AI risks, failures and consequences: Corporate governance for the AI era

Zofia Bednarz and Susan Bennett†*

Artificial intelligence (AI) tools can bring undeniable benefits for private sector organisations, however, they can also lead to significant harms. Focusing on the example of the financial industry, this article explores how these harms translate into risks for organisations in the context of AI applications that have direct implications for consumers. We ask what the use of such AI tools means in terms of risks for companies under current and emerging risk and governance frameworks, and what it means for directors discharging their duties under the Corporations Act 2001 (Cth) and general law. Drawing on practical examples and recent cases, we examine ways in which the risks of harm can arise and analyse the challenges organisations face when implementing AI governance systems to adequately reduce risks.

I Introduction

Artificial Intelligence (AI) is the buzzword of the moment. Since OpenAI's ChatGPT launched at the end of 2022, hardly a day passes by without some news about AI, its potential to transform business and our lives, its risks or upcoming regulatory interventions. Estimates of potential gains that AI will unlock in the near future continue to dazzle: generative AI (GenAI) could add up to $4.4 trillion in economic value to the global economy, according to the

* Lecturer, University of Sydney Law School; Associate Investigator, ARC Centre of Excellence for Automated Decision-Making and Society.

† FGIA, FIP. Principal of Sibenco Legal & Advisory and Founder and Director of InfoGovANZ. This article incorporates aspects of Dr Bennett's doctoral thesis, *Privacy and Data Protection: The Interaction of Meta-Regulation and Information Governance* (University of Sydney Law School, University of Sydney, 2023) <https://hdl.handle.net/2123/32008> (accessed 16 January 2025). The thesis includes an examination of the challenge of controlling personal information from the corporate governance perspective, where boards grapple with multifaceted strategic opportunities and risks arising from the intersection of technology, data and regulation. Effective information governance is shown to assist organisations with privacy meta-regulatory compliance and improve data management and information reliability across the organisation. The thesis argues information governance can also be used to improve information flows on nonfinancial risks to the board for good decision-making.

Earlier versions of this article were presented at the *Privacy Laws & Business Conference* (Cambridge, July 2023); *Let's Talk about Corporations Seminar* (University of Queensland and University of Sydney, August 2023); and *SCOLA 2024 Annual Conference* (University of New South Wales, 2024). The authors would like to thank the organisers of these events and the participants for the questions and feedback received, as well as anonymous reviewers for their very thoughtful and useful comments. The usual disclaimer applies.

management consulting firm McKinsey,[1] while some commentators consider the financial industry alone can expect AI to become a $130 billion industry by 2027.[2]

AI tools are opening up significant opportunities for financial sector organisations,[3] enabling them to slash costs and enhance operational efficiency. AI, including GenAI, and big data analytics are becoming a core capability for financial entities,[4] with the majority of financial institutions keenly exploring the benefits of the new technologies. Recent surveys indicate that as many as 80% are already implementing GenAI, and 86% of banks expect to be ready to go live with their AI use cases in 2024.[5] In this article, we focus mainly on AI uses that affect consumers of financial products and the implications this has for AI governance in financial sector entities.

Deploying AI solutions does not come without some significant risks to organisations, which can be commercial, reputational and/or legal in nature.[6] The narrative present in media around AI, and in particular GenAI, makes it look like a perfect solution for any organisation. In reality, deploying tools powered by new technologies requires a careful strategy and often a significant investment, not only in the technology, software and hardware necessary to run it but also in qualified staff, such as machine learning (ML) engineers, to maintain the operational environment. Failure to properly factor in all the costs, as well as risks, will inevitably lead to significant problems down the track.

Recent surveys show that in Australia, there is a sentiment of uncertainty in relation to AI use, going so far as corporate leaders believing that Australia is an 'AI-Wild West' since there is no AI-specific regulation.[7] This article counters that legal and regulatory frameworks that are currently in place already apply to AI governance in Australian corporations, including the financial industry.[8] No doubt the use of AI systems poses new legal and

---

1 O Bevan et al, 'Implementing Generative AI with Speed and Safety', *McKinsey & Company* (Web Article, 13 March 2024) <https://www.mckinsey.com/capabilities/risk-and-resilience/our-insights/implementing-generative-ai-with-speed-and-safety> (accessed 16 January 2025).

2 S Plummer, 'AI in Finance: Revolutionizing the Future of Financial Management', *datacamp* (Blog Post, 23 June 2024) <https://www.datacamp.com/blog/ai-in-finance> (accessed 16 January 2025).

3 This article focuses on Australian Prudential Regulation Authority (APRA)-regulated entities and financial sector 'corporations'/'companies' subject to the *Corporations Act 2001* (Cth) (*CA*) who are holders of an Australian financial services licence (AFSL). We use various terms to refer to the entities within the financial sector, including 'entities', 'institutions', 'organisations' and 'firms'. This includes private and public entities and statutory corporations that are regulated by APRA.

4 OECD, *Generative Artificial Intelligence in Finance* (OECD Artificial Intelligence Working Paper No 9, OECD Publishing, 2023) pp 13–18 <https://www.oecd.org/publications/generative-artificial-intelligence-in-finance-ac7149cc-en.htm> (accessed 16 January 2025).

5 IBM, *2024 Global Outlook for Banking and Financial Markets* <https://www.ibm.com/thought-leadership/institute-business-value/en-us/report/2024-banking-financial-markets-outlook> (accessed 16 January 2025).

6 L Solomon and N Davis, *The State of AI Governance in Australia* (2023) Report (Human Technology Institute, University of Technology Sydney) p 15 (*HTI Report*).

7 Above p 33.

8 As has been argued in the recent Australian legal literature on AI and financial services: see, eg, Z Bednarz and K Manwaring, 'Keeping the (Good) Faith: Implications of Emerging Technologies for Consumer Insurance Contracts' (2021) 43(4) *Sydney Law Review*, p 455

regulatory challenges, which are not always fully addressed in current regulations.[9] Still, abundant requirements currently apply to AI governance. In the submission to the Responsible AI Consultation, the Australian Banking Association pointed out that current risk governance frameworks 'consider many of the operational and technology risks that arise in non-AI contexts which equally apply in the implementation of AI tools'.[10] Robust information and AI governance can prepare financial entities to comply with current laws and emerging AI regulation, future-proof their AI technology solutions and deal with the risks arising from the use of AI and particularly, AI systems using personal information.

This article proceeds as follows. In Section II, we examine the risks of AI use by financial organisations beginning in II(A) with a brief recap of AI applications in the financial sector with the focus on consumer-facing applications. In II(B), we identify the main harms and risks the technology use poses. In Section III, we provide an overview of the legal and regulatory frameworks currently in place and applicable in relation to the use of AI tools. In Section IV, we analyse the relevant requirements placed on the financial sector entities, starting in IV(A) with the obligations placed on boards in the context of AI risk management arising from the statute (*Corporations Act 2001* (Cth) (*CA*)), as well as the regulatory guidance and standards. In IV(A)(1), we examine the obligation of boards to adequately monitor AI risks in accordance with the regulatory guidance and standard. In IV(A)(2), we then turn to consider the responsibility of the financial entity under s 912A(h) of the *CA* and the responsibility of directors for breaches of directors' duties under s 180(1) of the *CA*. Next, in IV(B), we examine the challenges of implementing a robust AI governance framework, analysing AI governance standards and frameworks (IV(B)(1)); the consequences of the technology integration failures (IV(B)(2)); the question of data and information governance challenges (IV(B)(3)); and the impact of the organisational culture on AI governance (IV(B)(4)). In Section V, we offer brief conclusions.

(Keeping the (Good) Faith); Z Bednarz and K Manwaring, 'Hidden depths: The effects of extrinsic data collection on consumer insurance contracts' (2022) 45(July) *Computer Law & Security Review*, 105667: p 1; Z Bednarz and K Manwaring, 'Insurance, artificial intelligence and big data: Can provisions of ch 7 of the *Corporations Act* help address regulatory challenges brought about by new technologies?' (2021) 36(3) *Australian Journal of Corporate Law*, p 216 (Insurance, artificial intelligence and big data).

