CFTE Centre for Finance, Technology and Entrepreneurship

CFTE ACADEMIC PAPER SERIES

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

Dirk A. Zetzsche | Douglas Arner
Ross Buckley | Brian W. Tang

VOLUME 1 2020

Artificial Intelligence in Finance:
Putting the Human in the Loop

Dirk A. Zetzsche[*]

Douglas Arner[**]

Ross Buckley[***]

Brian W. Tang[****]

February 2020

Finance has become one of the most globalized and digitized sectors of the economy. It is also one of the most regulated of sectors, especially since the 2008 Global Financial Crisis. Globalization, digitization and money are propelling AI in finance forward at an ever increasing pace.

This paper develops a regulatory roadmap for understanding and addressing the increasing role of AI in finance, focusing on human responsibility: the idea of "putting the human in the loop" in order in particular to address "black box" issues.

Part I maps the various use-cases of AI in finance, highlighting why AI has developed so rapidly in finance and is set to continue to do so. Part II then highlights the range of the potential issues which may arise as a result of the growth of AI in finance. Part III considers the regulatory challenges of AI in the context of financial services and the tools available to address them, and Part IV highlights the necessity of human involvement.

We find that the use of AI in finance comes with three regulatory challenges: (1) AI increases information asymmetries regarding the capabilities and effects of algorithms between users, developers, regulators and consumers; (2) AI enhances data dependencies as different day's data sources may may alter operations, effects and impact; and (3) AI enhances interdependency, in that systems can interact with unexpected consequences, enhancing or diminishing effectiveness, impact and

[*] Professor of Law, ADA Chair in Financial Law (Inclusive Finance), Faculty of Law, Economics and Finance, University of Luxembourg, and Director, Centre for Business and Corporate Law, Heinrich-Heine-University, Düsseldorf, Germany.

[**] Kerry Holdings Professor in Law and Director, Asian Institute of International Financial Law, Faculty of Law, University of Hong Kong; Adviory Board Member, Centre for Finance, Technology and Education.

[***] KPMG Law and King & Wood Mallesons Chair of Disruptive Innovation, Scientia Professor, and Member, Centre for Law, Markets and Regulation, UNSW Sydney. Professor Buckley chairs the Digital Finance Advisory Panel of the Australian Securities and Investments Commission (ASIC) however the views expressed herein are strictly his own, not those of ASIC.

[****] Founding Executive Director, LITE Lab@HKU, Faculty of Law, University of Hong Kong, Co-chair of the FinTech Association of Hong Kong's RegTech Committee, Co-Founder, Asia Capital Market Institute (ACMI) and member of the IEEE Global Initiative in Autonomous and Intelligent Systems' Policy Committee.

The authors thank the Hong Kong Research Grants Council Research Impact Fund and the Qatar National Research Fund National Priorities Programme for financial support.

explainability. These issues are often summarized as the “black box” problem: no one understands how some AI operates or why it has done what it has done, rendering accountability impossible.

Even if regulatory authorities possessed unlimited resources and expertise – which they clearly do not – regulating the impact of AI by traditional means is challenging.

To address this challenge, we argue for strengthening the *internal* governance of regulated financial market participants through external regulation. Part IV thus suggests that the most effective path forward involves regulatory approaches which bring the human into the loop, enhancing *internal* governance through *external* regulation.

In the context of finance, the post-Crisis focus on personal and managerial responsibility systems provide a unique and important external framework to enhance internal responsibility in the context of AI, by putting a human in the loop through regulatory responsibility, augmented in some cases with AI review panels. This approach – AI-tailored manager responsibility frameworks, augmented in some cases by independent AI review committees, as enhancements to the traditional three lines of defence – is in our view likely to be the most effective means for addressing AI-related issues not only in finance – particularly “black box” problems – but potentially in any regulated industry.

Contents

Introduction

The concept of artificial intelligence – AI – is the focus of much global attention today.[1] While AI has a long history of development, technological advances combined with ever-widening digitization have underpinned recent rapid and unprecedented evolution. Central to the "Fourth Industrial Revolution" and the "digitization of everything" is the impact of datafication – manipulation of digitized data through quantitative data analytics, including AI.[2]

From a positive standpoint, AI is expected to contribute to problem solving in and development of most sectors of the economy and society. PwC's optimistic expectations are that AI will boost global GDP by 14% or US$15.7 trillion – by 2030.[3] In the context of finance, Accenture estimates that banks can expect potential savings of between 20% and 25% across IT operations, including infrastructure, maintenance and development costs.[4] The combination of cost savings and enhanced efficiency combined with the potential for entirely new business models and opportunities explains why financial services companies are expected to spend a US$11 billion on AI in 2020, more than any other industry.[5]

At the same time, AI and automation are raising major concerns, ranging from widespread job losses[6] to the possible advent of the "singularity": the point at which the capacities of general AI surpass that of humans in essentially every way. These concerns have triggered an increasing range of analyses of the policy, legal and regulatory implications of AI, from ethical dimensions[7] to legal restrictions.[8] Central to many of these discussions are the role of humans in the evolution of AI: the necessity of involving people in using, monitoring and supervising AI in order to reduce the likelihood of problems arising and their severity. This is the idea of putting a "human in the loop", and it is the challenge at the heart of AI governance discussions in all sectors, all over the world.

[1] *See*, for instance, the literature survey by Bonny G. Buchanan, "Artificial intelligence in finance services" (The Alan Turing Institute, April 2019) < https://www.turing.ac.uk/sites/default/files/2019-04/artificial_intelligence_in_finance_-_turing_report_0.pdf>.

[2] *See* UK Finance and Microsoft, "Artificial Intelligence in Financial Services" (27 Jun. 2019) 5 < https://www.ukfinance.org.uk/policy-and-guidance/reports-publications/artificial-intelligence-financial-services>.

[3] PricewaterhouseCoopers, "Sizing the prize: What's the real value of AI for your business and how can you capitalise?" (2017) 4< https://www.pwc.com/gx/en/issues/analytics/assets/pwc-ai-analysis-sizing-the-prize-report.pdf>.

[4] AI Accenture, "Redefine Banking with Artificial Intelligence" (2018) 9 > https://www.accenture.com/_acnmedia/pdf-68/accenture-redefine-banking.pdf>.

[5] *See* International Data Corporation (IDC), report May 2019, cited by Amy Zirkle, The Critical Role of Artificial Intelligence in Payments Tech, 27 May 2019, < https://www.fintechnews.org/the-crirital-role-of-artificial-inteliigence-in-payments-tech/>.

[6] *See* Shelly Hagan, More Robots Mean 120 Million Workers Need to be Retrained', 6 Sept 2019, https://www.bloomberg.com/news/articles/2019-09-06/robots-displacing-jobs-means-120-million-workers-need-retraining (citing an IBM survey stating that 120 million jobs will be lost due to AI within the next 3 years).

[7] *See* 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 (eds), *Towards Digital Enlightenment* (Springer, 2018).

[8] *See*, as one of the earlier scholarly articles, Harry Surden, "Machine learning and the law", 89 Wash. L. Rev. 87 (2014).

In the context of AI and AI governance, one area which has until very recently received relatively less attention is the role of AI in finance.[9] This is surprising because finance has become one of the most globalized and digitized sectors – if not the most globalized and digitized sector – of the world's economy. It is also one of the most regulated of sectors, especially since the 2008 Global Financial Crisis. Not surprisingly, AI is already playing an important role in finance, and one that is only likely to grow due to the nature of the financial industry and the ongoing process of global digital financial transformation. As a result, issues around AI and AI governance are growing in significance in finance. Finance however as a result of regulatory developments since the Global Financial Crisis, also provides an important opportunity to address the human in the loop challenge.

This paper develops a regulatory framework for understanding and addressing the increasing role of AI in finance, focusing on human responsibility within the context of putting the "human-in-the-loop" as a core approach in addressing "black box" problems with AI.

Part I maps the various use-cases of AI in finance, highlighting why AI is developing so rapidly in finance. Part II highlights the range of potential issues which may arise as a result of the growth of AI in finance. Part III considers the regulatory challenges of AI in the context of financial services and the tools available to address them, highlighting the necessity of human involvement. Part IV argues that the most effective path forward involves regulatory approaches which bring the human into the loop, enhancing *internal* governance and reducing financial supervision as *external* governance.

In the context of finance, the post-Crisis focus on personal and managerial responsibility systems provides a unique and important external framework to enhance internal responsibility in the context of AI, by putting a human-in-the-loop[10] through regulatory responsibility, as enhancements to the traditional three lines of defence, augmented in some cases with AI review panels.

We argue in Part V that this approach is central not only to addressing AI in finance but also potentially in any regulated industry which faces "black box" challenges in the context of AI or other new technologies.

I. AI and Finance

To consider AI in finance we first consider AI and its increasingly rapid development before turning to the particular characteristics of finance which make it highly suitable for AI and the range of uses which are rapidly evolving.

[9] For recent treatment, *see* Tom C.W. Lin, *Artificial Intelligence, Finance, and the Law*, 88 FORDH. L. REV. 531 (2019) (summarizing risks and limitations of AI in light of financial regulation).

[10] For a proposed Human-In-the-Loop framework, *see* Brian W Tang, "The Chiron Imperative – A Framework of Six Human-in-the-Loop Paradigms to Create Wise and Just AI-Human Centaurs" in Sophia Adams Bhatti, Susanne Chishti, Akber Datoo and Drago Indjic (ed), *The LEGALTECH Book: The Legal Technology Handbook for Investors, Entrepreneurs and Fintech Visionaries* (Wiley, forthcoming 2020).

A. AI and the Digitization of Everything

The term AI covers a series of technologies and approaches, ranging from “if-then” rule-based expert systems, to the interdisciplinary approach of combining linguistics and computer science known as natural language processing (NLP), as well as the marriage of algorithms and statistics known as machine learning that results in pattern recognition and inference from being trained from data rather than explicit human instructions. The increasing complexities of the latter seem to progressively reduce the role of humans as AI systems expand from supervised learning to unsupervised deep learning neural networks, reinforcement learning, collaborative learning, transfer learning and generative adversarial networks (GANs).

AI has been the focus of attention periodically over the past five decades. However, a unique confluence of factors has dramatically altered its developmental trajectory and as a result AI’s evolution is raising an increasing range of issues, from the mundane to the existential.

There are five key factors which today empower the rapid development, training and evolution of AI: data, storage, communication, computing power, and analytics.

Rapid developments are noteworthy with regard to all of these factors: From the standpoint of data, the core aspect is digitization. It is only once data become available in digital form that the process of datafication – the application of analytics including AI – becomes effective. Thus, the “digitization of everything” at the heart of the Fourth Industrial Revolution is central to the rapid evolution of AI.[11] For datafication, the volume of data is important as well as its digitization: larger volumes of data are more effective in supporting datafication and in particular machine learning (ML) processes and the “training” of AI systems. Data storage, data storage quality and capacity have dramatically increased while costs have gone down. Thus, the volumes of data being digitally captured and stored now dwarf those captured and stored earlier. This combination of digitization and storage underpins datafication and AI.

Central to digital capture and storage are communications, with internet, mobile phones and the internet of things making it ever more possible to capture, store (locally and/or remoting), transfer, manipulate, and analyse data, increasingly on almost anything. With advances in computer vision, internet of things (IoT), analytics, and online and mobile penetration and usage, we can reasonably expect more and more data to be generated given that all these cloud connected devices have, compared to humans, effectively unlimited capacity to collect and store data.

Datafication also requires computing power and this has also increased dramatically, following Moore’s Law, with dramatic reductions in cost. The emergence of quantum computing, if realised, will open incredible new avenues of processing. Datafication – while relying on computing power – also relies on research and development into algorithms and analytical processes themselves and this is another area of very rapid development.

This digitization of everything lies at the heart of the Fourth Industrial Revolution, ever-falling storage prices, telecommunications that link us all and to the cloud, ever-increasing computing power, and innovative algorithmic and analytical development underlies the explosion in datafication processes, which all in turn fuel AI growth that

[11] See Klaus Schwab, *The Fourth Industrial Revolution* (World Economic Forum, 2016).

looks set to continue, to the extent where discussions of the potential of the singularity are no longer the realm of science fiction.

B. AI and Digital Finance

These features come together uniquely in the contxt of finance.

After five decades of digital transformation, encompassing digitization and datafication, finance is the most globalized, digitized *and* datafied segment of the world's economy. While financial services have always integrated technical innovation,[12] this is particularly true for the latest wave of innovation referred to as financial technology (FinTech).

This process can be seen across four major axes: the emergence of global wholesale markets, an explosion of FinTech startups particularly since 2008, an unprecedented digital financial transformation in developing countries particularly China, and the increasing role of large technology companies (BigTech) in financial services (TechFin).

While finance and technology have always developed in tandem, since the 2008 Global Financial Crisis the changes have been unprecedented, particularly in terms of speed of change and range of new entrants including FinTech and BigTech firms. Speed of change can be seen particularly in the role of new technologies, often summarized under the ABCD framework: AI / analytics, blockchain, cloud and data, which are co-evolving at an increasing rate within finance. Many would also add mobile internet and IoT to these factors. Digital financial transformation combined with certain other aspects of finance make financial services particularly, and perhaps uniquely, fertile for AI development: these aspects include data, financial resources, human resources, and incentives.

As we have seen, one major technological pillar of digital financial transformation is the large-scale use of data: the financial sector has thus cultivated, over a long period, the extensive structured collection of many forms of data (e.g. stock prices). Such data have been standardized and digitized since the 1970s, with new forms of capture and collection constantly emerging. As a result, data in finance provides particularly fertile ground for AI, and finance provides the incentives and resources for the application of ever more sophisticated forms of analytics to ever wider ranges of data.

Furthermore, AI tends to perform best in rule-constrained environments, such as games like chess or Go, where there are a finite – although perhaps very large – number of possibilities to achieve specified objectives. This is the environment in which AI seems to outperform humans with increasing rapidity. This environment exists in many aspects of finance, for instance stock market investment, where there are specific objectives (maximizing profit) and set parameters of action (the trading rules and regulatory system) combined with massive amounts of data. Add technological possibility, in terms of computing power and analytics, to the financial and human resources and incentives to use them and it is apparent why finance is already transforming so rapidly as a result of digitization and datafication, and why this is likely to increase with further development of AI.

[12] See Douglas W. Arner, Janos N. Barberis & Ross P. Buckley, "The Evolution of FinTech: A New Post-Crisis Paradigm?", 47(4) Georgetown Journal of International Law 1345, 1345-1393 (2016).

The latter three – financial resources, human resources and incentives are fairly obvious: financial intermediaries generate massive amounts of income for their stakeholders, including management, investors and employees. As a result, they attract some of the very best human resources into the industry. Those human and financial resources have very strong reasons to continually search for advantages and opportunities for profit and thus invest substantial amounts in research, analytics and technology, to such an extent that there is an entire academic field – finance – focusing exclusively on research in the area and with major teams at financial institutions, advisory firms and academic institutions heavily focused on continually developing better analytical models for finance and investment. Since the 1980s, this process has had a very strong quantitative focus, involving the application of analytics to financial and other forms of data, and it is in the area of data where finance is perhaps unique from the standpoint of AI.

