



MONASH University

Regulatory Approaches to AI in Australia: addressing the legal issues

David Geoffrey Watts
B.A., LL.B.

A thesis submitted for the degree of Doctor of Juridical Science
at
Monash University in 2025
Faculty of Law

Copyright notice

© David Watts 2025

Abstract

As the adoption of generative AI gains pace governments across the world, including in Australia, have been preoccupied with questions about how to design a regulatory architecture that maximises its economic and societal benefits while mitigating its harms. A common thread of the debate has been to ensure that any regulatory measures that are developed do not stifle innovation.

Initially the Australian government's approach was to adopt overarching regulation, informed by European AI regulation, based around a series of 'guardrails' that would apply to high-risk AI systems. More recently, it appears to be retreating from this position following work undertaken by the Productivity Commission, motivated by prioritising the productivity gains that could be hampered by overarching 'guardrails' legislation, and recommending against this approach as a last resort that should only be implemented where other, existing legal and regulatory frameworks are shown to be unsatisfactory. This thesis broadly supports the Productivity Commission's approach but for different reasons.

Drawing on rule of law concepts that suggest that a regulatory architecture should enable existing laws to apply to the artificial, this thesis argues that the starting point of Australia's approach to AI regulation should be constituted by a principle of equivalence where the artificial is regulated in an equivalent manner to human activity. Where there are disjunctions between the two, such as the lack of transparency, accountability and contestability of AI systems, additional regulation should be developed. This thesis identifies a series of foundational input AI risks that relate to AI training data, these being data protection and copyright risks. It argues that data protection reform is essential to account for these risks but that existing copyright law collective licensing mechanisms are capable of being adapted to the AI environment. It suggests that AI output risks should be controlled by strict liability concepts but with the addition of a safe harbour overseen by an AI Office under which AI developers and deployers would have the opportunity to lodge transparency and accountability material that could then be made available for consideration by the courts on a case by case basis. If found to be sufficiently transparent, safe harbour requirements would be satisfied and normal fault-based liability concepts would apply.

Declaration

This thesis is an original work of my research and contains no material which has been accepted for the award of any other degree or diploma at any university or equivalent institution and that, to the best of my knowledge and belief, this thesis contains no material previously published or written by another person, or any use of generative artificial intelligence technologies, except where due reference is made in the text of the thesis.

Print Name: David Watts

Date: 19 February 2026

Publications during enrolment

Some material in Chapter 2 of this thesis appears in Chapter 16 of Emeritus Professor Arie Freiberg's textbook, *Regulation in Australia*, (The Federation Press, 2025). I am acknowledged as the co-author of that chapter.

Acknowledgements

I wish to acknowledge the support and inspiration provided by Professor Emerita Moira Paterson, Professor Pompeu Casanovas, Professor Emerita Danuta Mendelson (dec'd) and Dr Bridget Bainbridge, each of whom have generously contributed their ideas and feedback at various stages during the research and writing of this thesis.

Table of Content

Chapter One – Introduction.....	8
Background.....	8
AI issues and risks.....	8
Public concerns about AI.....	10
Regulatory background issues.....	11
Key Terms: Defining AI.....	11
Regulation and the technology sector.....	12
The evolving international regulatory landscape.....	13
The current approach to AI regulation in Australia.....	15
Developing a regulatory approach for AI in Australia: a wicked problem?.....	19
Research question.....	21
Thesis methodology.....	21
Thesis structure.....	22
 Chapter 2 – What is Artificial Intelligence: Concepts, Terminologies, Technologies and Definitions.....	 25
Terminology and the ‘odd paradox’	26
Early development.....	27
Origins of AI.....	27
Weak AI: first wave.....	28
From Weak first wave to narrow second wave.....	29
Narrow AI: second wave.....	29
Neural Networks.....	31
Deep Learning.....	32
Black Boxes.....	32
Data.....	33
Training Data.....	33
Algorithms.....	34
Large Language Models (LLMs).....	35
Generative AI.....	36
Agentic AI.....	36
Narrow AI: some conclusions.....	37
Disciplines and Skills.....	37
Definitions of AI.....	38
Other approaches.....	40
Closing observations.....	41
 Chapter Three – The foundational risks posed by AI.....	 42
Background.....	42
Approach and scope.....	42
Foundational risks.....	45
Data scraping.....	46
Data Protection.....	53
Copyright.....	65
Fault-based liability: negligence.....	68
Administrative law and AI illegality.....	71
Discrimination.....	76
Conclusion.....	79
 Chapter Four- AI regulation and the issues it raises.....	 81
Introduction.....	81
AI regulation as a public policy issue.....	81
AI, innovation and disruption.....	82
Regulation and wicked problems.....	83
The lengthy gestation of AI: how it impedes AI regulation.....	84

Regulation and the broader tech sector.....	85
The innovation v regulation debate.....	88
Innovation and <i>ex ante</i> regulation.....	90
EU Regulation and the Innovation Debate.....	91
Chinese AI Regulation and the AI Debate.....	92
Learning from Experience.....	93
AI Ethics.....	93
AI governance.....	96
Conclusion.....	97
Chapter Five - Insights from international initiatives.....	99
Introduction.....	99
International law and related initiatives.....	99
China.....	102
European Union.....	108
United States.....	115
United Kingdom.....	118
Continuing concerns.....	119
What can Australia learn from these developments?.....	121
Chapter Six – A suggested way forward for Australia.....	123
Approach.....	123
The current position in Australia.....	124
AI regulation: first principles.....	126
Data protection.....	127
Copyright.....	131
AI liability.....	133
Strict liability as a default.....	134
What is a safe harbour?.....	136
An AI safe harbour.....	137
Indirect Effects.....	140
International Alignment.....	141
Conclusion.....	141

Chapter One – Introduction

Background

Although Artificial Intelligence (AI) technology has existed since the 1950s,¹ its rise to public prominence, expanding capabilities and rapidly increasing adoption by government, business and individuals is a comparatively recent phenomenon. The use of AI has grown gradually over the last decade with AI moving increasingly from the computer laboratory into the mainstream. This trend has accelerated with the rise of a wave of generative general-purpose AI (GPAI) systems, beginning when OpenAI released ChatGPT, ‘an era-defining product,’² in 2022. AI systems are now frequently being used to make decisions that have a significant impact on people’s lives in areas such as finance, healthcare, employment, housing, criminal justice, public administration and more.

There appears to be widespread agreement that AI has the potential to ‘improve social and economic well-being’³ and to enhance ‘quality of life, grow our economy and support a future made in Australia.’⁴ It is claimed that adopting AI in Australia will produce ‘multifactor productivity gains above 2.3% ... over the next decade, *though there is considerable uncertainty*.’⁵ This would translate into ‘about 4.3% labour productivity growth over the same period. But poorly designed regulation could stifle the adoption and development of AI and limit its benefits.’⁶

AI issues and risks

As the use of AI grows, a range of problems have been identified. AI systems can ‘hallucinate,’ that is, create incorrect, inaccurate and irrational results. The outputs can be biased, unsafe and unreliable. AI can produce and amplify discrimination. It can generate false and misleading content such as deepfakes that can cause harm to individuals and society. AI that works well in one context might fail in another. And there are signs that some

¹ Stephen Muggleton, ‘Alan Turing and the Development of Artificial Intelligence’ (2014) 27 AI Communications 3.

² Matteo Wong, ‘OpenAI Can Stop Pretending’, *The Atlantic*, 30 May 2025. See <<https://www.theatlantic.com/technology/archive/2025/05/openai-nonprofit-pbc/682979/>>.

³ Australian Government, Department of Industry, Science and Resources, *Safe and Responsible AI in Australia: Proposals Paper for Introducing Mandatory Guardrails for AI in High-risk Settings*, September 2024, p 2. See <https://storage.googleapis.com/converlens-au-industry/industry/p/prj2f6f02ebfe6a8190c7bdc/page/proposals_paper_for_introducing_mandatory_guardrails_for_ai_in_high_risk_settings.pdf>.