9 Above, see also K Weatherall et al, *ADM+S Submission to the Safe and Responsible AI in Australia Discussion Paper* (ARC Centre of Excellence on Automated Decision-Making and Society, 2023) pp 14–24 <https://consult.industry.gov.au/supporting-responsible-ai/submission/view/437> (accessed 16 January 2025).

10 R Luo, *ABA Submission — Safe and Responsible AI in Australia Discussion Paper* (Australian Banking Association, Sydney, 2023) p 2 <https://consult.industry.gov.au/supporting-responsible-ai/submission/view/498> (accessed 16 January 2025).

II AI risks

A Overview of AI applications in the financial sector

The use of AI tools leads to novel risks for organisations. First, this is because AI models are a subtype of automated processes, which can be characterised as:

> process[es] or procedure[s] performed with minimal human assistance, operating as a scale from full automation (no humans involved) to minimal automation (components of a process otherwise undertaken by a human are now undertaken by computer or robots).[11]

The failures of the Robodebt scheme provide an illustration of the risks related to automation.[12] The technology used in Robodebt is not considered AI: the Robodebt Royal Commission Report indicates that while:

> [t]he terms ... 'automated decision making' and ... 'AI' ... have been [used interchangeably] in commentary about the Scheme, ... they are distinct concepts. AI describes systems that are self learning and have a level of autonomy.
>
> ...
>
> While the Scheme did not use AI, it did use automaton. It involved a system of business rules with no ability to move outside of specific and defined action on the basis of the data received. It was extremely rigid; once the rules had been coded and set in place, the system itself would stay in place until the rules were changed by way of human interventon.[13]

Robodebt failures raised many of the issues of concern also linked to the widespread adoption of AI.

Second, AI systems go beyond automation, as these models allow machines to perform tasks requiring human-like 'thinking' or 'learning'. As Guihot and Bennett Moses point out:

> AI has been described as computers doing things that, if a human performed them, would require intelligence: learning, understanding language and seeing; perceiving or interacting with external environments.[14]

In the financial services context, a similar definition was accepted by the UK's Financial Conduct Authority (FCA) and the Bank of England (BoE), pointing to AI being 'the theory and development of computer systems able to perform tasks which previously required human intelligence'.[15]

---

11 M Guihot and L Bennett Moses, *Artificial Intelligence, Robots and the Law* (LexisNexis Butterworths, 2020) p 10.

12 'Robodebt' refers to an automated assessment and recovery of overpayments from welfare recipients operating in Australia from 2015 to 2019, which produced inaccurate results and was unlawful, resulting in catastrophic harms suffered by individuals subject to it: see Royal Commission, *Royal Commission into the Robodebt Scheme* (vol 1, 2023) Report (Royal Commission, Canberra).

13 Above p 472.

14 Guihot and Bennett Moses (n 11) p 14.

15 C Jung et al, *Machine Learning in UK Financial Services* (2019) Report (BoE and FCA) <https://www.bankofengland.co.uk/-/media/boe/files/report/2019/machine-learning-in-uk-financial-services.pdf> (accessed 16 January 2025).

Other AI definitions focus on systems that perform certain tasks against a set of objectives provided by a human. For example, the International Organization for Standardization (ISO) defines AI as:

> a technical and scientific field devoted to the engineered system that generates outputs such as content, forecasts, recommendations or decisions for a given set of human-defined objectives.[16]

For the purpose of the discussion in this article, we follow the definitions outlined and we focus our analysis on deployment of technologies that allow for new ways of doing things, including 'learning' and 'thinking', potentially leading to autonomous decision-making, as well as using new sources of data about people, obtaining new and more detailed insights from that data and doing any of these at a greater scale than ever before.

As a technology, AI encompasses a range of tools, techniques and application areas from computer vision to ML (large language models (LLMs), also referred to as GenAI, are an example of ML) to natural language processing (NLP).

There are many AI — and more broadly automation — use cases in the financial industry, with the great majority of financial firms already using AI in multiple ways.[17] Applications range from customer engagement to risk management and compliance. Uses that have the potential of directly affecting consumers include credit scoring and lending, customer- personalised product recommendations, underwriting of insurance contracts, chatbots or fraud detection.[18]

B Harms and risks flowing from AI use by financial sector entities

There is an increasing interest in how AI tools could be used in the decision-making processes of the board,[19] however, such considerations, especially focussing on AI potentially replacing human directors, remain (for now) largely theoretical. This article's main concern is what the use of AI tools, and especially tools directly affecting consumers, means in terms of risks for companies under current and emerging risk and governance frameworks. Furthermore, we explore the implications for directors discharging their duties under the *CA* and general law. In this context, it needs to be noted that an organisation defines its own purpose and strategic objectives and while there are strong motivations for implementing AI systems — to improve efficiencies, competitiveness and deliver returns to the

16 *Information Technology — Artificial Intelligence — Artificial Intelligence Concepts and Terminology* (1st ed, ISO/IEC 22989:2022, July 2022); see also, eg, *Regulation (EU) 2024/1689 of the European Parliament and of the Council laying down harmonised rules on artificial intelligence (AI Act)*, Art 3(1); OECD, *Recommendation of the Council on Artificial Intelligence* (OECD/LEGAL/0449, 3 May 2024).

17 See, eg, Jung et al (n 15) pp 9–11.

18 See, eg,: OECD (n 4) pp 13–18; European Insurance and Occupational Pensions Authority (EIOPA), *Big Data Analytics in Motor and Health Insurance: A Thematic Review* (2019) Report (EIOPA, Luxembourg) pp 29–41.

19 M Hilb, 'Toward artificial governance? The role of artificial intelligence in shaping the future of corporate governance' (2020) 24 *Journal of Management and Governance*, p 851.

bottom line — there is no positive duty to use any form of technology in the business practice,[20] even if AI tools can offer outcomes that are statistically superior to other possible solutions. The anticipated potential opportunities for efficiencies and profits need to be balanced against risks, which are not insignificant, for organisations and directors arising from AI. *HTI Report* points out that:

> [w]ithout proper governance, the rapid deployment of AI systems exposes organisations, employees, consumers, and the broader community to severe harms and significant risks. Organisations responsible for AI-related harms will be exposed to commercial, regulatory, and reputational risks.[21]

AI and automated systems can harm individuals and groups, with harms potentially affecting all aspects of social life, from work environments to healthcare, consumer experience and politics.[22] The harms, which can impact individuals, such as employees and consumers, but also organisations, and risks facing the corporations that use AI tools[23] are two sides of the same coin.

Sources of harms brought about by use of AI systems can be categorised in various ways.[24] For the purpose of this article, it is useful to distinguish between harms originating in the AI model itself and harms stemming from the way in which the system is used. While both types of issues can occur simultaneously, organisations using the AI tools in question will need to specifically consider the risks flowing from these two sources of potential harms.

Harms that originate in the AI model itself can be either related to the system's performance or to its inherent failures or fragilities.[25] The model itself can be biased or it may not be performing the task it was designed to do to a satisfactory level.[26] Furthermore, the system's fragility can lead to its crashing,[27] in addition to security failures exposing data held by the organisation deploying the model. The way in which those risks are addressed by the organisation will depend on the source of the system, that is, whether it is provided by a third party, or built in-house and who is responsible for its maintenance.

---

20 Cf A Godwin, 'A Duty to Use Artificial Intelligence? Learning from the Past and Hedging for the Future' in R T Langford (Ed), *Corporate Law and Governance in the 21st Century: Essays in Honour of Professor Ian Ramsay* (Federation Press, 2023) p 174, 174–193.

21 *HTI Report* (n 6) p 15.

22 D Acemoglu, 'Harms of AI' in J B Bullock et al (Eds), *The Oxford Handbook of AI Governance* (Oxford University Press, 2022) p 660, 660–706.

23 M E Kaminski, 'Regulating the Risks of AI' (2023) 103 *Boston University Law Review*, p 1347, 1357.

24 See, eg, a useful categorisation in the *HTI Report* (n 6) p 16.

25 Above pp 16–17.

26 For more detailed explanation regarding AI bias, see (and the literature cited there) Bednarz and Manwaring, 'Keeping the (Good) Faith' (n 8) pp 466–470.

