when collusion is highly profitable. In the end, this comes down to the unwillingness dimension discussed above in Part V(C)(2): personal responsibility requirements address these, particularly when supplemented by review committee, due diligence and explainability requirements that all come with enhanced documentation and potentially severe liability and director and managerial disqualification resulting from lack of oversight.

4 *Over-Deterrence of Innovation*

At the same time, manager responsibility may be too much of a good thing. If the regulatory burden excessively deters good managers from being involved in AI-based financial services, we may find a reduction in innovation in finance and corollary reductions in efficiency, access to justice and combatting of financial crime, and/or leadership by less thoughtful and reflective people serving as senior managers for financial services firms. Well-intentioned global regulation may also lead to unintended consequences that disadvantage financial institutions, fintechs and technology companies from emerging economies seeking to deploy AI.[209] Regulators must respond to this concern with proportional 'carrots' to incentivise and recognise good actors as well as 'sticks' for irresponsible conduct.[210] Personal responsibility liability systems should also include continuing education frameworks.

Individual responsibility could lead to decreased diligence in monitoring fellow key function holders. Conversely, collective responsibility could increase monitoring among key function holders, but lead to over-deterrence. This debate is underscored by the Australian Westpac bank scandal — a potent example of the potential magnitude of techrisk.[211] The bank had developed its own software to implement and govern remittances, and a relatively innocuous looking piece of software allegedly permitted 23 million anti-money laundering breaches.[212] The

---

206 World Economic Forum (n 60) 118.

207 Ibid.

208 Ibid.

209 Fintech Association of Hong Kong and LITE Lab@HKU, 'Joint Response to IOSCO Consultation on "The Use of Artificial Intelligence and Machine Learning By Market Intermediaries and Asset Managers"' (23 October 2020) <https://ftahk.org/system/files/2020-10/FTAHK%20LITELabHKU%20IOSCO%20AI%20Consultation_October%2023.pdf>.

210 Brian W Tang, 'Promoting Capital Markets Professionalism: An Emerging Asian Model' in Ross P Buckley, Emilios Avgouleas and Douglas W Arner (eds), *Reconceptualising Global Finance and its Regulation* (Cambridge, 2016) 357, 389.

211 Buckley et al (n 72) 40–1.

212 See above n 80 and accompanying text.

breaches attracted a massive financial penalty and arguably even more reputational damage for the bank.[213]

To avoid or limit over-deterrence of innovation, a compromise would include defining some collective core duties, while also imposing individual responsibility. This should apply to both board and corporate responsibility.

Regulators usually require finance experience as a precondition for licensing a financial entity. Technology start-up founders often have little experience in running a regulated firm. If regulators require this expertise of all key function holders in a start-up, innovation will be severely impaired. One obvious response is for regulators to require sufficient expertise and experience from the fintech start-up's board and key executives as a group. Therefore, some board members and executives can contribute the IT/AI expertise,[214] while others contribute experience in running regulated financial services firms. Gradually, all board members and executives should be able to meet the standards for seasoned financial intermediaries.

For personal responsibility in given areas, specific area-related expertise is required as one aspect of the fit-and-proper test. While it may make sense in a fintech start-up to take a balanced and proportionate approach to board and key executive requirements as a group, specifically mandated individual responsibility requirements, expertise and experience requirements would remain necessary in the licensing process.

5 *Differing Attitudes to AI and Technology in Financial Services*

Our final, and perhaps most important, recommendation goes to the cultural attitude of many in financial services towards AI and technology in general. There is much talk about the trust crisis in our modern world of fake news and low institutional credibility. But we do not need to trust AI more in financial services (or in medicinal care or criminal sentencing or other applications). We need AI to demonstrate its trustworthiness.

---

[213] Paul Smith, 'Westpac's Tech Mess Could Happen to Anyone', *Australian Financial Review* (online, 6 December 2019) <https://www.afr.com/technology/westpac-s-tech-mess-could-happen-to-anyone-20191204-p53gqq>; Stephen Bartholomeusz, 'Mission Impossible? Westpac Panel Highlights Directors' Dilemma', *The Sydney Morning Herald* (online, 4 June 2020) <https://www.smh.com.au/business/banking-and-finance/mission-impossible-westpac-panel-highlights-directors-dilemma-20200604-p54zfp.html>; James Frost and James Eyers, 'Westpac's Risk Culture Deemed "Immature and Reactive"', *Australian Financial Review* (online, 4 June 2020) <https://www.afr.com/companies/financial-services/westpac-scandal-sloppy-not-a-conspiracy-20200604-p54zbq>.

[214] Montagnani and Passador (n 166) 9 (including n 35), 40–41. Such skill may effectively contribute to the proper selection of the perfect 'black box' (more correctly, AI tool) for that specific company, and become as essential as a legal or economic background among directors now often is. More specifically, as a consequence, 'boards will incorporate these features and will be able to independently equip themselves with the most suitable composition to fully understand those mechanisms and, specifically, to choose the most suitable AI system for their specific company, thereby ensuring the utmost accountability of AI systems employed for predictive goals': at 42.