⁴ Ibid p 3.

⁵ My emphasis.

⁶ Australian Government, Productivity Commission, *Harnessing Data and Digital Technology Interim Report*, August 2025. See <<https://www.pc.gov.au/inquiries/current/data-digital/interim/data-digital-interim.pdf>>.

types of generative AI may pursue its own ends, rather than those of its developers, users and society in general.⁷

Apart from these (and many other) problems, the most widely deployed generative AI systems that utilise machine learning techniques are black boxes: they are systems whose internal workings are unknown to both their creators and users. Although their inputs and outputs can be known, what takes place within the black box is occluded. Black box AI is different from the disruptive technologies that preceded it – such as the printing press, steam power, aviation and telephonic and radio communications. In each wave of prior disruption, there were clear explanations about how the technology worked and what it did. However disruptive, its operations could be understood. Although black box AI can be described in general terms, what goes on inside the black box cannot.

At the same time, the design and deployment of AI challenges a range of legal norms such as data protection law,⁸ administrative law, copyright law, tort law, and discrimination law, all of which were developed on the basis that they covered human, not artificial, agency.

AI legal issues and risks arise across a multitude of domains and take various forms. Arie Freiberg lists more than seventy.⁹ They include cyber-abuse and cyber security threats, facilitation of child sexual abuse, deep fakes, identity theft, hate speech, consumer scams, AI-generated content that is represented as being real, cyber security risks and infringements of human rights. A comprehensive analysis of these legal risks would require a thesis for each.

This thesis is designed to take a broad approach to the legal risks of AI and to provide an account of how these might be addressed from an Australian perspective. This requires a focus upon a limited number of *foundational* AI risks, defined as risks that facilitate the production of a variety of downstream harms. For example, the way in which AI systems misuse personal data is responsible for contributing to a cascading number of harms, including all of those identified in this section. It is these foundational risks that are the focus of this thesis and its suggestions about how they might be addressed and mitigated through the legal process.

⁷ Bryan Robinson, *AI Goes Rogue: Do 5 Things if Your Chatbot Lies, Schemes or Threatens*, Forbes, 4 August 2025. See <<https://www.forbes.com/sites/bryanrobinson/2025/07/03/ai-goes-rogue-do-5-things-if-your-chatbot-lies-schemes-or-threatens/>>.

⁸ In this thesis, I use the expressions 'data protection' and 'information privacy', and 'personal data' and 'personal information', synonymously.

⁹ Arie Freiberg, *Regulation in Australia*, 2nd ed, 2025, pp 460-462

Analysis of these foundational risks includes two Australian case studies. The first relates to the Robodebt programme (an automated system for welfare overpayment detection and collection)¹⁰ and the more recent Targeted Compliance Framework (an automated system for cancelling income support based on failure to comply with required mutual compliance obligations).¹¹ Both of these illustrate the challenges posed to administrative law (and the rule of law itself) by the application of simple, first wave AI.¹² This includes the concerning repetition of government errors in applying legal rules, even in a comparatively simple AI environment. The second relates to the use of AI in the employment screening process in Australia. It demonstrates the opaque and, in some cases, unlawfully discriminatory use of AI to determine individuals' suitability for employment. Much of the underlying research into AI risks has taken place outside of Australia. In this thesis I have prioritised Australian research into local AI risks where feasible and appropriate.

Public concerns about AI

Analysis of public expectations and preoccupations about AI in Australia highlights societal concerns. A 2024 University of Queensland survey¹³ found that Australians are concerned about diverse risks, including from AI systems that are unsafe, untrustworthy and not aligned to human values; AI's threats to employment; cyber security risks and the development of lethal autonomous weapons. The survey also found that Australians support regulatory and non-regulatory action to address risks from AI. Actions nominated included the need to prevent dangerous and catastrophic outcomes from AI; requiring mandatory audits to make sure AI is safe before release; making sure that AI companies are liable for the harms they cause; and reducing AI-related job losses.

A 2025 survey undertaken by the University of Melbourne on *Australian Trust, Attitudes and Use of AI* found that 'Australians are more worried than optimistic or excited about AI'¹⁴ and that they 'expect oversight by government and existing regulators, co-regulation with industry, and international laws and regulations.'¹⁵ Forty six per cent 'believe risks of AI

¹⁰ Royal Commission into the Robodebt Scheme, Report, 2023. See <<https://robodebt.royalcommission.gov.au/publications/report>>.

¹¹ Commonwealth Ombudsman, *Automated in the Targeted Compliance Framework: When the Law is Changed but the System Isn't*, August 2025. See <https://www.ombudsman.gov.au/__data/assets/pdf_file/0017/320750/Automation-in-the-Targeted-Compliance-Framework.pdf>.

¹² First wave AI is AI where its 'intelligence' is built in. It is not capable of 'learning.' The distinction is discussed in greater depth in Chapter 2.

¹³ AI Governance in Australia, *2024 Australian Survey Assessing Risks from AI*. See <<https://aigovernance.org.au/survey/>>.

¹⁴ KPMG and the University of Melbourne, *Trust, Attitudes and Use of Artificial Intelligence: A Global Study 2025, Australian Insights*. See <<https://assets.kpmg.com/content/dam/kpmg/au/pdf/2025/trust-in-ai-global-insights-2025-australia-snapshot.pdf>>.

¹⁵ Ibid.

outweigh the benefits;¹⁶ seventy eight per cent are ‘concerned about negative outcomes from AI;¹⁷ and thirty seven percent ‘personally experienced or observed negative outcomes.

Regulatory background issues

Considering these concerns, including the perceived need to take regulatory action to mitigate AI risks, what is the best way forward for Australia? Developing a regulatory approach is complicated by a range of conceptual, policy, regulatory and legal issues. These include ambiguities associated with fundamental concepts such as defining AI; the need to appropriately balance innovation and regulation; accounting for the international regulatory environment; assessing the extent to which existing laws and regulatory instruments provide an appropriate regulatory foundation and, where they do not, what changes need to be made; and whether new, bespoke AI regulation is warranted and, if so, what form it should take.

Key Terms: Defining AI

Discussions about AI are replete with definitional ambiguities, the use of concepts that are poorly defined and the intermingling of marketing language with technical jargon and legal terms. Although these are discussed in greater detail later, it is necessary to address the definition of AI at the outset to provide clarity to this introduction.

There is no consensus definition of ‘Artificial Intelligence.’ AI can be seen as a variety of computer-based technologies that emulate what humans can do. Its development has occurred in three key waves:

1. ‘First wave AI’ covers relatively simple technologies that can excel at specific tasks within a limited context, such as playing chess or language translation. The ‘intelligence’ relies upon humans embedding the necessary knowledge within the AI system, enabling it to undertake a domain-limited task.
2. ‘Second wave AI’ which emulates human thought and agency using machine learning techniques involving statistical and probability analysis.
3. ‘Generative AI’ which produces new outputs across many domains.

¹⁶ Ibid.

¹⁷ Ibid.