27 Crashing is a relatively common problem with home smart devices that use AI systems, which often freeze and are vulnerable to power or network outages, leading to device malfunctioning or not rebooting correctly: see W He et al, 'When Smart Devices Are Stupid: Negative Experiences Using Home Smart Devices' [2019] *IEEE Security and Privacy Workshops*, p 150 <https://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=8844630&casa_token=YfIf1IBb5D0AAAAA:GGojUzFFA4ELi1etbKqvjtfd9UxzyunNhJMwcG1I7s8tcjEE65DbQIXx4VO_dYm4Is-Lsa9ZuQ> (accessed 16 January 2025).

The second group of harms relates to how an AI system is used. The potential issues here can be usefully presented under two headings proposed in *HTI Report*: '[m]alicious or misleading deployment'; and '[o]veruse, inappropriate or reckless use'.[28] Malicious or misleading deployment, for purposes of this article, mainly refers to AI systems that produce outputs that are either legally, or at least ethically, unfair and potentially misleading.[29] While such outcomes can be a result of some inherent flaws within the particular AI system itself, the focus here is on how the model is used. 'Price optimisation' practices in the insurance industry, consisting of charging different premiums from clients who are similar in terms of risk and costs, are an example. An AI model could be used to achieve such 'price optimisation', for example, through analysing a consumer's willingness to pay — or likelihood they will pay a certain price, rather than the real underwriting risk they present.[30] Such a use of an AI system by an organisation is intentional. While the organisation in question may believe they are complying with the legal and regulatory framework in place, regulators are increasingly likely to take action for similar practices, as the recent case brought by Australian Securities and Investment Commission (ASIC) against Insurance Australia Ltd shows.[31]

The second category: '[o]veruse, inappropriate or reckless use' includes a wide range of societal, economic and environmental harms.[32] This category includes situations where AI tools are used even though they are not needed, for example, because other solutions exist, which are less harmful for consumers. The Air Canada chatbot case below provides a good illustration. From the perspective of the organisation using the system, the harm resulting from overuse or inappropriate use of AI tools may not be intentional, however, it can be viewed as a by-product of the practices that an AI system enabled.

Overcollection of data linked to the use of AI is a related issue. Due to the very nature of AI systems, which can be described as 'data-hungry',[33] overcollection of data which is not strictly necessary to provide a service, but rather data gathered with a view of downstream use (such as training of an AI model) can also fit in this broad category of harms.

The harms outlined above and the risks that organisations implementing AI face are interrelated. The risks have been described as trifold: commercial,

28 *HTI Report* (n 6) p 16.

29 Note that the *HTI Report* also considers other malicious uses of AI systems, such as AI-powered cyberattacks or weaponisation of AI systems (see above). These issues, while no doubt important, are outside of the scope of our analysis since our focus is on the risks that financial organisations that use AI tools face.

30 EIOPA (n 18) p 39; FCA (UK), *General Insurance Pricing Practices: Final Report* (2020) Market Study No MS18/1.3 (FCA, London) p 21; see also the ongoing action Australian Securities and Investment Commission (ASIC) brought against an insurer for such practices: *ASIC v Insurance Australia Ltd ACN 000 016 722* (Federal Court of Australia, commenced 21 December 2023) (Amended Concise Statement).

31 Above.

32 *HTI Report* (n 6) p 16.

33 R Jesse McWaters, *The New Physics of Financial Services: Understanding How Artificial Intelligence is Transforming the Financial Ecosystem* (2018) Report (World Economic Forum) p 42.

regulatory and reputational;[34] and are usually present at the same time to a varying degree. A recent Canadian case provides a good illustration: Air Canada was ordered to compensate a customer who was given inaccurate information by the airline's chatbot.[35] The customer was misled into purchasing a full-price airfare, while they were eligible for a discounted ticket under the company bereavement fare policy, as they were travelling to attend a funeral. Air Canada tried arguing that it could not be held liable for information provided by the chatbot,[36] and that correct information regarding the airfare in question was available on the website. The Court found, however, that the chatbot's advice constituted misrepresentation. The compensation amounted to CAN$812.02.[37] The risk that materialised for the business in this case does not seem like a big one legally, or indeed commercially, but it could be quite significant in terms of the company's reputation. Further, in the financial industry context, the result of misleading advice provided by a chatbot could be much more damaging.

The AI risks that organisations need to deal with arise from planning and procurement through to implementation and then throughout its ongoing use. A report by the BoE describes three key stages of the AI lifecycle: data; models; and governance.[38] Governance is key to ensuring risks are properly addressed, both at the AI system level and the overall corporate governance of the organisation. Furthermore, the success of the AI system relies on the quality of the data used. This requires robust governance of data — from the training (or even pre-training) and refinement stages of the AI model to reinforcement learning, which can be achieved with human feedback[39] and ongoing governance of the AI system throughout its use. Different risks that the organisations must consider will arise at the subsequent stages of the AI models development and use, including model drift.[40]

The challenges of implementing AI to comply with current legal regulations have been revealed by a set of recent class action lawsuits in the US, providing an illustration of (potential) harms resulting from (alleged) AI system failures or misuses, affecting consumers of financial services and corresponding risks associated with the use of AI tools by financial sector entities.

The *Huskey v State Farm Fire & Casualty Co*[41] class action lawsuit alleges discrimination against Black homeowners in relation to home insurance products provided by State Farm. In the statement of claim, it is argued that

34 *HTI Report* (n 6) p 18.

35 *Moffatt v Air Canada* (2024) BCCRT 149 (CanLII).

36 Above at [27].

37 Above at [44].

38 BoE, *Artificial Intelligence and Machine Learning* (2022) Discussion Paper No DP5/22 (BoE, London).

39 Weatherall et al (n 9) p 50.

40 'Model drift refers to the degradation of machine learning model performance due to changes in data or in the relationships between input and output variables. Model drift — also known as model decay — can negatively impact model performance, resulting in faulty decision-making and bad predictions.': see J Holdsworth, I Belcic and C Stryker, 'What is Model Drift?', *IBM* (Webpage, 16 July 2024) <https://www.ibm.com/topics/model-drift> (accessed 16 January 2025).

41 *Huskey v State Farm Fire & Casualty Co* (ND Ill, Case No 22-cv-7014, commenced 14 December 2022) (Statement of Claim) (*State Farm*).

Black policyholders' claims were being subjected to greater suspicion and higher scrutiny than those for white homeowners, resulting in the Black policyholders having to submit extra paperwork, needing to interact more with the insurer and being less likely to have their claim processed in less than a month.[42] The reason for this — it is alleged — is the use of automated claims processing methods and ML algorithms by the insurer, that:

> make predictions and decisions about whether a claim might be fraudulent, how much scrutiny it requires, and how it should be processed, and in doing so, rely on (1) biometric data that function as proxies for race, such as physical appearance, genetics, and voice; (2) intrusive behavioral data that function as proxies for race, such as geolocation, social media presence, and browser search history; and (3) historical housing and claims data that are themselves infected with racial bias.[43]

The tools State Farm uses for predictive fraud modelling are provided by third parties, including the FRISS software that delivers real-time fraud assessment on every claim, assigning a risk score based on both internal insurers' data and external data, and applies techniques, such as AI text mining to 'search for questionable words and phrases and search for suspect language patterns'.[44] Automating fraud detection in insurance claims no doubt provides significant efficiency gains and thus profits to the insurer, however, it does not come without risks.

Bias and resulting discrimination are also a data governance problem. For example, an AI model that is learning continually[45] may create feedback loops. Such a model, while facially neutral, may perpetuate the existing social inequalities, and also amplify them,[46] potentially in an unjustified way (like the detection of 'questionable word patterns' mentioned). Such feedback loop amplification is argued in the lawsuit analysed. The plaintiffs allege that the:

> [b]iased antifraud algorithms become self-[fulfilling] — if there is racial bias in the claims that are identified as potential[ly] fraudulent and then investigated, there will be racial bias in the claims which are found to be fraudulent. Because you won't find fraud in a claim you don't investigate.[47]

Another recent class action lawsuit, *Stephens v Tesla Insurance Services, Inc*,[48] provides examples of potential harms caused by AI systems due to inaccurate input data. The plaintiffs argue that Tesla insurance is charging inflated premiums based on faulty sensors. The issue lies in the usage-based insurance offered, which uses 'safety factors' collected directly from the policyholder's Tesla vehicle to generate a 'safety score', which helps calculate the premium. It is alleged that those 'safety scores' are inflated — resulting in

---

42 Above at [3], [4], [21]–[25].

43 Above at [4].

44 Above at [37]–[40].

45 E Verwimp et al, 'Continual Learning: Applications and the Road Forward' (2023) *arXiv* <https://arxiv.org/abs/2311.11908> (accessed 16 January 2025).