While finance and technology have always developed in tandem, since the 2008 Global Financial Crisis the changes have been unprecedented, particularly in terms of speed of change and range of new entrants including FinTech and BigTech firms. As of today, not merely the quantitative hedge funds are using algorithms, computational power and alternative data sources in finance. Instead, digital transformation has now impacted every aspect of finance, almost everywhere in the world.

As a result of digitization and datafication, almost every aspect of finance provides a potential area for AI.

Due to ever-improving performance in data gathering, processing, and analytics, AI can be expected to increasingly affect all operational and internal control matters of financial intermediaries, from strategy setting,[13] to compliance,[14] to risk management and beyond.[15]

C. Finance Use Cases

Today, algorithms and AI in financial services are frequently recognized as being used on the front- or back-end of an increasing range of processes and functions in finance.[16]

These include:

(1) customer related processes

[13] *See* John Armour & Horst Eidenmüller, "Self-driving corporations?" European Corporate Governance Institute-Law Working Paper No. 475/2019, https://ssrn.com/abstract=3442447, at 15 (while "strategic questions considered at the C-suite level" are unlikely to justify machine learning analysis, given the insufficiency of available data, "external generic data can be used to assist in scenario planning.").

[14] *See* Kenneth A. Bamberger, *Technologies of Compliance: Risk and Regulation in a Digital Age*, 88 TEX. L. REV. 669, at 690-93, 701-02 (2010).

[15] *See* Saqib Aziz & Michael M. Dowling, *Machine Learning and AI for Risk Management*, *in* DISRUPTING FINANCE. PALGRAVE STUDIES IN DIGITAL BUSINESS & ENABLING TECHNOLOGIES 33 (Theo Lynn et al. eds., 2019).

[16] *See* e.g., Hong Kong Monetary Authority & PwC, *Reshaping Banking with Artificial Intelligence* (November 2019) <https://www.hkma.gov.hk/media/eng/doc/key-functions/finanical-infrastructure/Whitepaper_on_AI.pdf> ; Bank of England and Financial Services Authority, *Machine Learning in UK financial services* (October 2019) <https://www.fca.org.uk/publication/research/research-note-on-machine-learning-in-uk-financial-services.pdf>

- on-boarding customers – particulary retail – more quickly and with a better user experience through biometrics such as facial recognition[17]
- marketing of financial services to specific user groups[18]
- enhancing customer relationship management, e.g. by (1) delivering instant responses to credit applications, (2) offering faster and better affordability checks for mortgages, and (3) delivering client-specific services with enhanced information and data-driven analyses[19]

(2) operations and risk management

- supporting or applying statistical models, e.g. for the calculation of pay-outs[20]
- managing risk, in particular setting risk limits and conducting stress testing[21] and credit scoring[22]
- determining executive compensation[23]
- monitoring boards of director decision-making biases[24]

(3) trading and portfolio management:

- capital allocation[25]

[17] This is a core aspect of RegTech.

[18] *See* Dirk A. Zetzsche, Ross P. Buckley, Douglas W. Arner & Janos N. Barberis, *From FinTech to TechFin: The Regulatory Challenges of Data-Driven Finance*, 14 N.Y.U. J. L. & BUS. 393, 425-430.

[19] *See* AI Accenture, supra n 4, p. 13, 15, 17 (providing the example of an AI steering the SME client to the best qualified relationship manager for the SME's needs, based on an analysis of the SME's cash-flow and risk figures, and informing the relationship manager on the needs and background of the SME, ensuring un-interrupted services and advice).

[20] *See* Buchanan, supra n 1, at p. 2 (stating that Fukoku Mutual Life Insurance uses IBM's Watson Explorer AI to calculate pay-outs).

[21] *See* Financial Stability Board, "Artificial intelligence and machine learning in financial services" (Nov. 2017) 16 <https://www.fsb.org/wp-content/uploads/P011117.pdf > (summarizing Ai-based risk management and stress testing, and stating that one global corporate and investment bank is using unsupervised learning algorithms in model validation).

[22] *See* Oliver Wyman & China Securities Credit Investment Company, *China Credit-tech Market Report: Technology-Driven Value Generation in Credit-Tech,* 2019 <https://www.oliverwyman.com/content/dam/oliver-wyman/v2/publications/2019/apr/china-credit-tech-market-report-4.pdf>

[23] U.S.-based Equilar Inc. uses available compensation disclosures, performance targets and performance data, to generate "pay-for-performance" scores that can be used to determine whether an executive is over- or under-paid relative to executives of similarly situated companies. *See e.g.* Equilar's patent application for its "Equilar Pay for Performance Score", U.S. Patent Office, Patent Application Publication, Pub. No. US 2013/0159067 A1, Pub. Date: 20 Jun. 2013 (detailing the algorithms and data sources used for calculating the score).

[24] Venture capital firm Deep Knowledge Ventures assigned a (sort of) board position to an AI dubbed VITAL. VITAL scans prospective companies' financing, performance, IP and previous funding rounds. Its task is to identify overhyped projects. *See* Press Release, Deep Knowledge Venture's Appoints Intelligent Investment Analysis Software VITAL as Board Member – Hong Kong Venture Capital Fund Appoints Machine Intelligence as Board Member, 13 May 2014, *available at* https://globenewswire.com/news-release/2014/05/13/635881/10081467/en/Deep-Knowledge-Venture-s-Appoints-Intelligent-Investment-Analysis-Software-VITAL-as-Board-Member.html.. For a scholarly discussion of VITAL, *see* Luca Enriques & Dirk Zetzsche, Corporate Technologies and the Tech Nirvana Fallacy, ECGI Law Working Paper 457/2019; Michal S. Gal, *Algorithmic Challenges to Autonomous Choice*, 25 MICH. TECH. L. REV. 59, 61 (2018); Mark Fenwick, Wulf A. Kaal & Erik P.M. Vermeulen, *The "Unmediated" and "Tech-Driven" Corporate Governance of Today's Winning Companies* 42 n114, TILEC Discussion Paper No. 2017-009 (2017).

[25] *See* FSB, supra n 21, at 15 (summarizing the efforts to employ AI for optimizing risk-weighted assets

- financial services robo-advice[26]
- algorithmic trading[27]

(4) payments and infrastructure

- replacing human agents with chatbots in client communication[28]
- combatting fraud[29]

(5) data security and monetization

- document data extraction, for strategic or risk management purposes[30]
- automated threat prevention, detection and response, in particular through cybersecurity solutions[31]

(6) regulatory and monetary oversight and compliance

- transaction monitoring[32]
- detecting and reporting compliance breaches, for instance with regard to insider trading and market abuse[33]
- AML and know-your-customer checks (KYC)[34]
- Macroeconomic adjustments and fine-tuning[35]

(RWA) and margin valuation adjustment (MVA)).

[26] *See* Kokfai Phoon & Francis Koh, Robo-Advisors and Wealth Management, The Journal of Alternative Investments Winter 2018, 20 (3) 79-94; Jill E. Fisch, Marion Labouré, John A. Turner, The Emergence of the Robo-advisor, PRC Policy Paper; Tom Baker & Benedict G.C. Dellaert, Regulating Robo Advice Across the Financial Services Industry, 103 Iowa L. Rev. 713 (2018).

[27] *See* Andrei A. Kirilenko & Andrew Lo, Moore's Law versus Murphy's Law: Algorithmic Trading and Its Discontents, 27:2 Journal of Economic Perspectives 51-72 (2013); for an overview of the EU Framework in Art. 17 MiFID II see Tilen ČUK & Arnaud Van Waeyenberge, "European Legal Framework for Algorithmic and High Frequency Trading (Mifid 2 and MAR): A Global Approach to Managing the Risks of the Modern Trading Paradigm", 9:1 Eur. J. of Risk Regulation 136-153 (2018).

[28] *See* Amy Zirkle, The Critical Role of Artificial Intelligence in Payments Tech, 27 May 2019, < https://www.fintechnews.org/the-crirital-role-of-artificial-inteliigence-in-payments-tech/>.

[29] *See* blog Bizety.com, 'PayPal Deep Learning Methods Against Fraud', 18 Oct. 2016 (describing Paypal's deep learning algorithms that analyze thousands of data points (e.g. IP address, buying history etc.) in real time in order to identity theft, phishing attacks etc., and arguing that Paypal's fraud rate with 0.32% is one of the lowest in financial services, compared 1.32% as financial industry standard, citing the Lexis Nexis True Costs of Fraud Study 2016).

[30] *See* Buchanan, supra n 1, at p. 2 (stating that UK PropTech2 start-up Leverton applies AI to automatically identify, extract and manage data from corporate documents such as rental leases); FSB, supra n 21, at 21 (summarizing the efforts to employ AI for macroeconomic surveillance an data quality assurance).

[31] *See* https://www.infosecurity-magazine.com/next-gen-infosec/ai-future-cybersecurity/.

[32] *See* supra n. 29

[33] *See* FSB, supra n 21, at 19 (summarizing the efforts to employ AI for compliance and RegTech, and stating that one global corporate and investment bank is using unsupervised learning algorithms in model validation).

[34] *See* FSB, supra n 21, at 20 (stating that AI supports KYC checks primarily in two ways: "(1) evaluating whether images in identifying documents match one another, and (2) calculating risk scores on which firms determine which individuals or applications need to receive additional scrutiny. Machine learning-based risk scores are also used in ongoing periodic checks based on public and other data sources, such as police registers of offenders and social media services. Use of these sources may enable risk and trust to be assessed quickly and often cheaply. Firms can use risk scores on the probability of customers raising "red flags" on KYC checks.").

[35] *See* Okiriza Wibisono, Hidayah Dhini Ari, Anggraini Widjanarti, Alvin Andhika Zulen & Bruno

The adoption rate of AI and autonomous systems in finance is increasing rapidly. At the same time, the pain from skyrocketing compliance costs and sanctioning has induced financial institutions – from FinTech startups to global systemically important banks – to focus on back-office AI-solutions, in the form of RegTech. RegTech solutions include Amazon Alexa-like voice bots used by Credit Suisse for compliance queries[36], and bots at JP Morgan to review commercial loan contracts equivalent to 360,000 hours of work each year by lawyers and loan officers.[37] AI is also being applied to equities trade execution for maximum speed at best price at JP Morgan[38] and post-trade allocation requests at UBS[39], and to calculate policy payouts at Japan's Fukoku Mutual Life Insurance.[40] AI is also behind the trend to seek alternative data for investment decisions,[41] prompting the mantra "all data is credit data".[42]

D. A New Focus for Financial Regulators

In recent years regulators and policymakers have begun to consider the use of AI in finance.[43]

For instance, a World Economic Forum (WEF) report from August 2018[44] highlighted that the use of AI-enabled systems by financial institutions is promoting "new efficiencies" and delivering "new kinds of value". However, a tight focus on these new capabilities risked overlooking how financial services are shifting fundamentally, as financial institutions become "more specialized, leaner, highly networked and dependent on the capabilities of a variety of technology players."[45] Financial institutions need to develop new approaches to how they deal with their people, processes and data.[46] In this regard, the WEF suggests that collaboration amongst multiple stakeholders will be required to counter the potential social and economic risks

Tissot, The use of big data analytics and artificial intelligence in central banking, IFC Bulletin May 2019, <https://www.bis.org/ifc/publ/ifcb50.pdf>.

[36] "Credit Suisse has deployed 20 robots within bank, markets CEO says" (Reuters, 2 May 2017): <https://www.reuters.com/article/us-milken-conference-creditsuisse/credit-suisse-has-deployed-20-robots-within-bank-markets-ceo-says-idUSKBN17X2JC>.

[37] "JPMorgan Software Does in Seconds What Took Lawyers 360,000 Hours" (Bloomberg, 28 Feb. 2017): <https://www.bloomberg.com/news/articles/2017-02-28/jpmorgan-marshals-an-army-of-developers-to-automate-high-finance>.

[38]"JPMorgan develops robot to execute trades" (Financial Times, 31 Jul. 2017): https://www.ft.com/content/16b8ffb6-7161-11e7-aca6c6bd07df1a3c?mhq5j=e6

[39]"Robots enter investment banks' trading floors" (Financial Times, 6 Jul. 2017): <https://www.ft.com/content/da7e3ec2-6246-11e7-88140ac7eb84e5f1?mhq5j=e6>

[40] "This Japanese Company Is Replacing Its Staff With Artificial Intelligence" (Fortune, 6 Jan. 2017): <http://fortune.com/2017/01/06/japan-artificial-intelligenceinsurance-company/>

[41]"AI and Alternative Data: A Burgeoning Arms Race" (20 Jun. 2017): <https://www.waterstechnology.com/trading-technologies-andstrategies/3389631/ai-and-alternative-data-a-burgeoning-armsrace>

[42] *See* M. Hurley and J. Adebayo, "Credit Scoring In the Era of Big Data" 18:1 Yale Journal of Law and Technology: <http://digitalcommons.law.yale.edu/cgi/viewcontent.cgi?Art.=1122&contex t=yjolt >

[43] We discuss a range of others in the following sections.

[44] World Economic Forum, "The new physics of financial services: How artificial intelligence is transforming the financial ecosystem" (15 Aug. 2018) < https://www.weforum.org/reports/the-new-physics-of-financial-services-how-artificial-intelligence-is-transforming-the-financial-ecosystem>.

[45] WEF, supra n. 44, at 19.

[46] AI Accenture, supra n. 4, at 5-7.

accompanied by the use of AI-enabled systems in finance.[47] Similarly, in October 2019, the WEF addressed how the financial services industry can responsibly use AI, focusing on understanding the governance requirements and risks of using AI in financial services. In particular, AI explainability, systemic risk and AI, bias and fairness, the algorithmic fiduciary, and algorithmic collusion are considered as prominent sources of uncertainties and risks associated with the use of AI in financial services. In the main, the WEF was of the view that in developing AI, the strategy taken should focus on a willingness to consider new governance and regulatory approaches that take into account the complex nature of AI-enabled systems, rather than developing "new ethics" for the financial services industry.

Among regulators, the European Central Bank has focused on the matter since at least 2017[48] and announced in February 2019 that algorithmic trading, an early and leading use case of AI, "has been growing steadily since the early 2000s and, in some markets, is already used for around 70% of total orders."[49]

In October 2019, the Bank of England and UK Financial Conduct Authority (FCA) released a major survey looking at maching learning (ML) in the UK financial industry.[50] Based on responses from 106 regulated financial institutions, the key findings included:

- ML is increasingly being used in UK financial services, with two thirds of respondents reporting they already use it in some form.
- Deployment is most advanced in the banking and insurance sectors.
- ML is now used across a range of business areas from front-office to back-office, and is used most commonly in AML and fraud detection as well as in customer services and marketing, with some firms also using it in areas such as credit risk management, trade pricing and execution, and general insurance pricing and underwriting.
- Regulation is not seen as a major barrier – rather, the biggest constraints are legacy IT systems and data limitations.
- Firms thought that ML does not necessarily create new risks, but could be an amplifier of existing ones. Such risks, for instance ML applications not working as intended, may occur if model validation and governance frameworks do not keep pace with technological developments.
- Firms validate ML applications before and after deployment. The most common validation methods are outcome-focused monitoring and testing against benchmarks.
- Firms use a variety of safeguards to manage the risks associated with ML. The most common safeguards are alert systems and so-called "human-in-the-loop" mechanisms. These can be useful for flagging if the model does not work as

[47] WEF, supra n. 44. *See also* UK Finance, supra n. 2.