While each wave emulates human intelligence and agency, it does so in different ways and involves different risks.¹⁸

A variety of definitions of AI are proposed and in use, including the Organisation for Economic Cooperation and Development's (OECD's) influential definition of 'AI system' (which has been adopted in Australia's current policy work on AI) and the different, but broadly similar, definition found in the European Union's (EU's) AI Act. Notably, both definitions exclude first wave AI.¹⁹ As illustrated by the case study of the Robodebt scheme, the use of first wave AI caused significant damage in Australia. This thesis argues that a broader definitional approach is required, incorporating all three waves of AI development. It adopts the following definition of AI developed by the Supreme Court of Victoria:

A term describing a range of technologies and techniques used to computationally generate outputs that typically require human intelligence to produce.²⁰

This definition is broadly similar to that proposed by the Productivity Commission which defines AI as 'machine systems that perform tasks typically requiring human intelligence, such as learning, reasoning, and decision-making. Unlike traditional algorithms — where humans specify a task — AI models can now infer how to generate outputs from inputs, often with minimal human oversight.'²¹

Regulation and the technology sector

Definitional issues aside, government efforts to regulate technology, particularly information and communications (ICT)-related technologies, has proven to be difficult and highly contested. As a result:

The default approach of states today is not to interfere in nascent (likely disruptive) technologies but rather to let the market regulate unless there are public safety or security concerns.²²

¹⁸ Another stage of AI has been posited and is the end goal of much current AI research and development. This is known as the 'singularity' or 'Artificial General Intelligence' (AGI). AGI is a theoretical next step in the evolution of AI. It describes AI that matches or exceeds the cognitive abilities of humans across any domain: see, for example, DC Bikkasani, 'Navigating artificial general intelligence (AGI): societal implications, ethical considerations, and governance strategies' (2025) 5 *AI Ethics* 2021–2036, <<https://doi.org/10.1007/s43681-024-00642-z>>. As a theoretical concept it is beyond the scope of this thesis.

¹⁹ For example, Recital 12 of the EU AI Act states that it is not intended to apply to 'systems that are based on the rules defined solely by natural persons to automatically execute operations.'

²⁰ Supreme Court of Victoria, *Guidelines for Litigants: Responsible Use of Artificial Intelligence in Litigation*, May 2024. See <<https://www.supremecourt.vic.gov.au/sites/default/files/2024-05/AI%20Guidelines%20SCV.pdf>>.

²¹ *Harnessing Data and Digital Technology Interim Report* (above, n 6) p 11.

²² Julia Black and Andrew Murray, *Regulating AI and Machine Learning: Setting the Regulatory Agenda*, (2019) 10(3) *European Journal of Law and Technology* 1.

The technology sector developed in a ‘light-touch,’ permissive regulatory environment. The policy underpinnings for this consist of a range of arguments and philosophical positions that converged at about the time that the internet, and in particular, the world wide web, became widely accessible in the early 1990s. These arguments prioritise the benefits of technology over the risks and harms it can cause and claim that regulation stifles innovation. In western economies the tendency has been to rely on existing laws, market self-regulation, co-regulation and some legislation to address specific risks where necessary.

Australia has not always followed the international market-driven consensus. It has occasionally developed its own, bespoke approach. An example is the *Online Safety Act 2021 (Cth)* (OSA) which, although not specifically designed as an AI-specific initiative, helps to remediate online harm and improve the safety of online content. This can include content generated by AI. Part 4A of the OSA²³ is a world-first initiative that introduces an obligation on age-restricted social media platforms, through the development of co-regulatory industry codes, to prevent children under the age of sixteen years from holding accounts on those platforms. Equally, the exemptions for small business and employee records in the *Privacy Act 1988 (Cth)* (Privacy Act) represent a unique Australian approach to data protection.

The evolving international regulatory landscape

AI technologies are developing at pace. In response, governments are seeking ways to ensure that its benefits are optimally realised. As a result, regulatory proposals and approaches are often being bundled into holistic, broad policy initiatives that prioritise AI’s perceived benefits over regulation and risk mitigation.

Representing the speed with which AI policy and regulatory approaches can change, the US has retreated from the light-touch, market-based AI regulation it had developed in the period leading up to the election of its current administration at the end of 2024. Its replacement, the *US AI Action Plan*²⁴ is, *inter alia*, designed to remove bureaucratic red tape and ensure that AI development is not smothered in bureaucracy.²⁵ It is expressly anti-regulatory with the caveat that AI should protect free speech and American values.²⁶

²³ Part 4A comes into effect in December 2025.

²⁴ The White House, *Winning the Race: America’s AI Action Plan*, July 2025. See <<https://www.whitehouse.gov/wp-content/uploads/2025/07/Americas-AI-Action-Plan.pdf>>.

²⁵ Ibid p 3.

²⁶ Ibid p 4.

While market-based self-regulation has been the norm for the technology industry, the EU's regulatory posture, particularly over the last few years, has been to regulate online-related technology risks more actively,²⁷ including through specific AI regulation, the EU Artificial Intelligence Act (EU AI Act).²⁸ The EU AI Act provides risk-based regulation of *AI systems*. As noted above, it applies to second wave AI.²⁹ Until recently, its risk-based approach was viewed as the template, 'gold standard' for international AI regulation to the extent that Margot Kaminsky, amongst others, asserted that international AI governance was converging around this type of model.³⁰

In Australia, regulatory options for AI have generally been characterised as a choice between the US and the EU approaches. China's approach has largely been overlooked. Incremental and iterative in nature it shares a number of features with the EU approach, including specific transparency measures such as algorithm registration and training data disclosure. It emphasises the need for AI to develop in accordance with the rule of law and adopts measures to safeguard personal data and copyright as well as other initiatives designed to protect individuals from AI harms. Although China's regulatory architecture is implemented differently to the EU's, they share many common features and preoccupations. Importantly, China's AI regulation demonstrates that a vibrant, innovative AI sector that is challenging and, some argue surpassing, US AI technology has been developed in a regulated environment: Chinese experience tends to disprove the argument that regulation and innovation are incompatible and, from an Australian perspective, invites an inquiry into how well designed and thoughtful regulation can support innovation *and* provide appropriate protections from its harms.

Recent developments cast doubt on an international risk-based regulatory consensus. The most notable of these are the US *AI Action Plan*³¹ and China's *Artificial Intelligence Global Governance Action Plan*,³² which both emphasise the need to accelerate the development of

²⁷ Examples include the development of EU-wide regulatory measures such as the *General Data Protection Regulation* (GDPR) in 2018, which regulates data protection; the *Digital Services Act* in 2022, which regulates digital platforms and is designed to prevent illegal and harmful online activities and the spread of disinformation; the *Digital Markets Act* in 2022, which is designed to regulate 'gatekeeper' powers of large digital platforms and applies specific new competition laws that supplement existing EU competition laws.

²⁸ EU Regulation 2024/1689 laying down harmonised rules on artificial intelligence, 13 June 2024. See < https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=OJ:L_202401689 >.

²⁹ By recital 12, the EU AI Act is not intended to apply to 'systems that are based on the rules defined solely by natural persons to automatically execute operations.'

³⁰ Margot Kaminsky, 'Regulating the Risks of AI' (2023) 103 Boston University Law Review 1347, at 1351.

³¹ US, White House, *Winning the Race: America's Action Plan*, July 2025. See < <https://www.whitehouse.gov/wp-content/uploads/2025/07/Americas-AI-Action-Plan.pdf> >.

³² Central People's Government of the People's Republic of China, *Artificial Intelligence Global Governance Action Plan*, 26 July 2025. See < https://www.gov.cn/yaowen/liebiao/202507/content_7033929.htm >.

AI but differ as to whether they support a unilateral approach (US) or a global approach to building consensus standards and norms (China). Canada and the UK have recently abandoned the development of EU-style regulation.³³

The current approach to AI regulation in Australia

Like Canada and the UK, there are clear signs that Australia is considering a retreat from risk-based AI regulation. Up until August 2025, Australia's policy work on AI regulation had been undertaken primarily by the Department of Industry Science and Resources (DISR), which produced a *Proposals Paper for Introducing Mandatory Guardrails for AI in High-Risk Settings* in 2024³⁴ (2024 Proposals Paper) after several years of consultation. The government has not committed to implementing it.