46 Such mechanism has been documented, eg, in an AI model used to assess visa applications in the UK: 'Home Office drops "racist" algorithm from visa decisions', *BBC* (Online, 4 August 2020) <https://www.bbc.com/news/technology-53650758> (accessed 16 January 2025).

47 *State Farm* (n 41) at [44].

48 *Stephens v Tesla Insurance Services, Inc* (Cal, Case No 22CV031800, commenced 10 April 2023) (Statement of Claim).

higher premiums — because of random 'ghost' forward collision warnings (which is one of the safety factors mentioned) when there is no actual danger or any car in sight.[49] As a result, the insurer may charge higher premiums based on these 'unsafe' driving events that never actually occurred.

III General overview of currently applicable frameworks

While Australia currently lacks an AI-specific regulation, the use of AI within the financial sector is subject to a range of existing regulations. The Australian legal system is — to a great extent — technologically neutral and financial entities are subject to a legal framework that imposes obligations that apply independently of the type of technology or other decision-making processes used. The question of legal and regulatory gaps in terms of addressing new and emerging impacts of AI[50] is outside the scope of this article: we focus on the operation of current rules in relation to the current use of AI tools by financial firms. It should be noted that robust AI and information governance[51] will not only mitigate the current reputational, commercial and regulatory risks linked to the organisation's use of AI tools but will also prepare them to comply with future AI regulation.

In terms of existing laws and regulations, *HTI Report* usefully points to six broad areas of most relevance: privacy and data protection regime; consumer protection; cyber security; anti-discrimination laws; duty of care and negligence; and work health and safety.[52] These rules set out legal obligations that Australian corporations, including financial sector entities, have to comply with when using AI systems. This is in addition to corporate law requirements, as well as sector-specific regulations. It remains to be seen how onerous potential new rules regulating AI will be on top of all these obligations.[53] In the following section, we focus our analysis on requirements relevant to AI governance that financial sector entities need to consider, followed by the challenges of AI governance that organisations are currently facing.

IV Selected requirements in the context of the financial industry

A Directors' duties and risk management systems

1 Overview of regulatory requirements for adequate risk management systems

Under Australia's 'twin peaks', financial regulation vests in two separate regulators: ASIC, responsible for market conduct regulation and consumer protection, and the Australian Prudential Regulation Authority (APRA),

49 Above at [5]–[6].
50 See Weatherall et al (n 9) pp 11–24 for a detailed discussion of these gaps.
51 See Section IV(B).
52 *HTI Report* (n 6) pp 38–48.
53 Which is no doubt a concern that financial entities have: see, eg, Luo (n 10).

entrusted with the maintenance of the financial system's stability.[54] Under both these frameworks, there are regulatory requirements for boards of directors to ensure adequate risk management systems are implemented and monitored in relation to technology, including AI, and personal information used within AI systems. Where risk management systems are inadequate and/or not properly monitored by the board to ensure they comply with regulatory obligations, directors are exposed to potential regulatory action and other legal risks such as class actions.

Directors are responsible for overseeing and monitoring the management of a company's affairs, including compliance with all applicable laws. Both in general law and under statute, directors and officers have a duty to act with care and diligence.[55] Before taking a closer look at how directors' duties operate in the context of AI harms, we first consider the legal and regulatory requirements relevant to AI risk management.

In the context of the market conduct regulation overseen by ASIC, a number of requirements stand out, mainly under the *CA*, in addition to *ASX Australian Corporate Governance Principles* (*CGPs*).[56] A financial sector entity, as a holder of an Australian financial services licence (AFSL) pursuant to s 912A(1)(a) of the *CA* is required to do all things necessary to ensure the financial services covered by the licence are provided efficiently, honestly and fairly,[57] and to have adequate risk management systems in place (*CA* s 912A(1)(h)).

Listed companies should maintain a risk management framework pursuant to the *CGPs*, which are an annexure to the *ASX Listing Rules*.[58] The *ASX Listing Rules* require financial sector entities established in Australia to provide a copy of their annual report to the ASX and their security holders.[59] The annual report must include a corporate governance statement that discloses the extent to which the entity has followed the recommendations contained in the *CGPs*.[60] There are eight core *CGPs*; the seventh principle, which is to 'recognise and manage risk', states:

> A listed entity should establish a sound risk management framework and periodically review the effectiveness of that framework.

---

54 A Godwin and A Schmulow, 'Introduction: The Genealogy and Topography of Twin Peaks' in A Godwin and A Schmulow (Eds), *The Cambridge Handbook of Twin Peaks Financial Regulation* (Cambridge University Press, 2021) p 1.

55 *CA* s 180(1).

56 ASX Corporate Governance Council, *Corporate Governance Principles and Recommendation*s (4th ed, 2019) (ASX Corporate Governance Council) <https://www.asx.com.au/content/dam/asx/about/corporate-governance-council/cgc-principles-and-recommendations-fourth-edn.pdf> (accessed 16 January 2025) (*CGPs*).

57 For more detailed discussion of the 'efficiently, honestly and fairly' requirement in the context of AI, see Bednarz and Manwaring, 'Insurance, artificial intelligence and big data' (n 8) pp 229–232.

58 *ASX Listing Rules* Appendix 4G <https://www.asx.com.au/regulation/rules/asx-listing-rules.htm> (accessed 20 January 2025).

59 *ASX Listing Rules* 4.5, 4.7 <https://www.asx.com.au/regulation/rules/asx-listing-rules.htm> (accessed 20 January 2025).

60 *ASX Listing Rules* 4.10.3 <https://www.asx.com.au/regulation/rules/asx-listing-rules.htm> (accessed 20 January 2025).

The 'degree of care and diligence' required from directors in the exercise of their powers and the discharge of their duties under s 180(1) of the *CA*[61] in respect of AI use within an entity is informed by the *CGPs*.[62] Kingsford Smith points to the approaches taken by Owen J in *Bell Group Ltd (in liq) v Westpac Banking Corp* and Middleton J in *ASIC v Healey* following the approach of Austin J in *ASIC v Rich*, where 'corporate governance literature', including the *CGPs*, were considered relevant to determining liability and penalties.[63] More recently, Beach J referred to the 'regulatory guidance' provided by the *CGPs* in *ASIC v Mitchell (No 2)*.[64]

The commentary accompanying the *CGPs* sets out that '[o]ne of the key roles of the board of a listed entity is to monitor the adequacy of the entity's risk management framework' and that, 'this includes satisfying itself that the risk management framework deals adequately with contemporary and emerging risks, such as conduct risk, digital disruption, cybersecurity, privacy and data breaches, sustainability, and climate change'.[65] Clearly, AI, and particularly those systems that use personal information, is both an emerging and contemporary risk that needs to be adequately managed in order for directors to discharge their obligations under the *CA*.

The same is true for the prudential regulation. APRA sets out enforceable Prudential Standards, stating the minimum requirements entities are required to implement.[66] These standards cover five main areas: governance, risk management, financial resilience/business operations, resolution and reporting. In respect of AI governance, there are several relevant Prudential Standards including: *CPS 510 — Governance*; *CPS 220 — Risk Management*; *CPS 231 — Outsourcing*; *CPS 234 — Information Security*; and *CPG 235 — Managing Data Risk*.