[48] We discuss the Joint Report of the European Supervisory Authorities on the use of Big Data in Financial Services from March 2018 *infra*, at II.C.

[49] European Central Bank, "Algorithmic trading: trends and existing regulation", Newsletter 13 Feb. 2019, https://www.bankingsupervision.europa.eu/press/publications/newsletter/2019/html/ssm.nl190213_5.en.html >.

[50] Bank of England & Financial Conduct Authority, Machine Learning in UK Financial Services (Oct. 2019): https://www.fca.org.uk/publication/research/research-note-on-machine-learning-in-uk-financial-services.pdf.

intended (e.g. in the case of model drift, which can occur as ML applications are continuously updated and make decisions that are outside their original parameters).

- Firms mostly design and develop ML applications in-house. However, they sometimes rely on third-party providers for the underlying platforms and infrastructure, such as cloud computing.
- The majority of users apply their existing model risk management framework to ML applications and many highlight that these frameworks might have to evolve in line with increasing maturity and sophistication of ML techniques.

A 2019 survey by the Hong Kong Monetary Authority[51] and accounting firm PWC among the HK banking industry highlighted that:

- 89% of respondents (authorised banks) had adopted or planned to adopt AI applications.
- 92% of respondents planned to significantly expand their AI workforce in the next 5 years.
- Total capital investment in the area will rise by 70% in the next 5 years.
- The top 5 use cases include cybersecurity applications, client-facing chatbots, remote onboarding, biometric customer identification and personalised advertisements.
- 95% of the banks tend to partner with external technology firms for AI implementation, while 82% managed the research and development stage internally.
- The top three reasons for utilizing AI included improving customer experience, enhancing risk management and reducing costs.
- The major impediments for AI use in finance, include: lack of employees with AI expertise (70%), insufficient data (52%), design ethics of AI (48%), data privacy and security (44%) and legal and compliance challenges (44%).
- The top three risks identified were: (1) lack of expertise among employees, (2) biased decisions made by the AI models, and (3) lack of quality data.

Clearly, AI is playing an increasingly significant role in finance, a role which is set to increase. Looking forward, does this raise potential financial regulatory concerns?

II. The Risks of AI in Finance

AI raises many questions that are yet to be answered. General concerns without a particular financial services dimension, that could yet impact financial services, include for instance: (1) what happens to workers whose jobs are replaced by AI?, (2) how do we distribute the wealth created by machines in our societies and across borders?, (3) how does humankind maintain control of super-human AI systems?, and (4) which rights do we assign to robots, i.e. are we willing to grant robots human-like rights?[52] These macro issues with AI have a very important role in the financial sector and potentially in regulation, as we consider how we wish finance, the economy and our

[51] See Hong Kong Monetary Authority & PWC, Artificial Intelligence (AI) in Retail Banking (November 2019). < https://www.hkma.gov.hk/media/eng/doc/key-functions/finanical-infrastructure/Artificial_Intelligence_(AI)_in_Retail_Banking.pdf>.

[52] *See* Mirjana Stankovich et al, *Exploring Legal, Ethical and Policy Implications of Artificial Intelligence* (Sep. 2017).

societies to evolve as a result of the Fourth Industrial Revolution. These issues are central to discussions about AI and AI governance, within which finance plays an important role.

Our focus here however is on the more "micro" issues arising in the context of AI in finance. In particular, there is increasing awareness and analysis of the issues of fairness (including algorithmic bias), accountability and transparency (including "explainability") (sometimes summarized as "FAT") that arise with the implementation and evolution of AI.[53] These sorts of risks arise in particular as a result of "black box" issues: the view that AI develops independently and therefore its results are impossible to understand or accurately predict, highlighting the challenges of removing humans from AI systems.

In this section, we focus specifically on issues on the context of AI in finance from the standpoint of core financial regulatory objectives.[54] Using this lens of financial regulatory objectives, we categorize the major forms of risk as: data risks, cybersecurity risks, financial stability risks, and ethical risks.

Similar to the framework presented here, in December 2018, ACPR (Autorité de Contrôle Prudentiel et de Résolution – the French prudential regulatory authority within the Banque de France)[55] identified four major categories of risks associated with AI in finance:

(1) data processing risks associated with artificial intelligence;
(2) artificial intelligence and cybersecurity risks;
(3) the risk of players' dependency and the change of power relationships in the market; and
(4) challenges to financial stability and sovereignty.

ACPR further lists the governance and "explainability" of the algorithms, and challenges related to possible market restructuring, as further risks for supervisors.

The following sections consider these four major finance-related risks (data, cyber, financial stability, ethical) in light of the objectives of financial regulation.

[53] *See* eg, Brian W Tang, "Forging a Responsibility and Liability Framework in the AI Era for Regtech" in Janos Barberis, Douglas W Arner and Ross P Buckley, *The REGTECH Book: The Financial Technology Handbook for Investors, Entrepreneurs and Visionaries in Regulation* (Wiley, 2019), p 235; Yi Zeng, Enmeng Lu and Cunqing Huangfu, "Linking Artificial Intelligence Principles": <arXiv:1812.04814v1>; Jessica Fjeld, Nele Achten, Hannah Hilligoss, Adam Nagy and Madhulika Srikumar, "Principled Artificial Intelligence: Mapping Consensus in Ethical and Rights-Based Approaches to Principles for AI" (Berkan Klein Center Research Publication No.2020-1, 15 Jan. 2020): <https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3518482## >

[54] The objectives of financial regulation can be summarized as: financial stability, financial efficiency, financial integrity, customer protection, economic development and financial inclusion. Financial stability can be seen both negatively (as avoidance of crises) and positively (as appropriate functioning of the financial system). Financial integrity focuses on prevention of criminal activities and use of the financial markets for activities like money laundering and terrorist financing. Customer protection focuses on systems to prevent abuses of consumers. Financial efficiency, economic development and financial inclusion focus on how to support the positive functioning and role of the financial system. See Douglas W. Arner, *Financial Stability, Economic Growth and the Role of Law* (Cambridge University Press, 2007).

[55] ACPR, "Artificial Intelligence: Challenges for the Financial Centre" (Dec. 2018): <https://acpr.banque-france.fr/sites/default/files/medias/documents/2018_12_20_intelligence_artificielle_en.pdf>

A. Data Risks

Central to the potential of AI is its potential to process far more data than humans, and without two human weaknesses. First, AI treats past data with the same precision as more recent data; in contrast, humans tend to treat more recent data as more significant and neglect older data in line with declining memory. Second, AI, if correctly programmed, subjects all data to the same objective treatment, while humans tend to discriminate among certain datapoints based on their experience, values and other non-rational judgement patterns. In this limited sense that technology does not follow its own agenda, and is not itself subject to humans' cognitive biases, technology can be said to be unbiased.[56] As we discuss below, however, the results can still be objectionable, resulting in bias in treatment.

AI use is subject to a number of risks and idiosyncratically suffers from a number of deficiencies.

1. Data dependency

AI is data dependent. The results of AI use are only as good as the data with which the AI has worked. Data dependency can give rise to a number of deficiencies.

First, even with a wide range of data generated in diverse ways, the data pool analysed **may lack the data relevant for the task**.[57] As a principle, past data may have some predictive capacity, in the sense that one event is more likely than another, but lack the ability to determine, strictly speaking, the path of future events in detail. Probability must not be confused with certainty. As the value of high-quality information and the threats posed by information gaps both continue to grow, regulators should focus on the development of widely used and well-designed data standards.[58]

Second, the **data quality may be poor**. An often-repeated example in the field of AI research includes the use of training data from the Enron case for compliance AI. From today's perspective, the Enron data are outdated. Even at the time, the Enron case was a deeply unfortunate outlier, rendering the use of the Enron dataset quite inappropriate.[59] From a legal perspective, protected factors come under threat if AI discriminates based on factors, proxies for these factors, or other factors altogether, that all describe little more than *a* part of social and financial relations within society. For instance an algorithm that determines creditworthiness based on consistency of phone use (rather than complete economic and financial data), may discriminate against members of certain religions who tend to use their phones far less on one day each week, such as a Friday or Saturday.[60]

Third, the **data used for AI analysis may suffer from biases**. This may be due either to data selection issues ("dashboard myopia") or data reflecting biases persisting in

[56] *See* Gramitto Ricci, "The technology and archaeology of corporate law", Cornell Law School research paper No. 18-40 (2018), at 37-38, http://ssrn.com/abstract=3232816; Martin Petrin, "Corporate Management in the Age of AI" (UCL Working Paper No.3/2019), at 34-35.

[57] Enriques & Zetzsche, supra n. 24, at 32.

[58] *See* Berner & Judge, "The Data Standardization Challenge", in Arner et al., Systemic Risk (2019), at 135-149.

[59] Enriques & Zetzsche, supra n. 24, at 31.

[60] *See* Zetzsche, Buckley, Arner & Barberis, *From FinTech to TechFin*, supra n 18.

society at large (e.g. that males are more likely to work in tech).[61] Decision-makers with prejudiced views may mask these by wittingly or unwittingly using biased data.[62] Biased data could likewise be selected in efforts to enhance an executive's personal bonus or to reduce oversight within an organization.[63]

2. Data availability

Data availability, even with a wide range of data generated in diverse ways, may be limited. The data may exist, but not be collected, structured or available for digital analysis. This may happen for two reasons. First, data collection is expensive. Small financial services providers may focus on the collection of data they believe valuable, giving life to their biases as to which data is relevant. Second, large financial services providers may be unwilling to share data they have with other firms, given that the other firms may either sell the data or become competitors of the data originator in the future[64] (the problem open banking is designed to address). The issue of data availability and accessibility then intersects with the vast world of data privacy and protection regulation.

3. AI Interdependency

A variant of the data availability issue is the lack of data on how other AI perform similar calculations at the same time, and how their decisions influence the tasks performed by the first AI. Such behaviour can result in "herding", in which actors make use of similar models to interpret signals from the market.[65] Algorithms trading in millisecond trading windows simultaneously in unexpected situations in which their operating assumptions do not apply have resulted in extreme volatility events, referred to as flash crashes.[66] This has resulted in regulation addressing algorithmic trading across the world.[67]

We can imagine similar problems with robo-advisors, in which one AI may front-run another AI advisor'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, given the speed of events and that these algorithms will, again, be based on (sometimes) inadequate assumptions. Notwithstanding the former, the underlying issue remains that the original performance of calculations may turn out to be futile, or

[61] *See* Lin, supra n. 9, at 536-537.

[62] Solon Barocas & Andrew D Selbst, "Big Data's Disparate Impact" 104 CAL. L. REV. 671 (2016), at 692

[63] Enriques & Zetzsche, supra n. 24, at 30.

[64] Enriques & Zetzsche, supra n. 24, at 30.

[65] World Economic Forum, "Navigating uncharted waters: A roadmap to responsible innovation with AI in financial services" 62 (Oct. 23, 2019) < https://www.weforum.org/reports/navigating-uncharted-waters-a-roadmap-to-responsible-innovation-with-ai-in-financial-services>.

[66] *See*, Buchanan, supra n. 1, at 6. *See*, generally, on flash crashs Andrei A. Kirilenko, Albert S. Kyle, Mehrdad Samadi & Tugkan Tuzun, "The Flash Crash: High-Frequency Trading in an Electronic Market", 72:3 Journal of Finance 967 (2017).

[67] See references supra n. 27.

very harmful, whenever various algorithms perform and execute similar tasks simultaneously.

The alternative to uncoordinated behaviour, however, is more frightening: tacit collusion. If several self-learning algorithms find out that cooperation in capital markets is more profitable than competition, they could cooperate, i.e. manipulate information and prices to their own advantage. There is evidence for self-learning AI colluding in price setting,[68] and generally little reason to believe that multiple AI colluding in financial markets pricing is unlikely. The WEF has suggested financial institutions may potentially mitigate the risks of tacit collusion by (i) restricting their AI-enabled systems to communicate only with their own environments for "explicitly justifiable business purposes"; (ii) ensuring their AI-enabled systems' decisions are explainable by "valid, legal business reasons"; and (iii) requiring humans to oversee decisions made by AI-enabled systems.[69] These are all good suggestions, but may not always be sufficient to fully mitigate this substantial risk of AI interdependency, in particular if collaboration is profitable to the firm. Accordingly, it is not surprising that competition authorities in Europe and elsewhere are increasingly focussed on this issue of algorithms and collusion.[70]

B. Financial stability risks

The Financial Stability Board in 2017[71] analysed and summarized the possible financial stability implications of AI and ML as including the following:

- customer-focused uses – credit scoring, insurance and client-facing chatbots
- operations-focused uses – capital optimization, model risk management and market risk management
- trading and portfolio management – in trading execution and the scope of portfolio management
- regulatory compliance and supervision – applications by financial institutions for regulatory compliance (RegTech), uses for macroprudential surveillance and data quality assurance, uses and potential uses by central banks and prudential authorities (SupTech), and uses by market regulators for surveillance and fraud detection
- micro-financial analysis, including possible effects on financial markets, financial institutions, consumers and investors
- macro-financial analysis – market concentration and systemic risk importance of institutions, potential market vulnerabilities, networks and interconnectedness, and other implications.

[68] Ariel Ezrachi & Maurice E. Stucke, "Artificial intelligence & collusion: When computers inhibit competition" (2017) Univ. Ill. L. Rev. 1775.

[69] World Economic Forum, "Navigating uncharted waters", supra n. 69.

[70] *See* e.g., Bundeskartellamt and Autorite de la concurrence, *Algorithms and Competition* (November 2019) <https://www.bundeskartellamt.de/SharedDocs/Publikation/EN/Berichte/Algorithms_and_Competition_Working-Paper.pdf?__blob=publicationFile&v=5 >; UK Competition and Markets Authority, *Pricing algorihms: Economic working paper on the use of algorithms to facilitate colusion and personalised pricing* (8 Oct. 2018) <https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/746353/Algorithms_econ_report.pdf>

[71] *See* FSB, supra n. 21.

The FSB concluded that “AI and machine learning applications show substantial promise if their specific risks are properly managed”. In terms of financial stability, the FSB stressed that “network effects and scalability of new technologies may in the future give rise to additional third-party dependencies” and this “could in turn lead to the emergence of new systemically important players,” [72] up to the level of oligopoly or monopoly. Even more, some of these new market participants are currently unregulated and unsupervised. These third-party dependencies and interconnections could have systemic effects.[73] Further, the lack of interpretability or “auditability” of AI and ML methods has the potential to contribute to macroeconomic risk unless supervisors find way to supervise the AI. This is particularly challenging, given that “many of the models that result from the use of AI or machine learning techniques are difficult or impossible to interpret”[74] and AI-related expertise beyond those developing the AI is limited, in both the private sectors and among regulators.[75]

C. Cybersecurity

AI could be used to attack, manipulate, or otherwise harm an economy and threaten national security through its financial system directly and/or its impact on the wider economy.[76] For instance, algorithms could be manipulated in an effort to transfer wealth to foreign powers, to undermine an economy’s growth in an effort to create unrest, or to send wrong signals to trading units to seek to trigger a systemic crisis.[77] The cybersecurity dimension is all the more serious given that many financial services firms rely on a small group of technology providers, that give rise, by themselves, to a new form of risk we have termed Tech Risk.[78] This is amplified by the fact that many AI-enabled systems have not been tested in financial crisis scenarios.[79]

The most important way to address cybersecurity is to (1) invest in cybersecurity resources, including in-house expertise and training of employees, and (2) have protocols in place to cooperate swiftly with other financial intermediaries in a similar situation, to ensure fast detection of, and responses to, these attacks, with or without involvement of regulators.[80]

D. Ethics and Financial Services

Ethics in finance are a crucial concern.[81] Ethical issues came to the fore in the wake of the Global Financial Crisis and have received continued attention as a result of

[72] *See* FSB, supra n. 21, at 33-34.