The 2024 Proposals Paper argued that 'effective regulation'³⁵ was considered 'essential to securing the benefits of AI.'³⁶ Risk-based regulation was considered desirable, consisting of a mix of *ex ante* and *ex post* initiatives but with an emphasis on *ex ante* measures aimed at avoiding catastrophic harm before it occurs³⁷ and focused on high-risk applications.³⁸ Ten mandatory 'guardrails' were proposed to apply to organisations that develop or deploy high-risk AI systems and general purpose AI (GPAI) models.³⁹ The intent of the overall approach was to align 'with likeminded jurisdictions (to) reduce the compliance burden for both industry and government.'⁴⁰ The EU's AI Act and Canada's proposed *Artificial Intelligence Data Act* (AIDA) were given as examples of regulatory approaches with which Australia should align.⁴¹ Reporting in November 2024, the Australian Senate Select Committee on Adopting Artificial Intelligence broadly supported this approach, stating that 'the Australian Government (should) introduce new, whole-of-economy, dedicated legislation to regulate high-risk uses of AI.'⁴²

³³ These are discussed in greater detail in chapter 5.

³⁴ Australian Government, Department of Industry, Science and Resources, *Safe and Responsible AI in Australia: Proposals Paper for Introducing Mandatory Guardrails for AI in High-Risk Settings*, September 2024. See <https://storage.googleapis.com/converlens-au-industry/industry/p/prj2f6f02ebfe6a8190c7bdc/page/proposals_paper_for_introducing_mandatory_guardrails_for_ai_in_high_risk_settings.pdf>.

³⁵ See *ibid* n 2, p 5.

³⁶ *Ibid*.

³⁷ *Ibid* p 16.

³⁸ *Ibid* p 18.

³⁹ *Ibid* p 35.

⁴⁰ *Ibid* p 31.

⁴¹ The Australian Senate Select Committee on Adopting Artificial Intelligence broadly supported this approach, stating that 'the Australian Government (should) introduce new, whole-of-economy, dedicated legislation to regulate high-risk uses of AI.' Australian Senate, *Senate Select Committee on Adopting Artificial Intelligence*, November 2024, pp 46-47. See <[https://parlinfo.aph.gov.au/parlInfo/download/committees/reportsen/RB000470/toc_pdf/SelectCommitteeonAdoptingArtificialIntelligence\(AI\).pdf](https://parlinfo.aph.gov.au/parlInfo/download/committees/reportsen/RB000470/toc_pdf/SelectCommitteeonAdoptingArtificialIntelligence(AI).pdf)>.

⁴² *Ibid* pp 46-47.

However, in the period since then there has been little evidence of an Australian consensus favouring a risk-based or guardrails-type approach. Following the publication of the 2024 Proposals Paper, the Productivity Commission was tasked with undertaking a ‘Five Pillars Productivity Inquiry’⁴³ to inform the Commonwealth government’s efforts to boost economic productivity.⁴⁴ Pillar four related to ‘Harnessing Data and Digital Technology’ and included consideration of how new technologies such as AI and making better use of data can improve productivity. The Productivity Commission’s Interim Report on Harnessing Data and Digital Technology (PC Interim Report)⁴⁵ recommended that the Australian government ‘pause steps to implement mandatory guardrails for high-risk AI.’⁴⁶ Consultation on the PC Interim Report is underway with a final report to government due at the end of 2025.

Both the Commission and the Commonwealth have signalled that interest in a risk-based regulatory approach is waning. One of the Commissioners involved in developing the Pillar four work stream foreshadowed that the best way to regulate AI might be not to specifically regulate AI at all.⁴⁷ The Productivity Commission Interim Report recommends that ‘AI-specific regulation should be a last resort’⁴⁸ and that ‘[t]he Australian Government should only apply the ‘mandatory guardrails for high-risk AI’ in circumstances that lead to harms that cannot be mitigated by existing regulatory frameworks *and* where new technology-neutral regulation is not possible.’⁴⁹

Internationally and nationally, the AI regulatory landscape is, like AI itself, in a state of flux. Any consensus around EU-style AI risk-based regulation has either stalled or, more likely, is breaking down as governments worldwide increasingly prioritise the economic and productivity benefits of AI over regulation that might limit these benefits.

Other Australian initiatives have been progressing in parallel. The National Transport Commission has begun a project to ‘explore ways that artificial intelligence (AI) can be used to harmonise rail operating rules and reduce the regulatory and training burden on rail

⁴³ Australian Government, Productivity Commission, *Five Pillars of Productivity Inquiries*. See <https://www.pc.gov.au/inquiries/current/five-productivity-inquiries>.

⁴⁴ The Hon. Dr Jim Chalmers, Treasurer, *Public Consultation on Productivity Reform Priorities*, 19 May 2025. See <https://ministers.treasury.gov.au/ministers/jim-chalmers-2022/media-releases/public-consultation-productivity-reform-priorities>

⁴⁵ See *Harnessing Data and Digital Technology Interim Report* (above n 5) p 2.

⁴⁶ *Ibid* p 3.

⁴⁷ Stephen King, Commissioner, *The Best way to Regulate AI might be not to Specifically Regulate AI. Here’s why*, 20 September 2024. See <<https://www.pc.gov.au/media-speeches/articles/best-way-to-regulate-ai>>.

⁴⁸ *Harnessing Data and Digital Technology Interim Report* (above n 45) p 2.

⁴⁹ *Ibid* p 3 (my emphasis).

workers.⁵⁰ It has also undertaken a number of consultations on automated vehicle regulation.⁵¹

The Department of Health, Disability and Ageing published its AI inquiry, *Safe and Responsible Artificial Intelligence in Health Care – Legislation and Regulation Review Final Report* in March 2025 (DHDA AI Report).⁵² It concludes that:

AI technologies may present a significant productivity and public benefit opportunity for the Australian health, disability and aged care ecosystem and it is essential that these benefits are equitably realised through targeted innovation pathways and appropriate assurance mechanisms. Given health and care are high-risk settings for the deployment of AI and the rapid adoption of AI technologies across the sector have given rise to risks that must be responsibly managed to protect Australians and support safe use.⁵³

Published before the Productivity Commission Interim Report, the DHDA AI Report supports ‘economy-wide safeguards such as the proposed mandatory AI guardrails.’⁵⁴ It also highlights the difficulties associated with managing AI risks in healthcare settings where responsibility for delivering health and related services is distributed across the Commonwealth, states and territories and the private sector.

The Commonwealth does not deliver health services. This responsibility is primarily undertaken by the states and territories supplemented by the private sector. The Commonwealth’s role, and the health legislation it administers, is largely ‘administrative in nature’⁵⁵ such as the provision of medical benefits, the Pharmaceutical Benefits Scheme and My Health Record or the provision of program funding. Commonwealth ‘legislation is likely to require minor and technical amendments for clarity,’⁵⁶ but state and territory legislation for the use of AI in delivering healthcare will also need to be considered. The implications for AI regulation in Australia’s federal system of government where the states and territories are mainly responsible for delivering key services such as public, primary and acute health services, education and law enforcement has not figured prominently in the Australian legal

⁵⁰ National Transport Commission, *National Rail Action Plan Newsletter: April 2025*, May 2025. See <<https://www.ntc.gov.au/news/national-rail-action-plan-newsletter-april-2025>>.

⁵¹ For example, National Transport Commission, *The Regulatory Framework for Automated Vehicles in Australia*, February 2022. See <<https://www.ntc.gov.au/sites/default/files/assets/files/NTC%20Policy%20Paper%20-%20regulatory%20framework%20for%20automated%20vehicles%20in%20Australia.pdf>>.

⁵² Australian Government, Department of Health, Disability and Ageing, *Safe and Responsible Artificial Intelligence in Health Care – Legislation and Regulation Review Final Report*, March 2025. See <<https://www.health.gov.au/sites/default/files/2025-07/safe-and-responsible-artificial-intelligence-in-health-care-legislation-and-regulation-review-final-report.pdf>>.

⁵³ Ibid p 19.

⁵⁴ Ibid p 3.