CPS 510 sets out that 'the ultimate responsibility for oversight of the sound and prudent management of an APRA-regulated institution rests with its board of directors'.[67] *CPS 220* requires a risk management framework to be maintained and 'appropriately developed' to manage different types of 'materials risks' to ensure an entity meets its obligations.[68] Further, it makes clear that the board is 'ultimately responsible for the institution's risk management framework and for the oversight of its operation by

61 See below Section IV(A)(2).

62 See *CGPs* (n 56) Principle 7.

63 D Kingsford Smith, 'Governing the Corporation: The Role of "Soft Regulation"' (2012) 35(1) *University of New South Wales Law Journal*, p 378, 385–393; *Bell Group Ltd (in liq) v Westpac Banking Corp (No 9)* (2008) 39 WAR 1; (2008) 70 ACSR 1; [2008] WASC 239; *ASIC v Healey* (2011) 196 FCR 291; (2011) 278 ALR 618; [2011] FCA 717; *ASIC v Rich* (2003) 44 ACSR 341; (2003) 174 FLR 128; [2003] NSWSC 85.

64 *ASIC v Mitchell (No 2)* (2020) 382 ALR 425; (2020) 146 ACSR 328; [2020] FCA 1098 at [1400].

65 *CGPs* (n 56) p 27.

66 See *Australian Prudential Regulation Authority Act 1998* (Cth) s 8(1)(c), which empowers APRA to develop the administrative practices and procedures to be applied in performing its regulatory role and administration.

67 APRA, *Prudential Standard CPS 510 — Governance* (2019) (APRA) p 1.

68 APRA, *Prudential Standard CPS 220 — Risk Management* (2017) (APRA) cl 19 (*CPS 220*).

management'.[69] *CPS 220* also sets out a number of minimum requirements of a risk management framework, including a risk appetite statement and a specific requirement for 'a management information system(s) that is adequate, both under normal circumstances and in periods of stress, for measuring, assessing and reporting on all material risks across the institution'.[70]

The other relevant CPS standards underscore that APRA-regulated entities require boards to approve the outsourcing policy, 'which must set out the approach to outsourcing material business activities including a detailed framework for managing all such outsourcing arrangements'.[71] A 'material business activity' is defined as 'one that has the potential, if disrupted, to significantly impact the APRA-regulated institution's or group's business operations or its ability to manage risks effectively'.[72] This is determined by reference to a number of listed factors, including 'potential losses to the APRA-regulated institution's or group's customers and other affected parties in the event of a service provider failure'.[73] Therefore, in accordance with the requirements of the applicable CPS standards, boards need to ensure that all outsourcing of material business activities, which encompasses third-party AI systems, are subject to appropriate due diligence and that all risks arising must be appropriately managed;[74] that the outsourcing agreement address, inter alia, confidentiality, privacy and security of information;[75] that there is effective governance of data risk management in accordance with corporate governance frameworks;[76] and that the board is ultimately responsible for the entity's information security.[77]

2 AI harms: Responsibility of the entity and directors

The *CGPs* and the Prudential Standards make clear that boards are required to ensure that risk management frameworks are adequate to address risks arising from the use of emerging technologies, such as AI. As referred to above, s 912A(h) of the *CA* requires financial sector entities to have an adequate risk management system in place. ASIC has already successfully secured declarations against a financial sector entity where it was found that the cybersecurity systems in place were inadequate to protect sensitive personal information. In *ASIC v RI Advice Group Pty Ltd* (*RI Advice*),[78] action was taken against RI Advice, an AFSL holder, pursuant to ss 912A(1)(a) and 912A(1)(h) of the *CA*. There had been nine cybersecurity incidents. Prior to the sixth and most serious cyber incident involving a malicious attack in which unauthorised access was obtained to several thousand clients' personal information, RI Advice admitted that 'it did not have documentation, controls, and risk management systems that were adequate to manage risk in respect of

69 Above cl 9.
70 Above cl 23.
71 APRA, *Prudential Standard CPS 231 — Outsourcing* (2017) (APRA) (*CPS 231*).
72 Above cl 14.
73 Above cl 14(e).
74 *CPS 231* (n 71).
75 Above cl 29(j).
76 APRA, *Prudential Practice Guide: CPG 235 — Managing Data Risk* (2013) (APRA) cl 5.
77 APRA, *Prudential Standard CPS 234 — Information Security* (2019) (APRA) cl 13.
78 *ASIC v RI Advice Group Pty Ltd* (2022) 160 ACSR 204; [2022] FCA 496.

cybersecurity across its AR network'.[79] The unauthorised access occurred over a period of several months and resulted in the unauthorised use of personal information as notified by some clients. While RI Advice had subsequently developed measures to improve cybersecurity and cyber resilience, it admitted 'it took too long to implement and ensure such measures were in place'.[80]

Justice Rofe declared, in accordance with RI Advice's admissions, that the company had failed to ensure adequate cybersecurity measures were in place and/or adequately implemented. By failing to implement adequate cybersecurity and cyber resilience measures, it exposed its clients to an unacceptable level of risk.[81] The *RI Advice* case provides a salient warning to financial sector entities that measures need to be implemented swiftly, and steps must be taken to ensure that meta-regulatory requirements referring to 'adequate risks management systems' or 'adequate privacy and information security measures' are maintained.

In addition to the entity, its directors can also be found liable where AI caused harms. Under s 180(1) of the *CA*, a director or officer can be held liable for a corporate governance failure, if they fail to exercise care and diligence when there is a reasonably foreseeable risk of harm to the interests of the corporation. This formulation by Ipp J in *Vrisakis v ASIC*[82] was applied by Edelman J in *ASIC v Cassimatis (No 8)*,[83] who made clear that '[it] includes harm to *all* interests of the corporation' and stated:

> The interests of the corporation, including its reputation, includes its interests which relate to compliance with the law.[84]

In the Full Federal Court decision, Greenwood J, referring to the 'important observations' of the trial judge, noted that 'no proof of actual loss to the corporation is required by s 180(1) and harm to its interests, including "reputation," might also accrue without prospective loss'.[85] The reference to harm to the corporation's reputation is particularly important in the context of AI harms. In some of the examples discussed in Section II, the actual, quantifiable loss is likely relatively small for each individual harmed. However, damage to the reputation of the company may be significant, including adverse media and publicity, and may result in a loss of trust by customers, shareholders and other stakeholders.

Hanrahan and Yates identified three types of compliance failures in relation to the duty to monitor: a complete failure by the board to monitor, a failure to monitor after establishing a compliance system, and a failure by the board to respond adequately to '"red flags" raised by the systems and controls'.[86] This

79 Above at [62].

80 Above at [64].

81 Above at [65]–[66].

82 *Vrisakis v ASIC* (1993) 9 WAR 395 at 449–450; (1993) 11 ACSR 162 at 212; (1993) 11 ACLC 763.

83 *ASIC v Cassimatis (No 8)* (2016) 336 ALR 209; [2016] FCA 1023.

84 Above at [483].

85 *Cassimatis v ASIC* (2020) 275 FCR 533; (2020) 376 ALR 261; [2020] FCAFC 52 at [85], [87]–[88] (citations omitted).

86 P Hanrahan and R Yates, 'Directors' duties of oversight: Insights for Australia from recent developments in Delaware's *Caremark* jurisprudence' (2018) 33(2) *Australian Journal of Corporate Law*, p 185, 192.

would include scenarios where a director 'made a positive decision not to respond to a red flag, for example, because the likelihood and consequences of a breach [of the company's regulatory obligations] being detected would be less than the benefit derived by the company from the conduct', which would be a breach of the duty of loyalty and s 181 of the *CA*, as well as of the duty of care and s 180(1).[87]

In the context of organisations that have in place risk management frameworks, compliance programs and reporting systems to committees of the board and the board, oversight failures in relation to the use of new technologies will more likely include, for example:

- The failure to properly monitor data-driven technology systems, which may result in (even unintended) regulatory breaches, such as a failure of cybersecurity resulting in data breach, or implementation of a new technology or compliance system, or the integration of both, that leads to regulatory non-compliance; and
- Failures to respond to medium (amber) and high-risk (red) flags in board reports, for example, medium and high-risk ratings against regulatory compliance initiatives (eg, privacy compliance) or information security (eg, identification and mapping of personal data within an organisation or cybersecurity measures).