[73] For details *see* Lin, supra n. 9, at 544.

[74] *See* FSB, supra n. 21, at 33-34.

[75] *See* FSB, supra n. 21, at 33-34.

[76] For further examples *see* Lin, supra n. 9, at 538-539.

[77] *See* Ross P. Buckley, Douglas W. Arner, Dirk Zetzsche & Eriks Selga, *The Dark Side of Digital Financial Transformation: The New Risks of FinTech and the Rise of TechRisk*, __ SING. J. LEG. ST. __ (2020), *in press*.

[78] *See* Douglas W. Arner, Ross P. Buckley, and Dirk 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 Tears after the great crash* (McGill-Queen's UP 2019), at 69.

[79] *See* Buchanan, supra n. 1.

[80] *See TechRisk*, supra n. 77.

[81] See earlier focus on this after the global financial crisis, eg, Brian Tang, “Promoting Capital Markets

subsequent scandals including those relating to LIBOR, foreign exchange and most recently the entire Australian financial system. A number of ethical questions with a particular financial services dimension will, most likely, be addressed by future (financial) legislation so as to make AI-driven financial services stable and sound, and their risks balanced. These tend to fall into four areas: AI as non-ethical actor, AI's influence on humans, artificial stupidity and artificial maleficence, and more general ethical considerations.

1. AI as nonethical actor

Algorithms do not "feel" or have "values". Training machines in values seems difficult, since we humans often lack insights into the human psyche: ie, humans often cannot tell why they feel as they do in certain ways.[82] While some ethical concerns, such as the ban of interest under Shariah law, can possibly be codified in ways that could be adopted by algorithms, most human feelings are more subtle, and subject to change under specific circumstances, reflecting the human abilities to learn and adapt.

AI's lack of ethical foundation could create serious harm for the portfolio value of a given financial intermediary if the AI misprices reputational risk. For instance, Microsoft's AI bot, Tay, "originally designed to interact with people online through casual and playful conversation, ended up hoovering good, bad, and ugly interactions. Within 16 hours of launch, Tay turned into a brazen anti-Semite, stating, 'Hitler was right'."[83] If a launched product came to this conclusion, we would expect serious stock price reactions. Unforeseen reputational risk can also prompt sudden and deeply unhelpful rule changes with major financial consequences. A vivid example is the near-prohibition of certain diesel cars in the EU following the diesel scandals in the US, contrary to the evidence that diesel's carbon emissions are lower than those of cars using petrol, and that its other pollution effects can be reduced even further by employing certain filters.[84] Volkswagen's severe ethical shortcomings in this case were all too human, but software controlling engine performance in test situations could foreseeably be programmed by AI at some point in the future.[85]

Professionalism: An Emerging Asian Model" , in Ross P Buckley, Emilios Avgouleas and Douglas W Arner, *Reconceptualising Global Finance and Its Regulation* (Cambridge University Press, 2016), at 357

[82] For details *see* Enriques & Zetzsche, "Corporate Technologies", supra n. 24, at 34.

[83] *See* Elle Hunt, "Tay, Microsoft's AI chatbot, gets a crash course in racism from Twitter" The Guardian (24 Mar. 2016) <https://www.theguardian.com/technology/2016/mar/24/tay-microsofts-ai-chatbot-gets-a-crash-course-in-racism-from-twitter>; Dawson D & Schleiger E, Horton J, McLaughlin J, Robinson C∞, Quezada G, Scowcroft J, and Hajkowicz, (2019) Artificial Intelligence: Australia's Ethics Framework – Discussion Paper. Data61 CSIRO, Australia. < https://data61.csiro.au/en/Our-Research/Our-Work/AI-Framework>, at 31-32.

[84] European Environment Agency, "Explaining Road Transport Emissions: A Non-Technical Guide" (Jan. 2016) <https://www.eea.europa.eu/publications/explaining-road-transport-emissions>, p. 12; European Court of Auditors, "The EU's Response to the 'Dieselgate' Scandal", Briefing Paper (Feb. 2019) <https://www.eca.europa.eu/lists/ecadocuments/brp_vehicle_emissions/brp_vehicle_emissions_en.pdf >, [7] – [9].

[85] Capgemini Research Institute, "Accelerating Automotive's AI Transformation: How Driving AI Enterprise-wide Can Turbo-charge Organizational Value" 17-8 (Mar. 2019) <https://www.capgemini.com/wp-content/uploads/2019/03/Ai-in-automotive-research-report.pdf>.

The apparent risk is intensified by access to vast data accumulated on clients. The more data AI has about a certain person, the greater the risk of the AI nudging the human into certain behaviour, such as the purchase of an unsuitable financial product.

While some such unethical conduct could be mitigated through 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.[86] AI needs human guidance for ethical decision-making: humans-in-the-loop are a necessity.

2. AI's influence on humans

Human-AI interaction will require particular analysis in financial services. If, for instance, humans respond differently to AI information requests than they would to human requests, paradigms on which financial services legislation is based may need rethinking. This could pertain, for instance, to (1) product governance and target market concepts, (2) mandatory disclosure, (3) mandatory client / consumer protection rules, and (4) choice of law and courts.

AI can enhance or diminish human capacity. One obvious field in which AI can enhance human capacity is knowledge and education. AI as "augmented intelligence" could turn an uneducated, unskilled human into a skilled investor, by way of recommendations or substitution for human decision-making. The same is true for human decision-making errors revealed in behavioural finance literature: AI could be programmed to address biases that reflect the human tendency to rely on patterns rather than thinking, given that the hard task of thinking could be outsourced to the AI. For instance, AI could adjust for the human bias to stick to investments made rather than opt for reconsiderations based on data.

On the other hand, AI could decrease human capacity. For instance, to the extent that the human need to develop advanced math and other sophisticated data analytical capacities is lessened with appropriate programmes being widely available, we would expect humans to develop lesser data analytic capacities in time. This is supported by the WEF which suggests 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 to crises appropriately.[87] Our generally increasing lack of ability to remember telephone numbers or recall directions are vivid demonstrations of the effects of our increasing dependency on mobile phones today.

Both effects could be exploited in the financial services context. Coaching AI could be used to enhance financial and tech literacy of staff and investors, resulting in better resource allocation. Exploitative AI could ask clients to invest in overpriced, less valuable financial products that benefit only the product originator.

Obviously, the former can happen in a transparent or non-transparent, nudging manner. Research as to how humans respond to computer-generated incentives is ongoing and hints at serious risks for humans. Humans respond to certain communications with an enhanced degree of trust. AI can invest in relationships using an almost unlimited amount of resources, potentially generating a very high degree of trust. This illustrates

[86] *See further* Lin, supra n 9, at 537-538.

[87] World Economic Forum, "Navigating uncharted waters", supra n 69.

the level of responsibility AI developers bear, and the absolute necessity for ethical restrictions by way of rules and internal controls.

3. Artificial stupidity and artificial maleficence

How we can protect ourselves against AI mistakes and unethical behaviour is a major question. 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. A common example given in AI literature refers to the task of eradicating cancer, for which a machine could propose the eradication of humankind. While human-controlled machines hopefully will not do this, in time, can we be so confident about super-human machines? We draw similar examples from financial services. For instance, where certain conduct results in liability and consumers sue far more than institutional clients, a computer could decide to avoid consumer relationships, thereby financially excluding consumers and depriving them of the opportunity to use the financial system to hedge against the risks of mankind, ranging from poor health to unemployment and old age.

E. Risk typology: Framework of analysis

The risks of AI for finance outlined in this section fall into three major categories. (1) information asymmetry, (2) data dependency and (3) interdependency.

First, as to information asymmetry, AI enhances information asymmetry about the true functions and limits of certain algorithms as third party vendor or in-house AI developers typically understand the algorithms far better than the financial institutions that acquire and use them (including the institutions' governance mechanisms) and the supervisors that supervise the institutions. This is to some extent the result of the innovation of new technology, but also egotistic and commercial considerations and other current "black box" technologies often mitigate against developers making the algorithms as transparent or as explainable as possible.

Second, AI enhances data dependency as data sources are critical for it to operate and AI assessed one day may change its operations, effects and potentially discriminating impact on a later day when using a different data pool.

Third, AI enhances interdependency in the sense that AI can interact with other AI with unexpected consequences, enhancing or diminishing its operations in finance.[88]

The law is likely to address the risks of AI generating undesirable results by preventive regulation or corrective liability allocation. Suffice to say that drafting these rules and enforcing them in light of the incredibly rapid developments in AI is a serious challenge. Leaving aside the much discussed private law dimension and liability allocation,[89] we will focus in the following Part on regulatory tools.

[88] *See* Lin, supra n. 9, at 542.

[89] *See* on AI-related liability from a U.S. perspective Mark A. Lemley & Bryan Casey, *Remedies for Robots*, 86 U. CHI. L. REV. 1313 (2019) (suggesting to focus on on no-fault liability systems, or at least ones that define fault differently, to compensate plaintiffs for AI-inflicted harm); Ryan Calo, *Open Robotics*, 70 MD. L. REV. 571, 601–11 (2011) (proposing liability for open-source robots aiming at balancing the goals of fostering innovation and incentivizing safety); Rebecca Crootof, *War Torts: Accountability for Autonomous Weapons*, 164 U. PA. L. REV. 1347, 1389–1402 (2016) (discussing

III. Regulating AI in Finance: Challenges for External and Internal Governance

Markets and regulators have a number of means to address risks relating to financial services, ranging from private ordering and self-regulation to soft law approaches including recommendations to top-down command-and-control regulation. Financial supervision will be challenged by AI, requiring careful consideration of approaches which can best balance benefits and risks.

We begin with a discussion of the wide range of ethical frameworks which are being developed around the world to address the challenges of AI. Many of these however do not cater for the specific context of finance. We thus focus on approaches which are focusing specifically on AI in finance.

A. Ethical frameworks for AI

General frameworks addressing the question of the extent to which humans should be responsible when developing and dealing with AI are under development worldwide.[90] These clearly have direct relevance in the context of finance and financial regulation.

1. General frameworks

The UK House of Lords AI Select Committee defined five general principles of AI development and treatment in December 2017.[91] In April 2019, the European

robotic weapons systems and their potential legal liability); Karni A. Chagal-Feferkorn, *Am I an Algorithm or a Product? When Products Liability Should Apply to Algorithmic Decision-Makers*, 30 STAN. L. & POL'Y REV. 61 (2019); Yavar Bathaee, *The Artificial Intelligence Black Box and the Failure of Intent and Causation*, 31 HARV. J. L. & TECH. 889, 931–32 (2018); from an European angle EUR. PARL., EUR. PARL. RES. SERV., PANEL FOR THE FUTURE OF SC. & TECHN., A GOVERNANCE FRAMEWORK FOR ALGORITHMIC ACCOUNTABILITY AND TRANSPARENCY 52, 72-74 (Apr. 2019), https://www.europarl.europa.eu/RegData/etudes/STUD/2019/624262/EPRS_STU(2019)624262_EN.pdf (discussing no-fault/strict tort liability with varying degrees of liability depending on the transparency and criticality of the algorithmic systems and on AI certification by public authorities); as well as the contributions in LIABILITY FOR ARTIFICIAL INTELLIGENCE AND THE INTERNET OF THINGS (Sebastian Lohsse/ Reiner Schulze/ Dirk Staudenmayer, eds., 2019); Brian W Tang, "Forging a Responsibility and Liability Framework in the AI Era for Regtech" in Janos Barberis, Douglas W Arner and Ross P Buckley (ed), *The REGTECH Book: The Financial Technology Handbook for Investors, Entrepreneurs and Visionaries in Regulation* (Wiley, 2019), p 235.

Liability is also discussed in the context of liability for harm inflicted by autonomous vehicles, *see* Mark A. Geistfeld, *A Roadmap for Autonomous Vehicles: State Tort Liability, Automobile Insurance, and Federal Safety Regulation*, 105 CAL. L. REV. 1611 (2017); Kenneth S. Abraham & Robert L. Rabin, *Automated Vehicles and Manufacturer Responsibility for Accidents: A New Legal Regime for a New Era*, 105 VA. L. REV. 127 (2019); Bryant Walker Smith, *Automated Driving and Product Liability*, 2017 MICH. ST. L. REV. 1; A. Michael Froomkin & P. Zak Colangelo, Self-Defense against Robots and Drones, 48 CONN. L. REV. 1 (2015)

[90] *See* https://blog.einstein.ai/frameworks-tool-kits-principles-and-oaths-oh-my/. For the Australian framework *see* Dawson et al., Artificial Intelligence: Australia's Ethics Framework, supra n 83. See also IEEE Global Initiative on Ethics of Autonomous and Intelligence Systems, whose Ethically Aligned Design <https://standards.ieee.org/industry-connections/ec/ead-v1.html >

[91] *See* House of Lords, Select Committee on Artificial Intelligence, "Written evidence volume: AI in the UK: ready, willing and able?" (11 Dec. 2017): https://www.parliament.uk/documents/lords-committee/Artificial-intelligence/AI-Written-Evidence-Volume.pdf. The five principles include the commitment (1) to serve and benefit humanity, (2) intelligibility and fairness, (3) data privacy and an

Commission released *Ethics guidelines for trustworthy AI*, based around seven key requirements: human agency and oversight; technical robustness and safety; privacy and data governance; transparency; diversity, non-discrimination and fairness; societal and environmental well-being; and accountability.[92]

Most influentially so far, in May 2019, dozens of countries including the United States adopted the OECD AI Recommendation, the first intergovernmental standard for AI:[93]

> "The Recommendation identifies five complementary values-based principles for the responsible stewardship of trustworthy AI:
>
> - AI should benefit people and the planet by driving inclusive growth, sustainable development and well-being.
> - AI systems should be designed in a way that respects the rule of law, human rights, democratic values and diversity, and they should include appropriate safeguards – for example, enabling human intervention where necessary – to ensure a fair and just society.
> - There should be transparency and responsible disclosure around AI systems to ensure that people understand AI-based outcomes and can challenge them.
> - AI systems must function in a robust, secure and safe way throughout their life cycles and potential risks should be continually assessed and managed.
> - Organizations and individuals developing, deploying or operating AI systems should be held accountable for their proper functioning in line with the above principles."
>
> "Consistent with these value-based principles, the OECD also provides five recommendations to governments:
>
> - Facilitate public and private investment in research & development to spur innovation in trustworthy AI.
> - Foster accessible AI ecosystems with digital infrastructure and technologies and mechanisms to share data and knowledge.
> - Ensure a policy environment that will open the way to deployment of trustworthy AI systems.
> - Empower people with the skills for AI and support workers for a fair transition.
> - Co-operate across borders and sectors to progress on responsible stewardship of trustworthy AI."