⁵⁵ Ibid p 4.

⁵⁶ Ibid.

and regulatory debate. It raises the possibility of a patchwork approach to AI regulation that is jurisdiction and sector based.

The development of an Australian normative framework for AI regulation will inevitably await the completion of the Productivity Commission's work, and even then, in practical terms, a comprehensive regulatory package is unlikely before 2027.

Arguably, there are several reasons for exercising caution in enacting risk regulation for AI on its own. AI technology is complex and fast moving. Issues and risks selected now may be inappropriate as AI develops. Risk approaches work backwards from identifying and categorising levels of risk in advance and then applying legal and regulatory consequences to each. But selecting and prioritising risks is not straightforward when it comes to AI. Equally, assigning interventions and outcomes for each category of risk is a complex undertaking, one for which the EU AI Act has been widely criticised.⁵⁷ One argument is that its complexity will give rise to high compliance costs, thereby favouring large, well-resourced organisations at the expense of innovative small to medium enterprises. This criticism applies particularly to its risk classification rules and the uncertainties associated with risk self-assessments under Article 19 of the AI Act. That said, there may be elements of AI regulation where risk regulation is helpful and appropriate. One example would be having *ex ante* measures that operate to mitigate risks and which would constitute admissible evidence that reasonable steps have been taken to comply with a duty of care as a defence to an action for negligence.

Transposing a risk-based regulatory approach as contained in the EU AI Act into an Australian context is also problematic when the EU's overarching purposes do not directly translate into an Australian context. The EU AI Act's purposes encompass promoting the uptake of AI while ensuring a high level of protection of 'fundamental rights as enshrined in the Charter of Fundamental Rights of the European Union (the 'Charter'), including democracy (and) the rule of law.'⁵⁸ Respect for and protection of fundamental rights is a cornerstone of the EU's regulatory posture. However, unlike the EU and any of the other OECD members, Australia lacks a human rights framework that is embedded in Constitutional protections. Australia's human rights framework consists of a range of statutes that can be amended by the

⁵⁷ See, for example, Sandra Wachter, 'Limitations and Loopholes in the EU AI Act and the AI Liability Directives: What this means for the European Union, the United States, and Beyond' (2024) 26(3) *Yale Journal of Law and Technology* 671.

⁵⁸ EU AI Act, Recital 1.

legislature.⁵⁹ In the face of these limitations it is not clear that a risk-based approach anchored in fundamental rights can be readily transposed into an Australian legal context.

This thesis suggests a possible, alternative way forward, which involves making appropriate adjustments to existing legal regimes, coupled with a mandatory registration of algorithms and training datasets, based on initiatives undertaken by China.

Developing a regulatory approach for AI in Australia: a wicked problem?

AI does not manifest into a regulatory vacuum. In Australia, the common law and existing statute law apply. The fundamental issues are whether that existing legal framework adequately applies to AI and whether it comprehensively accounts for new AI-instantiated risks. If it does not, what adjustments are necessary? To date, there has been no gap analysis undertaken in Australia that distinguishes between existing protections against projected new risks and the challenges these pose for the legal system.

Decisions about how to regulate AI in Australia involve socio-political and economic assessments that inform a public policy process. Noting that legal instruments and interventions are one of the five main public policy instruments,⁶⁰ the design of legal and regulatory solutions needs to account for the fact that AI – or at the very least key aspects of AI – present as ‘wicked problems.’ A wicked problem is not just a difficult problem. A wicked problem is one that either cannot be defined or, at best, is ‘not open to easy formulation’.⁶¹ It creates difficulties in determining possible solutions because the problem is shifting or ‘there is no way of learning about the issue without trying potential ‘answers’ that come with unintended consequences.’⁶² As Arie Freiberg argues, it is ‘a problem that is complex, is difficult to define, has many interdependencies, may be unstable and evolving, has no clear and correct solutions, involves many stakeholders, and features responsibilities that rest with many organisations.’⁶³

Are the perceived economic and societal benefits of AI of such a magnitude that it should receive special regulatory treatment? In other words, should existing legal and regulatory protections that apply in the real, non-AI domain, be attenuated in the AI domain so that

⁵⁹ For the purposes of this thesis, State-based human rights charters have not been taken into account.

⁶⁰ These are discussed in Catherine Althus, Peter Bridgeman and Glyn Davis, *The Australian Policy Handbook 4th ed*, Allen & Unwin, 2007, Chapter 6.

⁶¹ *Ibid* p 54.

⁶² *Ibid*.

⁶³ Freiberg (above n 9) p 157.

there is a dual system of liability that imposes a higher legal standard on humans and a lesser one on AI, meaning that AI is *less accountable* than humans when they undertake the same or similar activities? What happens when decision-making is undertaken jointly?

Second wave AI is trained on data. 'Training data is AI's oxygen: an AI system cannot exist without it.'⁶⁴ Overwhelmingly, that data has been either scraped from the internet without permission or compensation or extracted from individuals using opaque contractual provisions and privacy policies that permit such use. An AI system cannot function without that data. Thus, uniquely amongst all commercial endeavours, the developers of machine learning-based AI appropriate the key resource it needs to function. Should they be entitled to what amounts to a 'free ride.?' Also missing from the Australian AI debate is any discussion about who, and in what circumstances, should bear the cost of AI when it goes wrong and causes harm. Overwhelmingly, the current system for liability for causing unacceptable harm is fault-based, for example through the law of negligence. In more limited circumstances, strict liability applies where the public policy concludes that, because of the serious nature of the harm involved, causation-based liability is preferable. This type of debate is usually informed by making an assessment about who should bear the cost of AI-caused harm. Should those who profit from developing and deploying AI bear the cost of the harm it causes, or should society do so instead? A mature discussion of these issues has not occurred in Australia resulting in an impoverished regulatory debate that oversimplifies the associated wicked problems.

Research question

It is the aim of this thesis to elucidate an optimal approach to the design of a regulatory regime for AI in Australia. In order to do so, the secondary questions, which will be addressed in the subsequent chapters of this thesis, are as follows.

- (1) What are the specific features of current AI and the foundational risks and harms flowing from the use of AI that need to be appreciated in designing an appropriate regulatory framework for Australia?

⁶⁴ Jennifer King and Caroline Meinhardt, 'Rethinking Privacy in the AI Era: Policy Provocations for a Data-Centric World,' Stanford University Human-Centred Artificial Intelligence White Paper, 22 February 2024, p 37. See <<https://hai.stanford.edu/policy/white-paper-rethinking-privacy-ai-era-policy-provocations-data-centric-world>>.

- (2) What considerations should be taken into account in developing a regulatory framework for Australia, having regard to the broader context of prior ICT regulation and key international developments in AI governance?
- (3) What insights are available from key international developments in relation to the governance of AI?
- (4) How might it be possible to address the risks and harms posed by AI without enacting a *sui generis*, risk-based regulatory regime as originally proposed for Australia?

Thesis methodology

The methodology of this research project is doctrinal. Doctrinal research in legal scholarship is characterised as a 'library-based' research design, which aims to develop solutions to an identified legal problem.⁶⁵

This research project draws on the disciplines of law and regulatory studies. The sources consulted for this research project include technical, legal and regulatory texts as well as research and analysis undertaken by international organisations and government and government-sponsored inquiries and reports. As its focus is on a regime suitable for Australia, it prioritises Australian research into AI risks. It also includes some comparative analysis, drawing the insights available from key overseas developments relating to the governance of AI.