A question arises as to whether inaccurate and/or incomplete reporting to the board, leading to inadequate governance over technology systems, may give rise to a successful action against directors and officers. It is clear that the assessment of the director's conduct against required standards can take into account 'how the company's business has been distributed between the directors and the company's employees, the informational flows and systems in place and the reporting requirements within the company'.[88] As a result, if the risk management system and reporting system is deemed to be adequate, the failure of management to adequately report risks may be relied upon by the relevant directors and officers to successfully argue that they had discharged their obligations under s 180(1). For example, the lack of information flow to the Westpac board appears to have been viewed as acceptable by an external advisory panel review into Westpac's board governance following the *Chief Executive Officer of the Australian Transaction Reports and Analysis Centre v Westpac Banking Corp* judgment,[89] which concluded:

> the processes of the Board were adequate and its receipt of information, and the timing of that information, were also adequate. What failed was that the information provided by management was sometimes misleading or omitted. What was not known by management could not be provided.[90]

---

87 Above p 207, referring to Robson J in *ASIC v Flugge* (2016) 342 ALR 1; (2016) 119 ACSR 1; [2016] VSC 779.

88 *ASIC v Mariner Corp Ltd* (2015) 241 FCR 502; (2015) 327 ALR 95; [2015] FCA 589 at [441].

89 *Chief Executive Officer of the Australian Transaction Reports and Analysis Centre v Westpac Banking Corp* (2020) 148 ACSR 247; [2020] FCA 1538 (*AUSTRAC v Westpac*).

90 Westpac Group, *Board Governance of AML/CTF Obligations at Westpac: The Advisory Panel Review* (2020) Report (Westpack Group, Sydney) p 30 <https://www.westpac.com.au/content/dam/public/wbc/documents/pdf/aw/media/westpac-releases-findings-into-

However, as the board is ultimately responsible for ensuring it has the relevant information, it can also be argued that a failure of management to provide accurate and relevant information is also a failure of the directors. As previously noted, *CPS 220* requires a management information system(s) for measuring, assessing and reporting on all material risks across the institution.[91]

Hanrahan and Bednall point out '[t]he vulnerability of the oversight governance model is, of course, in the quality of the information flowing to the board'.[92] Contravention of s 912A(1)(h) of the *CA* (which requires AFSL holders to have adequate risk management systems in place) will not automatically result in directors' liability under s 180(1) of the *CA*, which would be the consequence of the application of the 'two-tiered' stepping stones liability.[93] Rather, as Langford points out, the Court in *Cassimatis* considered 'the breach or potential breach of the law by the company as part of the corporation's circumstances referred to in s 180(1) [of the *CA*]'.[94] It is open to ASIC to take action against directors for oversight failures in corporate governance arising from inadequate information systems and/or regulatory breaches, including where there is found to be an inadequate risk management system in place to properly manage AI and related risks and a contravention of s 912A(1)(h) of the *CA*.

Such oversight failures to properly monitor technology and information systems are evident in the AUSTRAC action against Westpac.[95] The Advisory Panel Report, in particular, identified Westpac board's lack of prompt action on red flags as inadequate.[96] While ASIC did not take action against the directors and officers, Westpac incurred a $1.3 billion penalty arising from the AUSTRAC action, as well as ongoing adverse publicity throughout the case and following the judgment, the external review and the Advisory Panel Report, the subsequent and ongoing shareholder class action.

However, the regulator could take action against a financial sector entity, as well as its directors and officers, for misuse of personal information through inadequate AI and information governance due to the risk management system being inadequate to protect personal information and comply with regulatory obligations.

Reporting by senior executives to board committees and to the board is typically carried out in accordance with the risk policy, risk management framework and the risk appetite statement, which are set by the board. Embedding AI governance into the enterprise risk management framework enables the formal reporting requirements on information and AI risks in

---

austrac-statement-of-claim-issues-media-release.pdf> (accessed 20 January 2025) (*Westpac Advisory Panel Report*).

91 *CPS 220* (n 68) cl 23.

92 P Hanrahan and T Bednall, 'From Stepping-Stones to Throwing Stones: Officers' Liability for Corporate Compliance Failures after *Cassimatis*' (2021) 49(3) *Federal Law Review*, p 380, 402.

93 R T Langford, 'Data Explosion, Disclosure and Stepping Stones' in A Godwin, P W Lee and R T Langford (Eds), *Technology and Corporate Law: How Innovation Shapes Corporate Activity* (Edward Elgar Publishing, 2021) pp 99–125, 115.

94 Above.

95 *AUSTRAC v Westpac* (n 89); see below for further discussion of the case.

96 *Westpac Advisory Panel Report* (n 90).

accordance with the organisation's existing risk framework and reporting processes to the board or governing authority. This means that boards need to review the enterprise risk management framework and the overarching risk policy and adjust these formal governance documents to ensure that both AI and information capabilities are clearly identified as risks so that the performance of both AI and information capabilities are adequately reported.

In addition to the traditional siloed reporting from designated areas and senior executives, such as IT (CIO/CISO), Legal (general counsel), Data (chief data officer) and, for example, now AI/Tech Innovation (CAIO), we propose an additional report providing an integrated performance assessment and reporting of information within the organisation to enable the governing body to adequately monitor the interconnected risks in accordance with the organisation's risk appetite and strategic objectives.

In this way, the integration of AI and information governance with the enterprise risk management framework can enable information as a risk to be more holistically assessed and monitored by the governing authority in accordance with the organisation's risk appetite statement and overarching objectives. In addition to reducing AI and information risks and improving the value derived from information, effective information governance can also improve the quality of the reporting to the board and overall corporate governance of the organisation.

B The AI governance: Challenges for organisations

In light of potential actions against financial sector entities and directors discussed above, it is incumbent on directors to ensure that risk management systems are adequate to manage AI risks. In addition to the regulatory guidance requirements in the form of the *CGPs* and the APRA Prudential Standards for risk management, there is now an emerging plethora of guidance, specifically on AI governance frameworks, which can assist organisations in mitigating AI risks.

The guidance available, however, often comes short in the context of addressing the practical challenges for organisations in implementing effective AI risk management. This is especially evident in relation to four factors:

- first, the significant work and cross-disciplinary collaboration required to implement comprehensive AI risk management;
- second, the challenge of integrating data-driven technology with broader regulatory compliance as illustrated by AUSTRAC cases against the Commonwealth Bank of Australia (CBA) and Westpac;[97]
- third, the poor state of organisational information governance as revealed by high-profile data breaches; and
- fourth, organisational culture potentially impeding effective technology risk management.

97 *AUSTRAC v Westpac* (n 89); *Chief Executive Officer of AUSTRAC v Commonwealth Bank of Australia Ltd* [2018] FCA 930 (*AUSTRAC v CBA*); see below for more detailed discussion.

We consider these in the following sections.

1 AI governance standard and frameworks: Resources and work required for implementation

While AI governance standards are voluntary, they provide a detailed guide of the steps organisations may take to implement adequate risks management of AI systems. In December 2023, ISO and the International Electrotechnical Commission (IEC) jointly released *ISO/IEC 42001* (*AI Standard*), which is a certifiable framework for an AI management system to support organisations in the responsible development, delivery or use of AI systems.[98] In February 2024, Standards Australia announced it had adopted *ISO/IEC 42001*, stating it 'paves the way for AS 42001:2023 to be the key standard for AI governance in Australia'.[99] The *AI Standard* is described as 'applicable to any organization developing, providing, or using products of services that utilize an AI system,' and lists examples of six sectors, one of which is the finance sector.[100] It sets out principles covering the context of the organisation, leadership, planning, support, operation, performance evaluation and continual improvement, and recognises that the AI management system should be integrated with the organisation's 'management of risks and opportunities'.[101]

The *AI Standard* sets out similar AI responsibilities for those who direct and control the organisation, as required under the *CGPs*.[102] This includes requiring organisations to 'define and establish an AI risk assessment process'[103] and reporting on the performance of the AI management system.[104] The starting point is to establish an AI policy and AI objectives.[105]

Annex A, supplementing the principles set out in the *AI Standard*, provides a reference for setting 'control objectives and controls'. The controls state that the AI policy be documented, that it aligns with other policies and that it be reviewed 'at planned intervals or additionally as needed'. As an example of the level of detail provided, controls and guidance for 'Data for AI Systems'[106] cover the following: data development and enhancement of AI systems, acquisition of data, quality of 'Data for AI Systems', data provenance and data preparation. The guidance for 'Data for AI Systems' then provides

98 *Information Technology — Artificial intelligence — Management System* (1st ed, ISO/IEC 42001:2023, December 2023) (*AI Standard*).