Drawing on the OECD AI Recommendation, the G20 endorsed the G20 AI Principles in July 2019.[94] In September 2019, endorsing the OECD Recommendations the US Chamber of Commerce released Principles on Artificial Intelligence,[95] including a call for US businesses to abide by these standards.

adequate level of data protection and against data monopolization, (4) to allow all humans to be educated and flourish mentally, emotionally and economically alongside AI, and (5) to avoid any AI's programming aiming at the destruction or deception of human beings.

[92] Independent High-Level Expert Group on Artificial Intelligence Set Up by the European Commission, "Ethics Guidelines for Trustworthy AI" (Apr. 2019): https://ec.europa.eu/digital-single-market/en/news/ethics-guidlines-trustworthy-ai.

[93] OECD Council Recommendation on Artificial Intelligence**,** https://www.oecd.org/going-digital/ai/principles/

[94] https://www.g20-insights.org/wp-content/uploads/2019/07/G20-Japan-AI-Principles.pdf.

[95] *See* https://www.uschamber.com/sites/default/files/chamber_ai_principles_-_general.pdf.

In the meantime, there are many parallel AI ethics initiatives arising from the private sector and researchers, such as the IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems,[96] Future of Life Institute's Asilomar AI Principles[97], the Partnership on AI and the Montreal Declaration for responsible development of AI,[98] as well as a number of financial institutions.[99]

In China, the AI ethics initiatives have been more top-down, including the Beijing Academy of Artificial Intelligence's AI Principles in May 2019,[100] and the Ministry of Science and Technology National New Generation AI Governance Expert Committee's Governance Principles for a New Generation of AI in June 2019,[101] with increasing calls for cooperation over competition.[102]

2. Data protection and privacy

Data protection and privacy commissioners have increasingly viewed the governance of AI as within their purview. For instance, at the 40th International Conference of Data Protection and Privacy Commissioners in October 2018, the commissioners in their Declaration on Ethics and Data Protection in AI[103] endorsed six guiding principles as core values to preserve human rights in the development of AI:

(1) Fairness,
(2) Continued attention and 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 every individual, and
(6) Unlawful biases or discriminations arising from the use of data in artificial intelligence should be reduced and mitigated.

[96] See eg, IEEE Global Initiative on Ethics on Authomous and Intelligent Systems, *Ethically Aligned Design*: *A Vision for Prioritizing Human Well-being with Autonomous and Intelligent Systems*, Version II. (< https://standards.ieee.org/industry-connections/ec/autonomous-systems.html>

[97] Future of Life Institute, Asilomar AI Principles, 2017: < https://futureoflife.org/ai-principles/>

[98] Montreal Declaration for a responsible development of artificial intelligence (4 Dec. 2018) <https://www.montrealdeclaration-responsibleai.com/the-declaration>

[99] Institutions having adoped AI codes of conduct include, for instance, BNY Mellon, Deutsche Bank and Toronto Dominion.

[100] Beijing Academy of Artificial Intelligence (backed by the Chinese Ministry of Science and Technology and the Beijing municipality government) issued the Beijing AI Principles 28 May 2019: <https://www.baai.ac.cn/blog/beijing-ai-principles>

[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" (17 Jun. 2019) <http://most.gov.cn/kjbgz/201906/t20190617_147107.htm>: see China Daily English translation <https://www.chinadaily.com.cn/a/201906/17/WS5d07486ba3103dbf14328ab7.html>

[102] See e.g., New Economic Forum speech of China's former vice minister of foreign affairs Fu Ying, "Why the US should join China in Future-proofing AI Technologies", South China Morning Post, 5 Dec. 2019: <https://www.scmp.com/comment/opinion/article/3040435/why-us-should-join-china-future-proofing-ai-technology>

[103] "Declaration on Ethics and Data Protection in Artificial Intelligence", 40th International Conference of Data Protection and Privacy Commissioners in October 2018: <https://icdppc.org/wp-content/uploads/2018/10/20180922_ICDPPC-40th_AI-Declaration_ADOPTED.pdf>.

The Conference called for common governance principles on AI and a permanent working group on Ethics and Data Protection in AI to address the challenges of AI development.

Also relying on data protection principles, Article 22 of the European General Data Protection Regulation (GDPR) is seen as designed to require AI to perform ethically.[104] Entitled "Automated individual decision-making, including profiling", Article 22 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 or similarly significantly affects her or him. Caveats apply if the decision is necessary for the entering into, or performance of, a contract between the data subject and the data controller; is authorized by the EU or a member state to which the controller is subject and which provides for suitable measures to safeguard the data subject's rights and freedoms and legitimate interests; and is based on the data subject's explicit consent. Decisions should not be based on special categories of personal data unless suitable measures are applied to safeguard the data subject's rights and freedoms, and legitimate interests (see Article 9 GDPR). Where it is necessary for entering into or the performance of a contract, or where the data subject's consent is required, the data controller should institute suitable measures to safeguard the data subject's rights and freedoms, and legitimate interests. The data subject has the right to insist on human intervention on the part of the controller and to express his or her point of view to contest the decision.[105]

B. Financial Regulation and AI

Regulators globally have started to consider how AI impacts financial services and to issue regulatory guidance.

1. European Supervisory Authorities

In one of the first regulatory enquiries, in December 2016, the European Supervisory Authorities (ESAs) (European Banking Authority (EBA), European Securities and Markets Authority (ESMA) and European Insurance and Occupational Pensions Authority (EIOPA)) published a draft report on Big Data risks for the financial sector that included AI.[106] Of the 68 respondents, some stressed that "predictions based on Big Data can be flawed. It was also noted that [AI] could render the decision-making process less transparent and, in general, the intensity of the risks (…) could increase as a direct consequence of such new tools."[107] While most saw AI as an additional layer of Big Data analytics and a key tool to improve discovering patterns in data,

[104] *See* Jimmie Franklin, "GDPR has kept AI ethical, despite concerns" (IFLR, 2 Oct. 2019): https://www.iflr.com/Article/3896942/GDPR-has-kept-AI-ethical-despite-concerns.html.

[105] *See generally* Mirjana Stankovic el al, "Exploring Legal, Ethical and Policy Implications of Artificial Intelligence" Law, Justice and Development Draft White Paper (Oct. 2017).

[106] Joint Committee of the European Supervisory Authorities EBA, ESMA, EIOPA, Discussion Paper on the Use of Big Data by Financial Institutions, 19/12/2016, JC/2016/86.

[107] Joint Committee of the European Supervisory Authorities EBA, ESMA, EIOPA, Joint Committee Final Report on Big Data, JC/2018/0415 (Mar. 2018), <https://www.esma.europa.eu/sites/default/files/library/jc-2018-04_joint_committee_final_report_on_big_data.pdf>, at [50].

classification, evaluation and prediction, some stakeholders emphasized AI would add to the complexity, and incomprehensibility, of Big Data tools.[108]

The ESAs' final report in March 2018 found that, even when such techniques are used by financial institutions, in some respects "specific legislation in the field of data protection, cybersecurity and consumer protection is [best] positioned to address some of [AI] risks".[109] At the same time, the ESAs found that

> for the time being the current sectoral financial legislation sets requirements that are capable to address a number of risks specific to the use of Big Data techniques by financial institutions. Indeed a number of existing far reaching requirements, while not designed with the risks posed by the use of Big Data in mind, are applicable irrespective of the technological context.[110]

Given the ongoing implementation of legislation such as GDPR, the second Payment Services Directive (PSD2), the second Markets in Financial Instruments Directive (MiFID II) or the Insurance Distribution Directive, the ESAs refrained from recommending additional legislative steps, but focused on a data-oriented interpretation of existing sectoral legislation.

a. Organizational and prudential requirements

The ESA's interpretation focused, from organizational and prudential perspectives, on the following principles:

- Establishing and operating sound internal control mechanisms, effective procedures for risk assessment and effective control and safeguard arrangements for information processing systems.[111] The ESAs require financial institutions to allocate appropriate capital, human and IT resources to the implementation of Big Data from an operational standpoint.
- Ensuring continuity and regularity in the performance of their activities (and employing appropriate and proportionate systems, resources and procedures to this end).[112] The ESAs require that financial institutions address "the possible threats that may impact the continuity and the regularity of the performance of the financial institutions' activity."[113]
- Monitoring market activity and mitigating against counterparty or systemic risk or disorderly trading.[114] Investment firms and trading venues must ensure robust measures are in place to prevent algorithmic or high-frequency trading from disrupting the markets.
- Ensuring that reliance on a third party (i.e. outsourcing) does not impair the quality and the continuous performance of services.[115] The ESAs "stress that sectorial legislation requirements applicable to the outsourcing of important

[108] Joint Committee of the ESAs, supra n 107, at 98-99.

[109] Joint Committee of the ESAs, supra n 107, at p. 23.

[110] Joint Committee of the ESAs, supra n 107, at p. 23.

[111] Cf. Art. 16(5) MiFID II, Art. 18 Alternative Investment Fund Managers Directive (AIFMD), Art. 12 Undertakings for Collective Investment in Transferable Securities Directive (UCITS), Art. 5, 95 PSD2, Art.41, 44, 46 Solvency II.

[112] See Art. 16(4), 17 MiFID II, Art. 5, 95 PSD 2, Art. 41 Solvency II.

[113] Joint Committee of the ESAs, supra n 107, at p. 29.

[114] See Art. 17 MiFID II, Art. 79 CRD.

[115] See Art. 16 MiFID II, Art. 13 UCITSD, Art. 19(6) PSD II, Art. 38, 49 Solvency II.

functions of financial institutions do apply when an external provider is performing all or part of the outsourced functions through the use of (…) technologies."[116]

- Complying with record-keeping requirements,[117] given these requirements enable one to "reconstruct efficiently and evaluate the [tech] strategies/tools employed and ascertain compliance of financial institutions with all applicable regulatory requirements when providing services to consumers."[118]
- Taking steps to identify, prevent and manage conflicts of interests.[119] The ESAs acknowledge that the use of technology "can generate new contexts involving conflicts of interests, for instance from embedded biases or flaws in Big Data tools favoring firm's interests or certain clients over other clients."[120]

b. Business Principles

The ESAs further emphasize business principles requiring financial institutions to:

- Act honestly, fairly and professionally.[121] The ESAs insist that the "requirement to act fairly is of particular importance when the procedure or methodology being set-up or up-dated consists in the profiling of consumers."[122]
- Manufacture and distribute products and services which meet the needs of identified target clients and monitor such products.[123] Financial institutions should ensure that the use of data technologies to (i) identify target markets or (ii) assign a customer to a target market, is compliant with target market and product oversight requirements.
- Ensure that all information, including marketing communications, addressed by financial institutions to customers are fair, clear and not misleading.[124]
- Assess certain minimum, accurate and up-to-date, information about clients and products/services before providing certain services (e.g. suitability or appropriateness tests or creditworthiness assessments).[125]
- Preserve the interests of consumers when purchasing bundled or tied packages of products (in particular, client mobility and ability to make informed choices at the right time in the sales process):[126] "These provisions should prevent firms from using Big Data in order to promote bundled or tied packages of products

[116] Joint Committee of the ESAs, supra n 107, at p. 7.

[117] See Art. 17 MiFID II concerning algorithmic strategies. *See also* Art. 258(1)(i) Solvency II Delegated Regulation (EU) 2015/35, of Oct. 10, 2014. *See also* in the banking sector the Guidelines on outsourcing issued in Dec. 2006 by the Committee of European Banking Supervisors (CEBS) and the more recent Final Report of recommendations on outsourcing to cloud service providers published by the EBA in Dec. 2017.

[118] Joint Committee of the ESAs, supra n 107, at p. 30.

[119] Art 23 MiFID II, Art 17, 27, 28 IDD, Art 7 MCD. *See also* Art. 258(5) Solvency II Delegated Regulation (EU) 2015/35, of Oct. 10, 2014. *See also* EBA GL on product oversight and governance arrangements for retail banking products July 2015.

[120] Joint Committee of the ESAs, supra n 107, at p. 30.

[121] *See* Art. 24(1) MIFID II, Art. 17(1) IDD, Art. 7(1) MCD, Art. 12 AIFMD, Art. 14 UCITS.

[122] Joint Committee of the ESAs, supra n 107, at p. 30.

[123] Art. 16(3), 24(2) MiFID II, Art. 25 IDD, EBA GL on product oversight and governance requirements for manufactures and distributors of retail banking products, July 2015.

[124] *See* Art. 16 MiFID II, Art. 13 UCITS, Art. 19(6) PSD2*.

[125] *See* Art. 25 MiFID II, Art. 30 IDD, Art. 18, 20 MCD.

[126] *See* Art. 24(11) MiFID II, Art. 24 IDD, Art. 12 MCD, Art. 9 PAD, Art. 66, 67 of PSD.

which are not in the interests of clients."[127]

- Establish fair and efficient claims and complaints handling processes:[128] "This requirement is relevant to ensuring that Big Data analytics (e.g. tools enabling to predict more accurately whether a given consumer is likely or not to lodge a claim/complaint) do not lead to consumer detriment."[129]

c. Good practices

At the same time, the ESAs encourage the development and implementation of good practices with a view to "promoting a fair, transparent and non-discriminatory treatment of consumers and ensuring that Big Data strategies remain fully aligned with the interests of consumers."[130] Being summarized under somewhat loose headings, key aspects of good practices related to robust processes and algorithms, consumer protection and disclosures.

Demanding robust Big Data processes and algorithms, the ECB requires the "periodical monitoring of the functioning of Big Data procedures and methodologies as well as Big Data tools to adapt to technological developments and newly emerging risks".

Good practices pertaining to consumer protection require:

- the "periodical assessment whether Big Data based products and services are aligned with consumers' interests and where relevant, the review and adjustment of the Big Data tools",

- the "setting-up of procedures aimed at taking appropriate remedial actions when issues that may lead to consumer detriment materialize or are anticipated (notably in relation to the segmentation of consumers, e.g. impact on pricing or access of consumers to services due to increased segmentation of the target market)",

- the factoring of "potential risks associated with the use of Big Data together with the content of the financial institution's Big Data transparency policy when designing and enforcing the financial institution's complaint handling framework",

- the "adherence to and strict compliance with industry-specific codes of conduct under the GDPR",[131]

- "special attention to their policy in terms of processing of data gathered from social media platforms considering the varied level of understanding by consumers of privacy settings on social media accounts and the risks of inaccuracies in such data", as well as

- maintaining a balance between automated decision-making tools and human interventions.

[127] Joint Committee of the ESAs, supra n 107, at p. 31.

[128] *See* e.g. Art. 14 IDD, Art. 101 PSD2; Art. 26 MiFID II Delegated Regulation* requires firms to establish, implement and maintain effective and transparent procedures for the prompt handling of complaints.

[129] Joint Committee of the ESAs, supra n 107, at p. 32.

[130] Joint Committee of the ESAs, supra n 107, at p. 24.

[131] Financial institutions may choose to voluntarily join and adhere to approved codes of conduct or approved certification mechanisms, as an element to demonstrate compliance with GDPR (cf. Art. 24(3), 28(5), 40-43 GDPR).