Thesis structure

This thesis proceeds as follows. Chapter two provides a high-level account of AI in its main forms in order to address the requirement identified by Matthew Scherer that: '[a]ny AI regulatory regime must define what exactly it is that the regime regulates; in other words it must define artificial intelligence.'⁶⁶ Chapter three examines AI risks and harms, prioritising foundational issues and relevant Australian research, with a view to providing as broad a description as possible of the AI risks that require attention in Australia. Chapter four outlines the background to technology regulation in general, and for AI in particular. It notes the development of the prevailing 'techno-libertarian' approach to regulating the technology industry and its immense international political and economic power that has long led to it

⁶⁵ Terry Hutchinson, 'Doctrinal Research: Researching the Jury' in Dawn Watkins and Mandy Burton (eds), *Research Methods in Law*, Taylor & Francis Group, 2013, p 28.

⁶⁶ Matthew Scherer, 'Regulating Artificial Intelligence Systems: Risks, Challenges, Competencies, and Strategies' (2016) 29(2) *Harvard Journal of Law and Technology* 354,359.

avoiding or minimising regulation. The chapter argues that many of the dilemmas associated with the regulatory debate concerning AI are a residue of unsolved regulatory issues applicable to the internet. It also traces the contours of the most common argument used against technology regulation, that regulation impedes innovation, and subjects that argument to critical analysis. Chapter five discusses and analyses the insights available from an analysis of key overseas developments in relation to the regulation of AI, including the latest trend favouring a more pro-innovation approach.

Chapter six suggests a way forward for Australia. A starting point for considering an Australian approach to AI regulation is to understand what AI is and what its claims are. John McCarthy, one of the founders of AI, thought that it is the science and engineering of making intelligent machines.⁶⁷ Although a philosophical discussion about what constitutes human intelligence is beyond the scope of this thesis, one of the central claims of AI is that it is a suite of technologies that aims to emulate what humans can do but, in many cases, faster and more accurately: it can perform tasks that normally require human intelligence such as solving problems, producing text, images and other content, answering questions, speech recognition, visual perception, learning and applying knowledge and skills.

Where humans undertake these activities, they are legally responsible and accountable for them. Certain activities they undertake might be prohibited and constitute criminal offences such as cyber bullying.⁶⁸ Some might be illegal under discrimination law while others may be regulated under the *Privacy Act 1988 (Cth)* (Privacy Act). In each case the decision to regulate has been determined following a public policy process that has led to the development of regulation. Typically, that regulation assumes that it governs people, not machines.⁶⁹

By parity of reasoning, if AI engages in the same conduct or undertakes similar tasks or activities as humans, it should be equally responsible and accountable and subject to the same substantive regulatory responsibilities. Adopting this idea of equivalence between human and artificial intelligence as a foundation for a legal and regulatory investigation into

⁶⁷ Christopher Manning, Stanford University, Human-Centered Artificial Intelligence, 'Artificial Intelligence Definitions', 2020. See <<https://hai-production.s3.amazonaws.com/files/2020-09/AI-Definitions-HAI.pdf>>.

⁶⁸ For example, *Criminal Code Act 1995 (Cth)*, s 474.17.

⁶⁹ It is acknowledged that the common law has also evolved to cover new forms of liability, notably through the development of the tort of negligence and other forms of liability. Although the common law has a role to play in addressing AI liability and regulation, it is unlikely to do so at a pace that will provide an adequately timely basis for AI regulation.

AI means that it is anchored in existing and accepted policy positions. I refer to this as the principle of equivalence.

Accompanying this principle of equivalence is another fundamental principle of the Australian legal system: the rule of law. The rule of law mandates that the law applies equally and to all. The 'core of the existing principle is ... that all persons and authorities within the state, whether public or private, should be bound by and entitled to the benefit of laws publicly made, taking effect (generally) in the future and publicly administered by the courts.'⁷⁰ There is no reason why 'artificial persons' should be exempt from the rule of law if undertaking functions that natural persons would, could or might undertake. The fact that the artificial intelligence might be faster, more accurate or function at a scale beyond the capacity of humans makes no in-principle difference. The rule of law does not apply differentially to humans who can complete mathematical equations or any other activities faster or more accurately than others.

It also invites an inquiry into what, if any, additional measures need to be taken to account for the differences between the human and the artificial to preserve that equivalence or to enhance it if thought necessary. As stated above, AI does not manifest into a regulatory vacuum. Existing laws continue to apply to it. There is no reason that they should be discarded or modified unless they do not adequately cover AI. In other words, what is different about AI that necessitates additional, specific, *sui generis* legal and regulatory interventions? The Productivity Commission refers to this task as a 'gap analysis'⁷¹ and notes that this exercise has not been undertaken.

There is much to be learnt from examining international approaches to AI regulation. Of significance is that both the Chinese and EU regulatory systems apply to AI supplied into their markets. Any Australian AI developer will, by virtue of the 'Brussels effect' and a developing 'Beijing effect' be compelled to comply with their regulatory requirements if it wishes to enter those markets. Although US federal AI policy has become anti-regulatory, a number of US states are developing their own AI regulation that will also apply any AI

⁷⁰ Tom Bingham, *The Rule of Law*, London, 2010, p 8. It is noted that theoretical arguments about the concept of the rule of law and what it entails abound and are contested. A useful discussion of these issues in an Australian context can be found in a paper delivered by Bell P., 'The Rule of Law and the Constitution – a short overview', 2021 CCCS Constitutional Law Conference, 23 July 2021. See <https://supremecourt.nsw.gov.au/documents/Publications/Speeches/2021-Speeches/Bell_20210723.pdf>. A discussion of these arguments is beyond the scope of this thesis.

⁷¹ Australian Government, Productivity Commission, *Harnessing Data and Digital Technology Interim Report* (above n 6) pp 17-18.

supplied into their territories. In a practical sense, any Australian AI supplier may choose to comply with the highest level of regulation as a means of complying with them all.

Ultimately this thesis argues that Australia is best served by applying existing laws to AI and adapting them, where necessary, to account for the differences between AI and human agency. This finetuning will not be without difficulties, particularly for Australian data protection law. The thesis also argues that new laws, based on the Chinese regulatory scheme that requires algorithm registration and training data disclosure, should be implemented to encourage transparency and accountability. To ensure that there are adequate transparency incentives, compliance with these registration and disclosure requirements should be a pre-condition of, and function as a safe harbour for, continued fault-based liability in tort or under product or service liability laws. In the absence of such compliance, strict liability should apply to ensure that end users have practical, enforceable rights where AI causes damage and to hold AI to account.

Chapter 2 – What is Artificial Intelligence: Concepts, Terminologies, Technologies and Definitions

Introduction

It is not possible to undertake a legal, legal policy or regulatory analysis of AI without understanding and defining what it is.

This chapter traces the contours of ‘AI’ and identifies key concepts and technologies. It notes that AI consists of a variety of computational techniques and methods designed to solve problems, make decisions and engage in activities in a way that emulates human capabilities. In doing so, it briefly explains how AI has developed and identifies its main elements.

This background work is necessary because one problematic aspect of the current debate about AI regulation is that it focuses on the relatively recent proliferation of machine learning-based, large language model generative AI (GAI) such as ChatGPT⁷² almost to the exclusion of any others. But GAI is only one type of AI. Taking an overly narrow focus excises other forms of AI that can be harmful from the regulatory debate. AI is not homogenous. It is a ‘collection of interrelated technologies used for problem solving and to complete tasks that would otherwise require human intelligence.’⁷³ There are many different techniques and approaches that can be used to develop and deploy AI. These are not uniform. To account for these and properly frame a regulatory analysis it is necessary to consider AI broadly and to arrive at a workable and clear definition of what it is and is not.

There are a variety of ways to account for AI. One option is to explain it by reference to its evolution or history. Another is to describe its technologies and computational techniques. Yet another is to examine the disciplines it draws upon, which are embodied into machines to enable them to emulate human activity. None on its own constitutes a comprehensive basis for understanding the legal and regulatory issues that are engaged by AI. To avoid this problem, the analytic lens adopted in this thesis is to describe each. I suggest that a helpful

⁷² ChatGPT is an acronym for Chat Generative Pre-Trained Transformer. The ChatGPT web site is <<https://chatgpt.com>>.