99 'Standards Australia adopts the international standard for AI Management System, AS ISO/IEC 42001:2023', *Standards Australia* (Online, 16 February 2024) <https://www.standards.org.au/news/standards-australia-adopts-the-international-standard-for-ai-management-system-as-iso-iec-42001-2023> (accessed 20 January 2025).

100 *AI Standard* (n 98) annex D.1.

101 Above vi.

102 Above cl 5.1.

103 Above cl 6.1.2.

104 Above cl 5.3(b).

105 Above cl 5.2.

106 Above annex A.7, B.7.

implementation guidance for each of the aforementioned areas.[107] In respect of data for development and enhancement of AI systems, it provides:

> The organization should define, document and implement data management processes related to the development of AI systems.
>
> Implementation guidance
>
> Data management can include various topics, such as, but not limited to:
> — privacy and security implications due to the use of data, some of which can be sensitive in nature;
> — security and safety threats that can arise from data-dependent AI system development;
> — transparency and explainability aspects, including data provenance and the ability to provide an explanation of how data are used for determining an AI system's output if the system requires transparency and explainability;
> — representativeness of training data compared to operational domain of use;
> — accuracy and integrity of the data.
>
> NOTE: Detailed information of AI system life cycle and data management concepts is provided by ISO/IEC 22989.[108]

Establishing and maintaining an overarching AI policy to guide the use of AI within an organisation and set the overarching guardrails is the easy part. As can be seen from the implementation guidance for data management above, the devil in AI implementation is in the detail, particularly to ensure that the use of personal information complies with regulatory and ethical requirements. This requires sufficient resources and the requisite level of cross-functional and multi-disciplinary professional collaboration between different areas and skill sets in the organisation in order to achieve safe, data-driven technology innovation. However, as shown in the following sections, the practical challenge of technology integration, data governance and regulatory compliance has led to significant failures with significant regulatory consequences.

2 The technology integration challenge

The challenge of technology implementation in the financial sector has resulted in significant failures to comply with regulatory requirements. These cases highlight the risks and dangers for financial sector entities and directors arising from the use of technology and data, providing salient lessons of the challenges boards face in adequately managing the risks against the opportunities of AI technology.

The two Federal Court actions taken by AUSTRAC highlight governance failures by CBA and Westpac.[109] In summary, both cases concerned information and technology system governance failures resulting in breaches of the *Anti-Money Laundering and Counter-Terrorism Financing Act 2006* (Cth) (*AML/CTF Act*) for failures of reporting and the inadequacies of compliance systems, processes and oversight. Westpac was ordered to pay a

107 Above annex B (Implementation guidance for AI controls).
108 Above annex B.7.2.
109 *AUSTRAC v Westpac* (n 89); *AUSTRAC v CBA* (n 97).

$1.3 billion penalty on 21 October 2020, which is the largest fine in Australian corporate history.[110]

In Westpac's case, AUSTRAC alleged 23 million *AML/CTF Act* breaches for 'systemic noncompliance'.[111] AUSTRAC stated that:

> [t]hese contraventions are the result of systemic failures in its control environment, *indifference by senior management and inadequate oversight by the Board.*[112]

Westpac admitted that it was late in reporting to AUSTRAC and failed to provide payer names, as required under s 45(2) of the *AML/CTF Act*. This included, for example, the failure to 'detect activity on its customers' accounts that [was] indicative of child exploitation'.[113] The failures to retain records and backup files took over six years to be identified by Westpac's IT change management and assurance processes.[114] The fact that Westpac had failed to detect and rectify system issues when AUSTRAC filed its action in November 2019 was surprising given that AUSTRAC had initiated civil penalty proceedings against CBA in August 2017 for 'serious and systemic noncompliance' with the *AML/CTF Act*.[115]

In June 2018, CBA agreed to pay a $700 million penalty to resolve the proceedings with AUSTRAC over its breaches of the *AML/CTF Act* for failing to report deposits of AUD$10,000 or more. The Statement of Agreed Facts filed in the Federal Court revealed that by July 2015, CBA had evidence that crime syndicates had used a system flaw in its intelligent deposit machines that involved money laundering of the proceeds of drug importations.[116] CBA said the breaches were due to a system error with the machines, which failed to introduce a daily limit on how much customers could deposit. Even after it became aware of the suspected money laundering, CBA accepted that it did not monitor its customers to mitigate and manage money laundering and terrorism financing risks.[117] Justice Beach described some of the failures in the

---

110 *AUSTRAC v Westpac* (n 89) at [116], [191].

111 AUSTRAC, *AUSTRAC applies for civil penalty orders against Westpac* (Media Release, 20 November 2019) <https://www.austrac.gov.au/about-us/media-release/civil-penalty-orders-against-westpac> (accessed 20 January 2025).

112 *Chief Executive Officer of AUSTRAC v Westpac Banking Corp ACN 007 457 141* (Federal Court of Australia, NSD1914/2019, commenced 20 November 2019) (Concise Statement) <https://www.austrac.gov.au/sites/default/files/2019-11/20191120%20Westpac%20Concise%20Statement%20FILED%2019008953.pdf> (accessed 20 January 2025) (emphasis added).

113 Above at [24]–[25]; see also A Ferguson, 'Spectacular failure: This is one scandal Westpac can't play down', *Sydney Morning Herald* (Online, 20 November 2019) <https://www.smh.com.au/business/banking-and-finance/spectacular-failure-this-is-one-scandal-westpac-can-t-play-down-20191120-p53cbn.html> (accessed 20 January 2025).

114 *Chief Executive Officer of AUSTRAC v Westpac Banking Corp ACN 007 457 141* (n 112) at [112].

115 AUSTRAC, *AUSTRAC seeks civil penalty orders against CBA* (Media Release, 3 August 2017) <https://www.austrac.gov.au/austrac-seeks-civil-penalty-orders-against-cba> (accessed 20 January 2025).

116 *AUSTRAC v CBA* (n 97); *Chief Executive Officer of AUSTRAC v Commonwealth Bank of Australia ACN 123 123 124* (Federal Court of Australia, NSD1305/2017, commenced 3 August 2017) at [3], [16]–[17] (Statement of Agreed Facts and Admissions) <https://www.commbank.com.au/content/dam/caas/newsroom/docs/2018-06-04-CBA-AUSTRAC-SAFA.pdf> (accessed 20 January 2025).

117 Above.

CBA case as 'derived from a misapprehension of obligations imposed under the Act, an inadvertent failure to update and configure automated processes, and an error when merging data from two of the CBA's systems'.[118]

In both CBA and Westpac cases, their boards and senior management were responsible for ensuring they complied with the *AML/CTF Act*. In respect to the Westpac case, Beach J recognised that the board and senior management sought and received regular and detailed reports in relation to the identification, mitigation and management of the money laundering and terrorism financing risks. However, Beach J identified many deficiencies, including inter alia, 'weaknesses in Westpac's data management and technology systems in relation to AML/CTF compliance' with compliance and risk management functions not being adequately resourced.[119]

CBA and Westpac cases highlight how significant regulatory breaches and class actions can be brought about by failures in implementing and integrating technology systems with regulatory reporting requirements. These cases also illustrate the role of human errors in how data and information are governed. In addition, as noted in Section IV(A), actions may be taken by ASIC against directors for a breach of the duty of care and diligence to properly monitor data-driven technology systems that result in regulatory breaches or where there is a failure to adequately respond to medium (amber) and high-risk (red) flags in board reports. This provides a further reason for boards to ensure the risk management framework and reporting processes within the organisation are providing the board with the *right* information in respect of AI systems and information management.