Disclosure on the use of Big Data should ensure a high level of transparency towards customers concerning the use of Big Data technologies to process their data and promote "public awareness, consumer education on the phenomenon of big data and of consumers rights related to the use of Big Data by financial institutions."[132]

Remarkably, the ESAs did not stress two aspects of relevance to AI. First, the fact regulators may lack the means to monitor the limits of self-learning algorithms, and second, the role of senior management qualifications and responsibility. This will form the focus of the next sections.

2. Other Regulatory Approaches

An increasing range of other financial regulators are likewise engaging with AI. In chronological order:

- the Monetary Authority of Singapore introduced the new FEAT Principles to promote responsible use of AI and data analytics (considered below) in November 2018.[133]
- De Nederlandsche Bank issued principles for responsible use of AI, namely soundness, accountability, fairness, ethics, skills and transparency (or "SAFEST") in July 2019.[134]
- the WEF suggested in October 2019 that AI should be held to higher standards than humans and present systems as a result of the impact that AI can have on the financial services industry.[135]
- the HKMA issued its twelve "High-level Principles on Artificial Intelligence" in November 2019.[136]

a. Singapore

In November 2018, the Monetary Authority of Singapore (MAS) introduced the Principles to promote Fairness, Ethics, Accountability and Transparency (FEAT) in the use of AI and Data Analytics (AIDA) in decision-making in the provision of Singapore's Financial Sector[137]. These were updated in February, 2019 to reflect Singapore's Personal Data Protection Commission's Proposed AI Governance

[132] Joint Committee of the ESAs, supra n 107, at pp. 32-34.

[133] 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" (November 2018): https://www.mas.gov.sg/~/media/MAS/News%20and%20Publications/Monographs%20and%20Information%20Papers/FEAT%20Principles%20Final.pdf .

[134] De Nederlandsche Bank, "General Principles for Use of Artificial Intelligence in Finance" (25 Jul. 2019): https://www.dnb.nl/binaries/General%20principles%20for%20the%20use%20of%20Artificial%20Intelligence%20in%20the%20financial%20sector_tcm46-385055.pdf .

[135] World Economic Forum, "Navigating uncharted waters", supra n 69.

[136] Hong Kong Monetary Authority, "High-Level Principles on Artificial Intelligence" (1 Nov. 2019): <https://www.hkma.gov.hk/media/eng/doc/key-information/guidelines-and-circular/2019/20191101e1.pdf>.

[137] 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" < >. (12 Nov. 2018) <https://www.mas.gov.sg/~/media/MAS/News%20and%20Publications/Monographs%20and%20Information%20Papers/FEAT%20Principles%20Final.pdf > .

Framework[138] that had been issued in January 2019. The Proposed Model AI Governance Framework has two guiding principles, namely that organizations must ensure that decision-making using AI is explainable, transparent and fair, and that AI solutions should be human-centric. 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 minimizing inherent bias, and
(4) Customer relationship management, including disclosure, transparency, and explainability.

In November 2019, the MAS announced the creation of the Veritas framework to promote the responsible adoption of AIDA by financial institutions using open source tools as a verifiable way for financial institutions to incorporate the FEAT principles. With an initial consortium of 17 members, Veritas will initially focus on customer marketing, risk scoring and fraud detection.[139]

b. Hong Kong SAR

In Hong Kong, in May 2019, the HKMA encouraged[140] authorized institutions to adopt and implement Hong Kong's Office of the Privacy Commissioner for Personal Data's Ethical Accountability Framework for the collection and use of personal data,[141] and its Data Stewardship Accountability, Data Impact Assessments and Oversight Models that were introduced in October the prior year.[142]

In November, 2019, the HKMA's Banking Supervision department published its High-Level Principles on AI.[143] These Principles require that bank boards and senior management be accountable for the outcome of AI applications. In particular, the Principles reinforce that banks should:

(1) Possess sufficient expertise;
(2) Ensure appropriate level of explainability of AI applications;
(3) Use data of good quality;

[138] Singapore Personal Data Protection Commission, "A Proposed Artificial Intelligence Governance Model" (Jan. 2019) <https://www.pdpc.gov.sg/-/media/Files/PDPC/PDF-Files/Resource-for-Organisation/AI/A-Proposed-Model-AI-Governance-Framework-January-2019.pdf>.

[139] Monetary Authority of Singapore, "MAS Partners Financial Industry to Create Framework for Responsible Use of AI" (13 Nov. 2019) <https://www.mas.gov.sg/news/media-releases/2019/mas-partners-financial-industry-to-create-framework-for-responsible-use-of-ai#1>

[140] Hong Kong Monetary Authority, "Use of Personal Data in Fintech Development" (3 May 2019) <https://www.hkma.gov.hk/media/eng/doc/key-information/guidelines-and-circular/2019/20190503e1.pdf>

[141] Hong Kong Office of the Privacy Commissioner for Personal Data, Ethical Accountability Framework for the collection and use of personal data (24 Oct. 2018) <https://www.pcpd.org.hk/misc/files/Ethical_Accountability_Framework.pdf>

[142] Hong Kong Office of the Privacy Commissioner for Personal Data, Data Stewardship Accountability, Data Impact Assessments and Oversight Models : Detailed Support for an Ethical Accountability Framework (24 Oct. 2018): <https://www.pcpd.org.hk/misc/files/Ethical_Accountability_Framework_Detailed_Support.pdf>

[143] Hong Kong Monetary Authority, "High-Level Principles on Artificial Intelligence" (1 Nov. 2019) <https://www.hkma.gov.hk/media/eng/doc/key-information/guidelines-and-circular/2019/20191101e1.pdf>

(4) Conduct rigorous model validation;
(5) Ensure auditability of AI applications;
(6) Implement effective management oversight of third-party vendors;
(7) Be ethical, fair and transparent;
(8) Conduct periodic reviews and on-going monitoring;
(9) Comply with data protection requirements;
(10) Implement effective cybersecurity measures; and
(11) Implement risk mitigation and contingency plans.

A few days later, the HKMA's Banking Conduct Department issued Guiding Principles on Consumer Protection in respect of Use of Big Data Analytics and AI (BDAI) by Authorized Institutions.[144] These guiding principles reinforced a risk-based approach to BDAI and focussed on four major areas, namely governance and accountability, fairness, transparency and disclosure, and data privacy and protection.

The HKMA's High-Level Principles on AI clearly set forth the expectation that "The board and senior management of banks should appreciate that they remain accountable for all AI-driven decisions", and that "the roles and responsibilities of the three lines of defence in developing and monitoring the operations of AI applications should be clearly defined."[145] This was reinforced in the HKMA's BDAI consumer protection guidance.[146]

C. Possible Regulatory Approaches

Current regulation focuses on human conduct, imposes safeguards on presumed static systems the vulnerabilities of which are not examined frequently, and entrenches peremptory transparency and auditability requirements.[147]

While designed as "high-level frameworks", the very fact that these guidelines have been issued by financial supervisory authorities turns these into more than mere "recommendations", into law *de facto*, if not in form: financial institutions subject to supervision will find it difficult to evade these supervisory expectations, with or without an authority's rule making capacity. This justifies a closer look at the measures available to financial supervisors in regulating AI.

In the following sections, we focus on five examples: authorization of AI itself, outsourcing rules and e-personhood, the qualifications of core personnel, the role of AI with regard to key functions, and sanctioning rules.

1. Authorization of AI

Enhanced use of AI influences the conditions for authorization. In particular, if a business model seeking authorization relies on AI, the business and operations plan

[144] Hong Kong Monetary Authority, "Consumer Protection in respect of Use of Big Data Analytics and Artificial Intelligence by Authorized Institutions" (5 Nov. 2019): <https://www.hkma.gov.hk/media/eng/doc/key-information/guidelines-and-circular/2019/20191105e1.pdf>

[145] Hong Kong Monetary Authority, "High-Level Principles on Artificial Intelligence", supra n. 135, page 2 para 1.

[146] HKMA, supra n. 154, page 2: para 1: "The board and senior management of AIs should remain accountable for all the BDAI-driven decisions and processes."

[147] World Economic Forum, "Navigating uncharted waters", supra n 69.

must lay out both the functioning of the AI itself, and 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 around the globe currently 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 discussed in the literature, however, is the introduction of a licensing requirement for AI being used by financial intermediaries.[148] Another potential response is a mandatory insurance scheme for AI.

Currently, financial services authorities worldwide are themselves increasingly seeking to upskill and introduce supervisory technology or suptech to perform meaningful reviews of AI. Software to monitor a self-learning AI's conduct does not, to our knowledge, yet exist, and 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 use.

AI authorization may also have a number of undesirable side-effects. The most important one is that authorization is potentially harmful for innovation given authorization is costly and takes time. It is also uncertain how rules could be drafted to reflect the daily reality of AI programming that minor amendments and improvements take place on almost a daily basis. Re-authorization of the code in this case will increase costs even further, meaning only AI with major income potential will be developed, and minor improvements of existing AI may well be uneconomic. Finally, in the case of self-learning AI, the actual authorized code will not be performing in practice, as the definition of self-learning AI is that it further develops its code while performing its services. Any authorization will thus be always outdated.[149] While sandboxes may in some settings be useful instruments for supporting innovation and effective regulation,[150] the authority can at best assess the services performed while the AI is functioning under sandbox conditions, thereby neglecting its performance under real conditions.[151] At the same time, fostering AI-related RegTech is independent of an AI's authorization (or sandbox, as the case may be); as it notably requires data-related reporting and governance rules.[152]

2. Regulatory outsourcing rules and e-personhood

In regulatory rulebooks around the world, crucial supplier frameworks apply if the AI is owned and operated by a separate services provider. If this is the case, the crucial supplier should be subject to additional monitoring by the outsourcing intermediary.

[148] *See* Andrew Tutt, *An FDA for Algorithms*, 69 ADMIN. L. REV. 83 (2017).

[149] *See* Enriques & Zetzsche, "Corporate Technologies", supra n 24, at 56.

[150] World Economic Forum, "Navigating uncharted waters", supra n 69; RP Buckley, DW Arner, R Veidt & DA Zetzsche, "Building FinTech Ecosystems: Regulatory Sandboxes, Innovation Hubs and Beyond", *Washington University Journal of Law & Policy* (forthcoming 2020)*;* and DA Zetzsche, RP Buckley, DW Arner & JN Barberis, "Regulating a Revolution: From Regulatory Sandboxes to Smart Regulation", (2017) (1) *Fordham Journal of Corporate & Financial Law* 31 (2017).

[151] *See* Enriques & Zetzsche, "Corporate Technologies", supra n 24, at 56.

[152] *See* Dirk Zetzsche, Douglas W. Arner, Ross P. Buckley, Rolf H. Weber, "The Future of Data-Driven Finance and RegTech: Lessons from EU Big Bang II" , EBI Working Paper Series 2019/35, <https://ssrn.com/abstract=3359399 >.

The reality of much AI in financial services will, however, be that the AI is owned and operated in-house, by the financial intermediary's own staff. This prompts the question of the adequacy of the legal framework covering the AI.

One option for regulating in-house AI is the granting of limited legal personality to the algorithm itself, similar to a partial license, paired with minimum capital requirements. If the capital is depleted, for instance due to liabilities or regulatory sanctions, the algorithm needs to stop operations. The argument against such a limited e-personhood are similar to those against authorizing AI: The calculation of capital requires a clear delineation of risks created by the AI. If the limits of the function of the AI itself is in doubt, as is the case with regard to self-learning algorithms, regulatory capital will most likely be set too low or too high.

Further, authorities have less expensive ways to restrict the use of AI, even in the absence of an AI's own regulatory capital. These include imposing reporting requirements for AI-prompted damages upon intermediaries that employ AI, and responding to such reporting by issuing orders limiting, or prohibiting, the use of the AI.

3. AI as key function holder?

Another aspect of the fit and proper test refers to the use of AI as an executive or board member of the intermediary.[153] In this regard, legality and practical feasibility may be two different things. As to legality: 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 as a SPV's sole activity. In other jurisdictions, these functions must be occupied by humans. As to practical feasibility, we could envision the AI functioning as a board member for certain routine tasks (the literature discusses the example of securitization vehicles in a corporate group), as well as for monitoring and supervisory services of a procedural nature, but would ask for a human board majority in order to ensure continuing operations when, and if, challenges exceed the limits of the programming of the AI.

Notwithstanding this, any rules allowing AI to assume some or all key functions of a financial intermediary must respect the existing limits of AI. This is particularly true for compliance monitoring. AI, on a stand-alone basis, is poorly adapted to handle compliance matters. The reason lies less in the lack of ethical screening abilities, and in the way rules are drafted: rules are incomplete on purpose. The law is full of vague terms such as "fair", "adequate", "just", "reasonable person" etc. These terms are used to ensure adjustment to an ever-changing world. Financial services are, however, a heavily regulated environment with plenty of rules and hence a lot of vagueness originates from these broad terms. These terms cannot be defined in 1/0 (yes/no) terms, and their meaning changes from context to context. If AI functioned as a compliance officer, we would thus expect inaccurate monitoring, widespread misreporting, and mispricing of risks all arising from vagueness in the law.[154]

[153] *See* note on VITAL *supra n* 24.

[154] *See* Enriques & Zetzsche, "Corporate Technologies", supra n 24, at 34-35.

4. Fit and Proper Test

One field where AI will most likely influence regulatory practice is the fit and proper test for key function holders (i.e. senior management or executives) as well as the board of directors. AI will impact existing licensing conditions in two respects. First, some existing requirements may be less necessary if an AI is doing the job. If in fact most decisions are taken by AI why should supervisors review a human executive's credentials?

Second, new requirements will reflect the greater reliance on AI, and some office holders may have new qualifications. For instance, 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.

We have argued that there is little merit in reviewing AI itself in the context of AI authorization (*supra*, at III.B.4.); the same argument applies to assessing how fit and proper an AI may be. The increasing use of AI will, however, impact on the fit and proper test of humans functioning in AI-heavy financial institutions. This will almost certainly require modifications to existing regulatory approaches: AI experts may have accumulated their AI experience outside of the financial sector, for instance within a major e-commerce firm, given that technical innovation useful for financial services takes places in these firms. If financial supervisors insisted on their three year standard in financial firms, the supervised entities may find little tech expertise for hire. Authorities may need to modify their experience requirements for the financial sector, choosing to value high level AI experience in other sectors, so as to strengthen the firm's internal controls.

Given we believe senior management qualifications to be one of the most important regulatory tools in responding to the use of AI in financial services, we discuss these matters in more detail in the next part (IV.).

5. Sanctioning

Financial regulation imposes sanctions, some directed at at the institution's overall conduct and some others at staff member conduct. Usually, financial supervisors need to show some type of negligence or ill intent on the side of the financial institution in order to impose a sanction, with a deficiency of risk management system providing a fall-back option for sanctioning in case any harm has materialized. In the AI age, these cases will be increasingly hard to make. Where AI fails and even supervisors are incapable of establishing an AI's processes and limits with certainty; determining the culpability standard and burden of proof to be applied that will impose prudent sanctions while retaining incentives to innovate is going to be very difficult. After all, the nature of innovation is that innovations fail or do harm. Executives can do little more than select AI to the best of their abilities; where these abilities authorized under the fit and proper test fail potential sanctions may have exercised little steering effect, even if sanctions are possible under the broad "failure of risk management" rationale.