⁷³ Genevieve Bell, Jean Burgess, Julian Thomas, Shazia Sadiq, (24 March 2023), *Rapid Response Information Report, Generative AI - Language Models and Multimodal Foundation Models*, Australian Council of Learned Academies, p 22. See <<https://acola.org/wp-content/uploads/2023/06/acola-rapid-response-information-report-generative-AI.pdf>>.

way in which to frame definitional issues is to consider AI as consisting of successive ‘waves’ of functionalities that evolved, beginning in the late 1940s.

Terminology and the ‘odd paradox’

Multiple terms may be used to describe AI or its key components, not always with precision. Examples include ‘Big Data’⁷⁴ and ‘Advanced Analytics,’ the former often being used to describe large data sets accompanied by advanced computing techniques, the latter being used to describe ‘a collection of sophisticated techniques and tools that are used to analyze large volumes of data, uncover hidden patterns and provide actionable insights.’⁷⁵ In other words, computing techniques that encompass the statistical and probability models that underpin machine learning. These terms are often appropriated by those who market AI, obscuring their content and overall relevance to AI.⁷⁶

A final consideration concerns Pamela McCorduck’s ‘odd paradox,’ whereby computational techniques that were once considered new and innovative become viewed as commonplace as newer techniques are developed.⁷⁷ What seemed ‘intelligent’ at one point in time (for example, machines that can surpass the best chess players or predictive text), quickly recedes into the background as more complex functionalities evolve.⁷⁸ This is a key reason why GAI receives so much attention at the expense of ‘weak or AI. As one of the founders of AI, John McCarthy, is reported as saying ‘[a]s soon as it works, no one calls it AI anymore.’⁷⁹ However, this is not helpful when conducting a legal or regulatory analysis of AI.

⁷⁴ The European Commission states that big data ‘refers to large amounts of data produced very quickly by a high number of diverse sources. Data can either be created by people or generated by machines, such as sensors gathering climate information, satellite imagery, digital pictures and videos, purchase transaction records, GPS signals, etc. It covers many sectors, from healthcare to transport and energy. See European Commission, ‘Digital Single Market Policy, Big data’, (January 2020). See <<https://ec.europa.eu/digital-single-market/en/big-data>>. There are several other definitions or descriptions. One of the most prominent is that developed by the US National Institute of Science and Technology (NIST), which states that Big Data is a ‘term used to describe the large amount of data in the networked, digitized, sensor-laden, information-driven world’ and, additionally, that it ‘refers to the need to parallelize the data handling in data-intensive applications.’ See National Institute for Science and Technology, *NIST Big Data Interoperability Framework: Volume 1, Definitions* (version 3) October 2019, p iii. See <<https://nvlpubs.nist.gov/nistpubs/SpecialPublications/NIST.SP.1500-1r2.pdf>>. The European Commission definition is preferred as it does not conflate the magnitude of data with computing techniques. These are conceptually different concepts.

⁷⁵ Cole Stryker, IBM, ‘What is Advanced Analytics?’ 10 July 2024. See <<https://www.ibm.com/think/topics/advanced-analytics>>.

⁷⁶ The phenomenon of the emerging technological ‘hype cycle’ is well known through the research work undertaken by a variety of advisory firms such as the Gartner Hype Cycle. See <<https://www.gartner.com.au/en/methodologies/gartner-hype-cycle>>.

⁷⁷ Pamela McCorduck, *Machines Who Think: A Personal Inquiry into the History and Prospects of Artificial Intelligence*, Boca Raton, 2019, p 423. This phenomenon and the ‘hype cycle’ referred to in the previous paragraph are linked.

⁷⁸ This phenomenon is also noted by the Organisation for Economic Cooperation and Development in its discussion about ‘What is an AI system?’ See <<https://oecd.ai/en/ai-principles>>.

⁷⁹ Moshe Vardi, ‘Artificial Intelligence: Past and Future,’ (2012) 55 Communications of the ACM, p 5.

Early development

There is nothing new about machines that can emulate human capabilities. Within western traditions, Pascal developed a mechanical calculating machine, the Pascaline, that could add and subtract in the mid-seventeenth century.⁸⁰ Leibniz improved it by adding multiplication and division functions.⁸¹ In the early nineteenth century, Babbage designed an automatic mechanical calculator, the Difference Engine, to generate error-free logarithmic tables.⁸² He later developed a steam and mechanically powered machine, the Analytical Engine, which included a mechanical equivalent of a memory and processor as well as an ability to interpret programming instructions using punched paper cards.⁸³ Hollerith augmented punched card technology so that it read and stored data embodied in cards and undertook mathematical calculations using the stored data powered by electromechanical means.⁸⁴ Hollerith's technology was used in successive US Censuses beginning in 1890 and, later, in Europe. His company eventually amalgamated with others to become International Business Machines (IBM).

Electronics then replaced mechanical processes offering broader, more flexible and powerful computational functionalities. Although there is debate about when the first general purpose electronic computational machine was developed, many consider that Colossus, developed in England to assist in decrypting German enigma cyphers, was the starting point.

Origins of AI

Modern AI dates to theoretical work undertaken in the late 1930s and 1940s:

The first work that is now generally recognised as AI was done by Warren McCulloch and Walter Pitts (1943) ... they drew on three sources: knowledge of the basic physiology and functions of neurons in the brain; a formal analysis of propositional logic due to Russell and Whitehead; and Turing's theory of computation.⁸⁵

In a 1950 paper,⁸⁶ Alan Turing posed the question, 'Can machines think?' He proposed the 'Imitation Game'⁸⁷ which involved an evaluator engaging in a text-based conversation with a human and a machine: if the evaluator could not discern a difference between their

⁸⁰ HistoryofInformation.com, 'Blaise Pascal invents a calculator.' See <<https://historyofinformation.com/detail.php?entryid=440>>.

⁸¹ HistoryofInformation.com, 'Leibniz Invents the Stepped Drum Gear Calculator.' See <<https://historyofinformation.com/detail.php?id=394>>.

⁸² Oscar Schwartz, 'Untold History of AI: When Charles Babbage Played Chess with the Mechanical Turk,' IEEE Spectrum, 10 March 2019. See <<https://spectrum.ieee.org/untold-history-of-ai-charles-babbage-and-the-turk>>.

⁸³ Ibid.

⁸⁴ Geoffrey Austrian, *Herman Hollerith: The Forgotten Giant of Information Processing*, Columbia University Press, 1982, pp15, 124, 418.

⁸⁵ Stuart Russell and Peter Norvig, *Artificial Intelligence: A Modern Approach*, Global edition, Pearson, 2021, p35.

⁸⁶ Alan Turing, 'Computing Machinery and Intelligence,' *Mind*, Vol. LIX. No.236., Oct 1950.

⁸⁷ The Imitation Game is also known as the Turing Test.

responses the machine won the Imitation Game and was ‘intelligent’: it had successfully imitated human capabilities. Turing’s paper also introduced concepts such as machine learning and reinforcement learning.⁸⁸ In doing so he suggested that it would be preferable to ‘create human-level AI by developing learning algorithms and *then teaching the machine* rather than by programming its intelligence by hand.’⁸⁹

The first reference to the term ‘Artificial Intelligence’ occurred at a Dartmouth College workshop sponsored by the US Defence Advanced Research Projects Agency (DARPA) in 1956. It proposed:

... a 2 month, 10 man study of artificial intelligence be carried out during the summer of 1956. The study is to proceed on the basis of the conjecture that *every aspect of learning or any other feature of intelligence* can be in principle so precisely *described* that a machine can be made to *simulate* it ...⁹⁰

Although the conference itself did not lead to significant AI breakthroughs, it set the scene for the development of what became known as ‘Weak AI.’⁹¹

Weak AI: first wave

Following the Dartmouth conference AI researchers focused on developing machines that solved ‘tasks considered indicative of intelligence in humans, including games, puzzles, mathematics, and IQ tests.’⁹² These types of AI involved ‘a general-purpose search mechanism trying to string together elementary reasoning steps to find complete solutions.’⁹³ But methods that were successful in limited domains of reasoning failed on more difficult problems that involved more complex reasoning.