Topol, in his authoritative review of medical AI, states that '[t]he state of AI hype has far exceeded the state of AI science, especially when it pertains to validation and readiness for implementation in patient care'.[215]

Spiegelhalter's recent article illuminates these issues succinctly and we recommend it highly.[216] In his words:

> It seems reasonable that, when confronted by an algorithm, we should expect trustworthy claims both:
>
> 1. *about* the system—what the developers say it can do, and how it has been evaluated, and
> 2. *by* the system—what it says about a specific case.[217]

Spiegelhalter suggests anyone seeking to purchase or use an AI system should ask these questions about it:

1. Is it any good when tried in new parts of the real world?
2. Would something simpler, and more transparent and robust, be just as good?
3. Could I explain how it works (in general) to anyone who is interested?
4. Could I explain to an individual how it reached its conclusion in their particular case?
5. Does it know when it is on shaky ground, and can it acknowledge uncertainty?
6. Do people use it appropriately, with the right level of skepticism?
7. Does it actually help in practice?[218]

These questions strongly appeal for their directness and simplicity. We have seen senior finance professionals, including in some major Australian banks, unwilling to insist on what their organisation really needs in its AI, and accept instead assurances or explanations from AI developers that they would not accept from other service suppliers. The reason seems to be the apprehension or lack of understanding many senior people have about AI and technology generally. In one of the most regulated of all industries, financial services, these attitudes are inappropriate. Spiegelhalter's seven questions provide a highly useful checklist in this regard. What is needed at the most senior levels of major banks, and within their in-house legal departments, is a cultural shift. Instead of the hesitancy and apprehension that often characterises current approaches to AI and technology more generally, these tools need to be approached with confidence, humility and the understanding that they can and must be held to perform at the required standards, and can be built to do so, if the procuring institution insists.

---

[215] Eric J Topol, 'High-Performance Medicine: The Convergence of Human and Artificial Intelligence' (2019) 25(1) *Nature Medicine* 44, 51.

[216] David Spiegelhalter, 'Should We Trust Algorithms?' (2020) 2(1) *Harvard Data Science Review* <https://hdsr.mitpress.mit.edu/pub/56lnenzj/release/1>.

[217] Ibid [2 (Trust and Trust Worthiness)] (emphasis in original).

[218] Ibid [5 (Conclusions)].

VI Conclusion

The financial services sector globally is one of the leaders in AI use and development. However, AI comes with numerous technical, ethical and legal challenges that can undermine the objectives of financial regulation with respect to data, cybersecurity, systemic risk and ethics — in particular, relating to black box issues.

As shown, traditional financial supervision focused on external governance is unlikely to sufficiently address the risks created by AI, due to: (1) enhanced information asymmetry; (2) data dependency; and (3) interdependency. Accordingly, even where supervisors have exceptional resources and expertise, supervising the use of AI in finance by traditional means is extremely challenging.

To address this weakness, we suggest that internal governance of financial institutions be strengthened to impose personal responsibility requirements to put a human in the loop. This approach is based on existing frameworks of managerial responsibility that evolved in the aftermath of the 2008 Global Financial Crisis and of a seemingly continuing stream of ethically questionable behaviour across the world in finance. These frameworks should be cognisant of and consistent with broader data privacy and human-in-the-loop approaches beyond finance.[219] From a financial supervisor's perspective, internal governance can be strengthened largely through a renewed focus on senior managements' (or key function holders') personal responsibilities and accountability for regulated areas and activities, as designated for regulatory purposes. These key function holder rules — particularly if enhanced by specific AI due diligence and explainability requirements — will assist core staff of financial services firms to ensure that any AI is performing in ways consistent with the senior managers' personal responsibilities. The key function holders or managers-in-charge are responsible for themselves, their area of supervision, their staff, their third-party contractors, and their technology, including AI.

This direct personal responsibility encourages due diligence in investigating new technologies, their uses and impact, and on requiring fairness and explainability as part of any AI system, with attendant dire personal consequences for failure. For a financial services professional with direct responsibility, demonstrating appropriate due diligence and explainability will be key to a personal defence in the event of a regulatory action. This approach will also prove helpful to address the other data, cybersecurity, systemic risk, and ethical issues relating to AI in finance, particularly when combined with new AI review committees that can augment the decision-making and collective responsibility of senior management and the board.[220]

Importantly, this approach — while a natural evolution in the context of financial regulation — also has great potential for addressing AI concerns in any

---

[219] Tang (n 196).

[220] Also within the board of directors, sound AI governance should be fostered and encouraged. Montagnani and Passador propose, at an operational level, enhancing board characteristics to this end, and, at a systemic level, avoiding the neglect of the ethical implications of using AI systems: Montagnani and Passador (n 166) 40–43. Creating an ad hoc committee (a tech committee) and assigning it tech-related functions, or specifically providing other board committees (mainly the audit or the risk management committees) with such tasks, might help in this regard.

other regulated industry facing black box issues arising from AI. While this does not necessarily address the macro issues emerging as a result of the Fourth Industrial Revolution, it will at least ensure that humans are central to the evolution of AI in already regulated industries. As it seems inevitable that AI will play a growing role in our lives and world it is imperative that we put humans in the loop in this human–machine relationship.[221]

[221] Tang (n 196).