This brings us to the broadly discussed question of how to sanction an AI. Withholding compensation, naming and shaming, and financial penalties have little meaning for AI. In a similar vein, director disqualification, the equivalent of a "death penalty", as well as civil and criminal liability, provide limited steering effect for AI in the current form, unless the AI is programmed to have a desire to survive.

Hence, the sanctioning system must be reconsidered and include how to set proper incentives for the AI itself. AI-adapted financial regulation would possibly (i) require blame-free remediation in which organizations are able to learn from failures and make improvements, (ii) encourage forward-thinking collaboration between industry players to promote early detection and the avoidance of unexpected failures in AI systems, and (iii) employ fit-for-purpose explainability in which frameworks are utilised to decide "if" explainability is a requirement (thereby assisting organizations to prioritize their AI's objectives) and "how" explainability should be achieved given the wide range of AI use-cases.[155] Only where a conduct infringes said "if" and "how" rules would sanctions apply.

IV. Putting the Human-in-the-loop into Finance

While regulators expect financial institutions to deploy AI in a responsible manner and therefore develop and become accustomed to using new tools and solutions to safeguard the financial system,[156] we have shown that AI poses particular challenges from a regulatory standpoint: not all forms of financial services regulation are well-suited to ensuring the responsible use of AI, given the enhanced severity of information asymmetry, data dependency and interdependency.

In particular, given challenges of the "black box" problem in AI for regulatory and supervisory authorities, we argue in this section that measures focusing on personal responsibility requirements that put the "human-in-the-loop", should instead be the focus of regulating AI-enabled systems in finance.

Two particular approaches seem to be gaining increasing currency. The first involves the use of technology (including AI) to monitor staff behaviour and identify issues ideally before they arise (which should be seen as a form of RegTech). As we have argued elsewhere, we understand RegTech as logical consequence of enhancing Fintech; FinTech cannot work without proper RegTech in place. This is not the place to repeat this argument.

We thus turn to the second approach. This involves an increasing range of regulatory systems based on personal responsibility of designated senior managers for areas under their supervision – so-called "senior manager", "manager in charge", "key function holders" or "personal responsibility" systems. We argue that regulators should utilize and strengthen these external governance requirements in order to require "human-in-the-loop" systems for internal AI governance.

This approach builds on existing trends in financial regulation which have developed as a result of the Global Financial Crisis, LIBOR and forex scandals. These frameworks seek to produce cultural change and an ethical environment in financial institutions through personal responsibility of directors, management and, increasingly, individual managers.

We suggest that such personal responsibility frameworks should be supplemented to include responsibility for AI, including a non-waivable AI due diligence and explainability standard. Finally, we discuss particularities of an AI-adjusted personal

[155] *See* AI Accenture, supra n 4, at 18; UK Finance, supra n 2, at 10-13; World Economic Forum, "Unchartered Waters", supra n 65, at 21.

[156] World Economic Forum, "Navigating uncharted waters", supra n 65.

responsibility framework to ensure appropriate incentives. Such systems are particularly suited to "black box" issues but are also an effective approach for the range of major financial risks we identify in terms of data, cybersecurity, systemic risk, and ethics.

A. External Governance Requirements to Transform Internal Governance and Culture: Personal Responsibility Frameworks in Finance

Over the past decade, most major financial jurisdictions have imposed, or are in the process of imposing, director and manager responsibility frameworks through financial regulation. The EU has developed a framework for internal governance, the UK, Australia, and Hong Kong have implemented manager responsibility regimes, and Singapore and the US have proposed regimes.

1. European Union

The EU joint internal governance guidelines were published by the EBA and ESMA 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. 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.[157]

The EBA and ESMA internal governance guidelines, and EIOPA's guidelines on systems of governance,[158] apply to all kinds of financial services institutions regulated under EU law, notably credit institutions, investment firms, managers of collective investment schemes, insurance undertakings and financial holding companies. These guidelines govern the conduct of the management body and key function holders. "Key function holders" is a term that refers to persons with significant influence over the direction of the institution that are not part of the management body. The management body and key function holders are to possess good repute, independence, honesty, integrity, knowledge, skills, and experience. Members of the management body must have sufficient time to perform their functions including understanding the business of the institution, its main risks, and the implications of the business and risk strategy.[159]

Responsibilities of the management body (in particular the CEO and other key executives) include setting, approving, and overseeing implementation of the overall

[157] These guidelines are to be read in conjunction with other guidelines and associated materials. See European Banking Authority, "Final Report - Guidelines on internal governance under Directive 2013/36/EU" (20 Sep. 2017) EBA/GL/2017/11, 5-7; "EBA and ESMA provide guidance to assess the suitability of management body members and key function holders" (26 Sep. 2017) <https://eba.europa.eu/eba-and-esma-provide-guidance-to-assess-the-suitability-of-management-body-members-and-key-function-holders>; Commission Delegated Regulation (EU) No 604/2014.

[158] EIOPA, Guidelines on Systems of Governance: <https://eiopa.europa.eu/GuidelinesSII/EIOPA_Guidelines_on_System_of_Governance_EN.pdf> (content-wise, these guidelines are essentially the same as the EBA and ESMA guidelines, only the older solvency framework for insurance undertakings from 2009 required a different wording.)

[159] European Banking Authority & European Securities and Markets Authority, "Guidelines on the assessment of the suitability of members of the management body and key function holders" (Mar. 21, 2018) ESMA71-99-598 EBA/GL/2017/12, 3 para 6, 5 para 3, 6, 11 para 26, 13 para 37, and 14 paras 39 and 41.

business strategy and the key legal and regulatory policies, the overall risk strategy, internal governance and control, risk capital, liquidity targets, remuneration policy, key functional holders' assessment policy, internal committees functionality, risk culture, corporate culture, conflict of interest policy, and the integrity of accounting and financial reporting systems.[160] The management body is also accountable for the implementation of the governance arrangements that ensure effective and prudential management of the institution, and promote the integrity of the market and the interests of clients.[161]

Key function holders such as heads of internal control functions including risk management, compliance and audit functions have a key role in ensuring that the institution adheres to its risk strategy, complies with legal and regulatory requirements, and ensures robust governance arrangements.[162] 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 in relation to the institution's core values and expectations of staff. Effective communication should promote an environment of open communication, welcoming challenges in the decision-making processes, encouraging a broad range of views, allowing for the testing of current practices, stimulating a constructive critical attitude, and promoting an environment of open and constructive engagement throughout the entire organization.[163] The principal of proportionality applies to all governance arrangements, consistent with the individual risk profile and business model of the institution.[164]

2. United Kingdom: Senior Managers and Certification Regime

The UK's Senior Management regulatory regime evolved from the overall EU framework and has been highly influential internationally. Compliance with the regime is subject to firms and individuals being authorized by the UK Prudential Regulation Authority (PRA) and the Financial Conduct Authority (FCA). Authorized firms are required to ensure that individuals who perform PRA-designated senior management functions are approved.[165] Authorization 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 (FSMA).[166]

[160] European Banking 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" (26 Sep. 2017) EBA/GL/2017/12, 18-20 para 23.

[161] European Banking Authority & European Securities and Markets Authority, "Guidelines on the assessment of the suitability of members of the management body and key function holders" (21 Mar. 2018) ESMA71-99-598 EBA/GL/2017/12, 6, 11 para 26, 13 para 37, 14 paras 39 and 41, and 31 para 110.

[162] European Banking 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" (26 Sep. 2017) EBA/GL/2017/12, 11 para 33.

[163] European Banking Authority, "Final Report - Guidelines on internal governance under Directive 2013/36/EU" (20 Sep. 2017) EBA/GL/2017/11, 34 para 98.

[164] European Banking Authority & European Securities and Markets Authority, "Guidelines on the assessment of the suitability of members of the management body and key function holders" (21 Mar. 2018) ESMA71-99-598 EBA/GL/2017/12, 9 para 20.

[165] Pursuant to s. 59 of the Financial Services and Markets Act 2000.

[166] Conduct rules apply to the senior management functions specified by the PRA and FCA pursuant to s. 63 of the FSMA. See Bank of England, 'Senior Managers Regime: approvals'

Following the promulgation of the Commissioned Delegated Regulation (EU) No 604/2014, the PRA replaced the Approved Person Regime with the Senior Managers and Certification Regime (2016 SMCR) in March 2016. The 2016 SMCR is regulated by the PRA and the FCA and applies to all individuals who perform a "Senior Management Function" at banks, building societies, credit unions, and PRA-designated investment firms. The 2016 SMCR was expanded to cover insurance firms in November 2018, and expanded again, for FCA-regulated financial institutions, to apply to asset managers and designated activities of investment firms (Extended SMCR) from December 2019.[167]

The 2016 SMCR applies to UK deposit takers, PRA-designated investment firms, and UK branches of foreign banks. It is structured around: (1) a Senior Managers Regime for individuals who require regulatory approval (i.e. senior management functions and prescribed responsibilities); (2) a certification regime for regulated firms to assess the fitness and propriety of employees carrying out a "significant harm" function; and (3) conduct rules which apply to most bank employees.[168]

Senior managers are each required to have a clear and succinct statement of responsibilities. These include prescribed responsibilities listed by the regulator. Conduct rules for senior managers specify a "Duty of Responsibility" by taking "reasonable steps" to ensure that the business of the firm is controlled effectively and complies with the regulatory framework. Senior managers must take reasonable steps to ensure that any delegation of responsibility is assigned to an appropriate person and oversee an effective discharge of the delegated responsibility. A senior manager must disclose any information of which the PRA or FCA would reasonably expect notice.[169] The FCA has clearly expressed that the 2016 SMCR is not intended to subvert collective responsibility or collective decision-making.[170]

Conduct rules encourage a healthy culture whereby all financial services 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 proper standards of market conduct. Firms are accountable for employee conduct and are required to notify the regulator of any breach of the conduct rules.[171]

The scope of the Extended SMCR is slightly wider than the 2016 SMCR. 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, in terms of asset management firms, the value for money

<https://www.bankofengland.co.uk/prudential-regulation/authorisations/senior-managers-regime-approvals>.

[167] B. Reynolds, T. Donegan, S. Dodds & J. Adams, "The UK's Expanded Senior Managers and Certification Regime: Key Issues and Action Plan For Brokers, Advisors and Asset Managers" (8 Jul. 2019) Shearman & Sterling <https://www.shearman.com/perspectives/2019/07/the-uks-expanded-senior-managers-and-certification-regime-key-issues-and-action-pan>.

[168] Linklaters, 'SMCR for deposit takers and PRA-designated investment firms' <https://www.linklaters.com/en/insights/publications/smcr/smcr/smcr-for-deposit-takers-and-pra-designated-investment-firms>.

[169] KPMG, "Individual Accountability: Global regulatory developments in financial services" (July 2018), 4-5.

[170] Allen & Overy, "The UK Senior Managers and Certification Regime: Themes, trends and challenges from the first three years" (March 2019), at 17.

[171] Debevoise & Plimpton, "The UK's Senior Managers and Certification Regime" (18 Feb. 2019), para 4.1.

assessments, independent director representation, and acting in investors' best interests. This last point is applicable to managers of authorized (retail) funds.[172]

Ultimately, it is broadly recognized that these considerations also apply to the board,[173] where the need for upskilling similarly applies.

3. Australia: Banking Executive Accountability Regime

The Australian Prudential Regulation Authority (APRA) administers the Banking Executive Accounting Regime (BEAR).[174] Steps are being taken for Australian Securities and Investment Commission (ASIC) to co-regulate the BEAR obligations with APRA. Given ASIC is a conduct-based regulator, it appears well suited to regulate BEAR's conduct requirements.[175]

The BEAR came into effect on 1 July 2018 for large banks and 1 July 2019 for smaller banks (collectively, authorized deposit-taking institutions).[176] Both authorized deposit-taking institutions (ADIs) and individual accountable persons (IAPs) have responsibilities under BEAR. The ADI must provide individual accountability statements to APRA which clearly outline individual responsibilities and provide an accountability map outlining how accountability is allocated across an institution (based on size, risk profile, and complexity). IAPs are accountable for their actual or effective responsibilities for the management or control of a significant or substantial part, or aspect of, the ADI's operations or an ADI group. Specifically, IAPs have obligations to: act with honesty and integrity, and with due skill, care, and diligence; deal with APRA in an open, constructive, and co-operative way; and take reasonable steps in conducting their responsibilities to prevent matters arising that would adversely affect the ADI's prudential standing or prudential reputation.[177]

4. Hong Kong: Securities Firm Managers in Charge/Senior Management

In relation to Hong Kong securities firms, senior management are defined as directors and "responsible officers" of a corporation, and "Managers-in-Charge" (MICs). Licensed corporations are required to appoint an MIC as the person primarily responsible for each core function, overall management oversight, key business lines, operational control and review, risk management, finance and accounting, information technology, compliance, and AML/CFT. For each core function there should be at least

[172] ibid para 2.4.

[173] *See* e.g., Financial Conduct Authority, "Artificial Intelligence in the Boardroom" (Insight, 1 Aug. 2019) <https://www.fca.org.uk/publication/research/research-note-on-machine-learning-in-uk-financial-services.pdf>

[174] BEAR is outlined in an information paper which recommends that it be read in conjunction with the requirements for accountability in Part IIAA of the Banking Act 1959, and an accompanying Revised Explanatory Memorandum. See APRA, "Information Paper: Implementing the Banking Executive Accountability Regime" (17 Oct. 2018), 4.

[175] ASIC, "ASIC update on implementation of Royal Commission recommendations" (19 Feb. 2019) <https://download.asic.gov.au/media/5011933/asic-update-on-implementation-of-royal-commission-recommendations.pdf>, 5 & 11.

[176] BEAR is set out in Part IIAA of the Banking Act 1959.

[177] APRA "Information Paper: Implementing the Banking Executive Accountability Regime" (17 Oct. 2018), sub-s 1.2.

one MIC responsible, although one MIC 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 SFC" (hereinafter, SFC Code of Conduct) states that senior management shall bear primary responsibility for ensuring the maintenance of appropriate standards of conduct and adherence to proper procedures by the firm. When determining responsibility in relation to a business operation, a person's actual and apparent authority shall be considered to determine responsibility and the degree of responsibility.[178] The Board shall approve and adopt a formal document clearly setting out, amongst other roles, responsibilities, accountability, and the reporting lines of senior management.[179]

Paragraph 14.1 of the SFC Code of Conduct specifies that senior management of a licensed corporation should properly manage the risks associated with the business of a corporation, including performing periodic evaluation of its risk processes, understanding the business nature of the corporation, its internal control procedures and its policy on the assumption of risk; and understanding the extent of their own authority and responsibilities.[180] Senior management are ultimately responsible for the adequacy and effectiveness of the corporation's internal control systems which include information management, compliance, audit or related reviews, operational controls, and risk management.[181] MICs should be aware of other codes and guidelines which impose responsibilities pursuant to section 193(3) of the Securities and Futures Ordinance (Cap. 571).[182]

5. United States: Proposed Senior Management Guidance for banks

In early 2018, the US Federal Reserve issued proposed senior management guidelines. The guidelines cover the senior management of large banks, bank-like institutions, and non-bank Systemically Important Financial Institutions (SIFIs). When the guidelines are formalized, they will build upon the independent risk management framework in Regulation YY which, in turn, implements certain provisions in sections 165 and 166 of the Dodd-Frank Wall Street Reform and Consumer Protection Act.[183]

Senior management is defined as the core group of individuals directly accountable to the board of directors for the sound and prudent day-to-day management of the firm. For foreign-bank holding companies, senior management refers to those individuals inside or outside the US who are accountable to the intermediate holding-company board, US risk committee, or global board of directors with respect to their US operations.