3 The data and information governance challenge for AI systems

In addition to technology integration failures, information governance failures revealed in high- profile data breaches point to the practical challenge of data governance for AI projects. While the strategic imperative may be to implement innovative AI technology, an impediment is the accuracy and quality of data to be used. If the data being processed includes outdated and/or inaccurate information, the insights and information derived from the use of AI will likely also be inaccurate.[120] This may impact the achievement of the strategic goals through the use of AI. Further, personal information used in AI systems, especially if not for the purpose(s) for which it was collected, may be in breach of privacy regulatory requirements and/or consumer law if it constitutes misleading or deceptive conduct because it is inconsistent with an entity's privacy policy.[121]

The Optus, Medibank and Latitude data breaches highlighted the over-retention of personal data and the risks involved for organisations subject to regulatory investigations by the Office of the Australian Information Commisioner. Multiple class actions have been commenced against these companies.[122] In the Latitude data breach, of the approximately 14 million

118 *AUSTRAC v Westpac* (n 89) at [71].
119 Above at [170].
120 See examples discussed in Section II(B).
121 See *Australian Privacy Principle (APP) 6 Privacy Act 1998* (Cth) (*Privacy Act*) and *Competition and Consumer Act 2010* (Cth) Sch 2 s 18.
122 Optus data breach 22 September 2022: Optus, *Optus notifies customers of cyberattack*

records disclosed, '6.1 million records dating back to at least 2005 were also stolen, of which ... 94 percent were provided before 2013'.[123] While the strategic focus has been on technology innovation, these high-profile data breaches are a reminder to boards that they are responsible for ensuring effective cybersecurity risk management exists to protect the organisation's technology and information assets. They also highlight that boards need to ensure there is effective information governance to protect personal information in compliance with existing regulatory requirements, including the obligation to de-identify or dispose of personal information when it is no longer required, thus avoiding over-retention of customer data, particularly of past customers.[124]

In addition to the governance of 'Data for AI Systems', organisations should have in place, as part of their enterprise-wide information governance, an information and data lifecycle management program to manage data from collection to its secure archival and disposal. With the ever-increasing volumes of data being stored by organisations, this requires a collaborative and planned program that is executed on an ongoing basis. Taking a piecemeal approach to managing data for use in AI systems brings inherent risks. These include using inaccurate data in innovation projects, as well as increasing risks in the event of a data breach for failures to comply with requirements to have adequate security in place to protect information, as well as complying with data minimisation requirements. Robust information governance provides the mechanisms and enables the collaboration required across organisational silos (eg, Records, Privacy, IT and Business Operations) for meeting record-keeping obligations, including disposal of data in accordance with privacy regulations and to reduce risk in the event of a data breach.

4 The impact of organisational culture on AI governance

Apart from the practical challenges of technology integration and the need for robust data governance, an organisation's culture can impact on the effectiveness of AI governance system that is put in place. For example, the costly failure of the Robodebt scheme and the ensuing scandal shone a light on the impact of workplace culture and its devastating consequences. This included harms to vulnerable people, overestimating savings, underestimating costs, a failure to understand the risks involved, challenges brought in the Administration Appeals Tribunal and a class action against the Commonwealth government, which settled for $112 million.[125]

---

compromising customer information (Media Release, 22 September 2022) <https://www.optus.com.au/about/media-centre/media-releases/2022/09/optus-notifies-customers-of-cyberattack> (accessed 20 January 2025); Medibank data breach 20 October 2022: E Ritchie, 'Medibank cyber incident response', *Medibank* (Online, 20 October 2022) <https://www.medibank.com.au/livebetter/newsroom/post/medibank-cyber-incident-response> (accessed 20 January 2025); Latitude data breach 16 March 2023: Latitude, *ASX Announcement: Cybercrime Update* (2023) Report (Latitude) <https://investors.latitudefinancial.com.au/DownloadFile.axd?file=/Report/ComNews/20230316/02644401.pdf> (accessed 20 January 2025).

123 Latitude, *Cybercrime Update* (Media Release, 27 March 2023) <https://www.latitudefinancial.com.au/about-us/media-releases/cybercrime-update-27-03-2023.html> (accessed 20 January 2025).

124 See *Privacy Act* APP 11.2.

125 *Prygodicz v Commonwealth (No 2)* (2021) 173 ALD 277; [2021] FCA 634.

Commissioner Holmes AC SC stated:

> The failure to confront fundamental flaws in the operation of the Scheme appears to have been a product of the culture within DHS at the time.[126]

The Royal Commission Report referred to a survey for public servants carried out by the Community and Public Sector Union in which 'nearly all members who responded to the survey raised concerns within DHS about the legality of the scheme, but were told that the legal advice was that it could proceed'.[127] Robodebt illustrates the importance of a culture that enables problems to be openly raised, discussed and resolved, including with independent legal advice where appropriate.

The culture of not giving bad news was found to extend to in-house lawyers whose professional duties required them to maintain their independence.[128] A former principal legal officer explained:

> A very siloed type of culture. You were responsible for what you were responsible for and stayed within those bounds. ... [I]t wasn't our role to turn our mind to broader risks than what was being explicitly asked.[129]

The role of professionals, particularly in-house lawyers who have a duty to be independent, provides important safeguards. AI and technology projects using personal data are required to comply with a myriad of regulations, including, as shown in the AUSTRAC cases, regulatory reporting and record-keeping. This requires technology, privacy, legal and record-keeping to work together with the business areas driving implementation and those who will be dealing with enquiries and/or complaints once the system is implemented. In contrast to the cultural issues identified in Robodebt of silos and discouraging bad news or red flags, an organisational culture that encourages staff at all levels to contribute their knowledge, including identifying risks and/or problems as they emerge, enables appropriate action to be taken to remediate the risks and problems. Providing there are adequate risk, audit and board reporting processes in place, it also will enable risks and problems to be reported in accordance with the risk management framework.

V Conclusions

As AI becomes increasingly embedded within the financial industry, directors will need to ensure that the use of technology complies with regulatory obligations. Recent cases, lawsuits and other examples discussed illustrate the substantial risks that can materialise for financial firms using AI tools. Robust AI governance and risk management are required to properly manage those risks.

The hype around AI has been reignited by the popularisation of GenAI for use in business operations. In the drive to 'innovate' and apply AI to a range of tasks,[130] there is a risk that more traditional solutions, including non-AI

126 Royal Commission (n 12) p 124.
127 Above p 392.
128 Above p 519.
129 Above p 124.
130 See, eg, M Chui et al, 'What Every CEO Should Know about Generative AI', *McKinsey*

technology, may be disregarded even if they may produce better outcomes while reducing risks. Furthermore, to discharge their duties, and in particular the duty of care and diligence, company directors must carefully weigh potential gains and risks that implementing a new AI system may bring about. For example, mere statistical superiority of an AI tool in relation to a specific task, will not always mean it is overall the best solution for the company, as the possible harms, risks and unknowns need to be considered.[131]

The AI regulatory landscape is rapidly evolving, although at a slower pace than technological advances. Therefore, it is crucial for financial sector entities to regularly review their risk management processes and AI and data governance to ensure compliance with current and new regulatory obligations as they arise. Importantly, directors need to ensure that the risk management and information governance frameworks and processes are adequate to sufficiently inform them of the risks arising from the AI tools used in the organisation. This should allow the directors to properly monitor AI and technology systems.

The issue of adequacy of reporting to the board to enable directors to be properly informed to make effective decisions is shaped by the risk management framework. Given the requirements set out in the *AI standard*, the technology integration challenges revealed by the CBA and Westpac cases, and the over-retention of personal data revealed in the Optus, Medibank and Latitude data breaches, it follows that boards need to consider whether their current risk management framework enables the directors to monitor information risks arising from the interconnection between data, information, technology and regulatory compliance. This is a rapidly evolving area of corporate governance that would benefit from further research.

Digital (Web Article, 12 May 2023) <https://www.mckinsey.com/capabilities/mckinsey-digital/our-insights/what-every-ceo-should-know-about-generative-ai> (accessed 20 January 2025), which is an example of an increasing pressure on corporate leaders to adopt GenAI to tasks that are currently achieved through different means.

131 Emerging research in the context of medical devices and doctor's duty of care provides a useful lesson: mere statistical superiority of an AI device when compared with a different solution — for example, a human radiologist describing medical imaging — is not enough for the foreseeability of harm to be established, and thus there is no positive duty for a doctor to use that AI device: A Berz, 'Confidentially Transparent: Balancing Accountability and IP Protection in AI Deployment', presentation at the *Intersections of Private Law Colloquium* (University of Sydney, 7 April 2024).