Senior management are responsible for managing the day-to-day operations of the firm and ensuring safety and soundness, and compliance with laws, regulations (including consumer protection), and internal policies and procedures. The two key responsibilities of senior management are overseeing the activities of the firm's

[178] SFC, 'Circular to Licensed Corporations Regarding Measures for Augmenting the Accountability of Senior Management' (Dec. 16, 2016), paras 1, 5, 7, 8, and 9.
[179] ibid para 28.
[180] ibid para 14(b).
[181] ibid para 14 (c). Referring to the Internal Control Guidelines.
[182] ibid para 14 (19).
[183] Federal Reserve, "Proposed Supervisory Guidance" (11 Jan. 2018) [Docket No. OP-1594] 83 *Federal Register* 8, 1353 <https://www.govinfo.gov/content/pkg/FR-2018-01-11/pdf/2018-00294.pdf>.

business lines (individually or collectively); and the firm's independent risk management and system of internal control. There are additional responsibilities for certain senior managers, such as the chief risk officer in relation to independent risk management and the chief audit executive in relation to the internal audit function.

Senior management are responsible for maintaining and implementing an effective risk management framework and ensuring that risk is appropriately managed in a manner consistent with the firm's strategy and risk tolerance. Furthermore, senior management is responsible for promoting and enforcing prudent risk-taking behaviours and business practices. Senior management should periodically assess the firm's risk-management framework and ensure that the framework is comprehensive and appropriate for the firm's business lines and changes in economic and market conditions. Effective communication and information sharing should be maintained across the entire firm, including providing timely, useful, and accurate information to the board.[184]

6. Singapore: Proposed Senior Manager Guidelines

In June 2019, the MAS issued Proposed Guidelines on Individual Accountability and Conduct (IAC Proposed Guidelines). Senior managers are responsible for the day-to-day operations of a financial institution in Singapore.[185] The IAC Proposed Guidelines state that senior managers are responsible for the management and conduct of "core management functions" (CMFs), for the actions of their staff, and the conduct of the business.[186] Financial institutions should apply CMF definitions which reflect the actual responsibilities of a particular senior manager.[187] Responsibility is described as "principles-based" and therefore a list of mandatory responsibilities has not been issued.[188] MAS states that the level of responsibility should reflect the senior manager's roles in relation to the financial institution's Singaporean operations.[189] Senior managers are responsible regardless of their title or whether they are based overseas.[190] Material Risk Personnel are also covered by the IAC Proposed Guidelines.

B. Adressing the Knowledge Gap

The trend in financial services regulation appears clear: increasing personal responsibility for senior management and other individuals responsible for regulated activities within financial institutions. Such frameworks should also apply to AI.

We suggest that such personal responsibility frameworks provide the basis of an appropriate system to address issues arising from AI in finance, in particular the three challenges of AI (information asymmetry, data dependency and interdependency). We propose that manager responsibility framework need to be expanded to specifically incorporate responsibility for AI in regulated activities, thus mandating a "human-in-the-loop". This should be extended to specifically mandate due diligence and explainability requirements. Such an approach could be augmented in many cases through the addition of AI review committees. Such an approach is highly effective in

[184] Ibid.
[185] IAC Proposed Guidelines (6 Jun. 2019), para 3.3.
[186] ibid paras 1.1 & 3.1.
[187] ibid para 3.23. For a definition of CMFs in relation to Senior Management, see ibid Annex C, 50ff.
[188] ibid para 3.23.
[189] ibid para 2.25
[190] ibid para 3.5.

addressing black box issues but also in providing a framework to address the four core financial risks we identify relating to data, cybersecurity, systemic risk, and ethics.

1. AI review committees

In order to address the information asymmetry as to AI's functions and limits, regulators should take advantage of an important practice emerging in some non-financial companies. These companies have created independent AI review committees to provide cross-disciplinary and impartial expertise to such companies developing and utilising AI.[191] Some of these committees or boards have been quite impactful, such as 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.[192] The impact of others have been less,[193] or remain to be seen.[194] In any case, these boards are designed to augment decision-making and do not detract from the ultimate responsibility vested in management and the board regarding AI governance.

2. AI Due Diligence

The second tool that reinforces and supports manager responsibility is mandatory AI due diligence. Due diligence is meant to include a full stock-taking 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 is the standard prior to AI employment, while AI explainability is the standard to meet throughout the use of any AI.

In order to reflect data dependency one part of the due diligence is a mapping of the data sets used by the AI, including an analysis of data gaps and data quality.

AI due diligence is a result of individual responsibility systems: the necessity of the individual having performed sufficient due diligence in exercise of their responsibilities to avoid liability for any failures which arise, whether from internal governance systems, employees, third parties, or IT systems.

[191] See Brian W Tang, "Independent AI Ethics Committees and ESG Corporate Reporting on AI as emerging corporate and AI governance trends" in Ivana Bartoletti, Susanne Chishti, Anne Leslie and Shan M. Millie (ed), *The AI Book: The Financial Technology Handbook for Investors, Entrepreneurs and Visionaries in Regulation* (Wiley, forthcoming 2020).

[192] See Rick Smith, "The Future of Face Matching at Axon and AI Ethics Board Report", Axon, (27 Jun. 2019) < https://global.axon.com/company/news/ai-ethics-board-report >/.

[193] See e.g., "Google Quietly Disbanded Another AI Review Board Following Disagreements", Wall Street Journal (16 Apr. 2019).

[194] See Facebook's new Oversight Board: "Establishing Structure and Governance For an Independent Oversight Board" (Facebook, 17 Sep. 2019) < https://about.fb.com/news/2019/09/oversight-board-structure/ >. Megvii Technology Limited, one of the first pure-play AI companies from China seeking to be listed, has set up an AI Ethics Committee: see Megvii Technology Limited, Application Proof filed with the Stock Exchange of Hong Kong, p.3 <https://www1.hkexnews.hk/app/sehk/2019/100283/documents/sehk19082500082.pdf>

3. AI Explainability

Explainability requirements are necessary as a minimum standard for humans-in-the-loop in AI use, i.e. demanding that the function, limits and risks of AI can be explained *to someone*, at a level of granularity that enables remanufacturing of the code.[195]

In principle, the someone should be the member of the executive board responsible for the AI (thus relying on the managers' incentive to avoid sanctions) or an external institution, in particular regulators, supervisors and courts.

We thus propose, in a first step, to introduce explainability requirements for the responsible managers; and a requirement that the explainability be documented. In a second step, supervisory authorities may review compliance with the explainability requirements. In this way manager responsibility systems will be buttressed by explainability systems which are in turn a necessary result of personal responsibility and accountability to regulators on an individual level for regulated functions. Individual managers will have to be able to explain their own decisions, the actions of their employees and contracts, and of their IT systems.

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

A number of concerns must be considered in the context of promoting the personal responsibility model. These include: (1) the inability to control AI well by internal governance, (2) overdeterrence, and (3) how to deal with FinTech start-ups.

1. Inability to control autonomous AI internally

If AI cannot be controlled by external monitors, such as financial supervisory authorities, it could be argued that AI cannot be monitored effectively and kept under control by senior management not directly involved in AI coding and operations. In this case, key staff will lack control over the AI, just as will external supervisors.

Methods of internal control include, for instance, 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. If key function holders / senior management are well aware of their responsibilities, in most cases these governance tools will be imposed with a view to controlling AI since the key managers' income expectation and future cash-flow opportunities depend on meeting their responsibility.

More importantly, personal responsibility / liability systems place the responsibility for areas of regulated conduct under the responsibility of specific individuals, thus meaning that an individual is directly responsible from a regulatory standpoint for regulatory breaches which arise in their area of responsibility. Thus, the individual will have strong incentive to monitor and understand their functional area, their staff, their third parties contractors and suppliers, IT systems. Once such understanding of responsibility develops, a culture of due diligence and explainability should evolve to address the "black box" problem. In cases where it does not, the individual and board will nonetheless remain responsible for developments.

[195] *See* on explainability World Economic Forum, "Navigating uncharted waters", supra n 69, at 32.

Naturally, the manager responsibility model requires including key people, for purposes of AI development, in the responsibility concept. Hence, key developers (to the extent the solution is developed internally) must be included in the net of responsibility. As we have argued in relation to TechRisk, an individual should also be designated with regulatory responsibility for IT and tech systems, for similar reasons and to achieve similar results.[196]

The manager responsibility concept may prove ineffective in two cases. First, if the developers lose control over self-learning AI, as can occur if, for instance, self-learning AI taps into unexpected data pools, and produces unexpected correlations. However, the production of unacceptable and unexpected outcomes can be countered by switching off the AI, an outcome which should be incentivised through personal responsibility requirements. Accordingly, given the risk of global systemic risk and impact on lives that finance plays, all AI used in finance should be programmed so as to be able to be switched off: the responsibility model should be designed to ensure that this indispensable requirement is in the code, and, most importantly, the organization needs to be able to function with the AI turned off. A contingency plan is vital and needs to include (a) the option to switch off the AI, and (b) the measures that will be instituted to deal with the consequences of doing so (such as manual, instead of algorithmic, trading, manual loan portfolio allocation, etc.).

Second, the responsibility concept fails if developers develop a super application that is so clever it can deceive human beings entirely by continuing to function even if developers activate the pre-programmed off switch. The sanctions for such a superapp behaving in this way must be so severe that developers have every incentive to ensure it is impossible. This does not mean that such a super app will never be built, but the manager responsibility concept should ensure even if it is developed outside of regulated financial services (such as through cloud service providers), that should be an important consideration in its adoption and implementation within regulated financial services.

2. Overdeterrence

Manager responsibility could be too much of a good thing. If the regulatory burden deters managers from being involved in AI-based financial services, we may find only reckless and unreflective people developing AI for financial services and serving as senior managers for financial services firms, resulting overall in weaker, rather than better, governance. Regulators must respond to this concern with proportionate responses to apparently irresponsible conduct including into human contributions to failure such as which person failed to perform the AI due diligence or bypassed the explainability requirements. An initial assessment will take place in the context of managerial requirements, including fitness and properness. Personal responsibility / liability systems should also include frameworks of continuing education as well as ongoing fit and proper requirements in order to balance this risk.

Facing the choice between individual and collective responsibility, individual responsibility concepts could lead to less diligence in monitoring fellow key function holders. Collective responsibility, by contrast, could increase monitoring among key function holders, but lead to overdeterrence. This debate is live amidst the blame allocation arising from the ongoing Westpac scandal in Australia that is being attributed

[196] See *TechRisk*, supra n. 77.

to a relatively low key piece of software that led to allegedly some 23 million AML breaches.[197] A compromise would include defining some collective core duties while also imposing individual responsibility. This is clearly the case in both board responsibilities as well as corporate responsibility, both thereby putting in place collective responsibility as well as individual responsibility systems of internal governance via external regulatory requirements.

3. FinTech start-ups

A third concern relates to FinTech start-ups. Usually, regulators require experience and management skills in finance as a precondition for licensing a financial entity. Start-up staff often have little experience in running a regulated firm. If regulators require this expertise of all key function holders, innovation will be severely impaired.

The obvious response is for regulators to require sufficient expertise *and* experience from the start-up's board and key executives, *as a group*. Under this whole board and executive concept, some board members and executives can contribute the IT / AI expertise while others contribute their experience in running a regulated financial services firm. After a certain time in the business, 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 startup to take a balanced and proportionate approach to board and key executive requirements as a group, specific regulatorily mandated individual responsibility requirements, expertise and experience requirements would remain necessary as part of the licensing process.

V. Conclusion

The financial services industry is one of the leaders in the use and development of AI and going forward AI is likely to become an ever more important technology for financial services firms. However, AI comes with a number of very substantial technical, ethical and legal challenges that can undermine the objectives of financial regulation, from the standpoint of data, cybersecurity, systemic risk, and ethics, in particular in the context of black box issues.

As we have shown, traditional financial supervision focussed on external governance is generally unlikely to be highly effective in addressig the risks created by AI. This is because of three main regulatory challenges: (1) enhanced information asymmetry about the AI; (2) data dependency; and (3) interdependency with other AI. Accordingly, even where supervisory authorities have exceptional resources and expertise, supervising the use of AI in finance by traditional means of financial supervision is extremely challenging.

In order to address this weakness, we suggest that the internal governance of financial institutions be strengthened through imposing personal responsibility requirements to put a "human-in-the-loop", based on existing post-Crisis frameworks of managerial responsibility. These should ideally be cognisant of and consistent with broader data

[197] See Paul Smith, "Westpac's mess could happen to anyone" (Australian Financial Review, 6 Dec. 2019): < https://www.afr.com/technology/westpac-s-tech-mess-could-happen-to-anyone-20191204-p53gqq>.

privacy and human-in-the-loop approaches beyond finance.[198] From a financial authority's point of view, the strengthening of internal governance can be achieved, for the main part, through a renewed supervisory focus on senior managements' (or key function holders') personal responsibilities and accountability for regulated areas and activities for which they are designated responsible for regulatory purposes as well as key input from external AI experts and stakeholders. These key function holder rules, particularly if enhanced by specific due diligence and explainability requirements, will assist core staff of financial services firm to ensure that the AI under their control is performing in ways consistent with their personal responsibilities. If it is not, they will nonetheless be responsible. That is the nature of personal responsibility systems: the manager etc in charge is responsible for themselves, their area, their staff, their third party contractors, and their IT, including AI. This encourages – as a result of direct personal responsibility – due diligence in investigating new technologies, its uses and its impact and on requiring explainability systems as part of any AI system – or IT system for that matter. This is necessary from the standpoint of an individual who has potential direct responsibility in the event of a regulatory action for any failure: due diligence and explainability will be the key to a personal defence. Likewise, a similar approach would be incentivized in the context of regulatory use of AI: the necessity of defending any enforcement action in court requires due diligence in development and use of AI for regulatory purposes as well as explainability systems in order to defend their actions. While clearly effective in the black box context, this also addresses other data, cybersecurity, systemic risk, and ethical issues in the context of AI in finance, particularly when combined with centralized AI review committees to address issues of collective responsibility of the board and more broadly.

Importantly, this approach – while a natural evolution in the context of financial regulation – also has great potential for addressing AI concerns in other regulated industries through the regulatory requirement for "human-in-the-loop" personal human responsibility systems. While it does not necessarily address the macro issues which are emerging as a result of the Fourth Industrial Revolution, the digitization of everything, and AI, it does at least make sure that humans are centrally involved in the context of the evolution of AI in regulated industries, providing for personal understanding and responsibility to address many of the core micro issues, and puts us in a better position to understand the potential macro issues as they arise.

[198] See eg. Tang, *ibid*, n10.