Weak AI’s ‘intelligence’ is programmed into it. Its functionality does not extend to ‘learning.’ It is referred to as weak because its techniques:

... do not scale up to large or difficult problem instances. The alternative to weak methods is to use more powerful, domain-specific knowledge that allows larger reasoning steps and can more easily handle typically occurring cases in a narrow area of expertise. One might say that to solve a hard problem, you have to almost know the answer already.⁹⁴

⁸⁸ The significance and meaning of these concepts are discussed later in this chapter.

⁸⁹ Stuart Russell and Peter Norvig, (above n 85) p36. (My emphasis).

⁹⁰ John McCarthy, Marvin Minsky, Nat Rochester, Claude Shannon, *A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence*, 31 August 1955. See <<https://raysolomonoff.com/dartmouth/boxa/dart564props.pdf>>. (My emphasis).

⁹¹ Stuart Russell and Peter Norvig, (above n 85) p 36.

⁹² Ibid.

⁹³ Ibid, p 40.

⁹⁴ Ibid, p 40.

One of the most successful instantiations of more powerful, domain-specific knowledge focused on a narrow area of expertise were expert systems. These involve the use of a knowledge base and an inference engine to solve specific problems, often including decision trees. Examples of these include various income tax advisor services used in the 1970s and 1980s.⁹⁵ Although weak and limited in scope, expert systems can undertake calculations and make decisions faster and more accurately than humans, thus emulating human capabilities through their apparent ability to reason. Significantly, the Robodebt programme presented as a case study in Chapter 3, was an example of a very simple expert system that caused significant damage and required a Royal Commission to unravel what went wrong. That it was able to do so was facilitated by *ex ante* transparency where the decision-making could be explained and its illegality determined.⁹⁶

From Weak first wave to narrow second wave

AI research progressed in spasmodically in the 1960s and 1970s. There were periods of AI optimism followed by failures to meet expectations leading to several AI ‘winters’ where progress slowed. However, many of the approaches and technologies that now underpin advanced AI were invented by the late 1980s. For example, the first computer neural network, the Perceptron, was developed in 1957; the first pattern recognition algorithms were written in 1967, and Explanation Based Learning (EBL) that enabled computers to analyse training data, a necessary requirement for machine learning AI using large language models, was first developed in 1981.⁹⁷

Narrow AI: second wave

Second wave AI is based on machine learning techniques designed to overcome the limitations of weak AI combined with advances in computing capabilities, the greater availability of machine learning software, more data sources and greater access to them:

[c]omputer performance increases over the last decade have enabled the success of machine learning, in combination with large data sets, and software libraries.⁹⁸

⁹⁵ See, by way of background, David Brock, *Learning from Artificial Intelligence's Previous Awakenings: The History of Expert Systems*, Association for the Advancement of Artificial Intelligence, 39 (3) AI Magazine, 2018

⁹⁶ Royal Commission into the Robodebt Scheme, *Report of the Royal Commission into the Robodebt Scheme*, 7 July 2023. See <<https://robodebt.royalcommission.gov.au>>.

⁹⁷ Bernard Marr, ‘A Short History of Machine Learning – Every Manager Should Read,’ Forbes, 19 February 2016. See <<https://www.forbes.com/sites/bernardmarr/2016/02/19/a-short-history-of-machine-learning-every-manager-should-read/>>.

⁹⁸ Defence Advanced Research Projects Agency, ‘AI Next Campaign.’ See <<https://www.darpa.mil/work-with-us/ai-next-campaign>>.

This second ‘wave’ of AI, sometimes also known as ‘narrow’ AI, is the dominant type of AI that currently preoccupies technologists, end users, policy makers and lawyers. It is the term used in this thesis for AI that can emulate ‘thinking’ and ‘learning’ and includes technologies such as GAI and Agentic AI.⁹⁹

The second rising wave of AI ... arrived ... after clear evidence that learning and perception capabilities are crucial for complex AI systems but missing in knowledge-based expert systems ... [The] second-generation AI paradigm was based on machine learning, which we now call *shallow* due to the lack of abstractions constructed by many-layer or “deep” representation of data ... In such shallow machine learning, engineers do not need to be concerned with constructing precise and exact rules as required for first-generation AI systems. Rather, they focus on statistical models, or simple neural networks as an underlying engine and then automatically learn or “tune” the parameters of the engine using the training data to make them handle uncertainty and generalize well from one condition to another and one domain to another.¹⁰⁰

The key distinguishing feature between first and second wave AI is that the latter is capable of ‘learning’ whilst the former is not. This capability is enabled by machine learning.

Machine learning

Machine learning:

... refers to making computers execute tasks through processes of ‘learning’ that derive inspiration from (but are not reducible to) human intelligence and decision making. Through analysis of large volumes of data and examples, ML algorithms can autonomously improve their learning over time. ML relies on algorithms ranging from basic decision trees through to artificial neural networks that classify information by mimicking the structure of the human brain. Predictive and anticipatory forms of ML are qualities that differentiate AI from previous forms of automation.¹⁰¹

This type of AI ‘involves using statistical learning and optimization methods that let computers analyse datasets and identify patterns. Machine learning techniques leverage data mining to identify historic trends and inform future models.’¹⁰²

There are three components to machine learning: an algorithm or a set of algorithms, training data and a model. An algorithm is a set of procedures. In machine learning, an algorithm learns to identify patterns after being trained on a large set of examples – the training data. Once a machine-learning algorithm has been trained, the result is a machine-learning model. The model is what people use.¹⁰³

⁹⁹ The OECD’s definition of AI referred to in n 68 comprehensively defines second wave AI.

¹⁰⁰ Li Deng, *Artificial Intelligence in the Rising Wave of Deep Learning: The Historical Path and Future Outlook*, (2018) 35 IEEE Signal Processing Magazine, Vol.1, p172 at p 173.

¹⁰¹ Toby Walsh, Neil Levy, Genevieve Bell, Anthony Elliott, James Maclaurin, Ivan Mareels, Fiona Wood, (2019) *The Effective and Ethical Development of Artificial Intelligence: An Opportunity to Improve our Wellbeing*, Report for the Australian Council of Learned Academics, p 34. See <https://acola.org/wp-content/uploads/2019/07/hs4_artificial-intelligence-report.pdf>.

¹⁰² University of California Berkeley, School of Information, ‘What is Machine Learning (ML)?’, 26 June 2020. See <<https://ischoolonline.berkeley.edu/blog/what-is-machine-learning/>>.

¹⁰³ Saurabh Bagchi, ‘What is a black box? A computer scientist explains what it means when the inner working of AI are hidden’, The Conversation, 22 May 2023. See <<https://theconversation.com/what-is-a-black-box-a-computer-scientist-explains-what-it-means-when->

The reliance of this type of technology on statistics and probability can frequently involve the logical fallacy that correlation equals causation.

Machine learning can undertake functions without being programmed to do so. It emulates learning based on statistics, probability and mathematical modelling.

For the machine to be ‘trained’ or to ‘learn,’ the algorithm is fed huge amounts of data; these training sets can affect whether the AI algorithm functions as it should and are (sic) compliant with regulations and standards.¹⁰⁴

Machine learning also encompasses ‘neural networks’ and ‘deep learning.’

Neural Networks

A neural network is a type of machine learning model:

... that makes decisions in a manner similar to the human brain, by using processes that *mimic the way biological neurons work* together to identify phenomena, weigh options and arrive at conclusions.¹⁰⁵

The network consists of layers of nodes (artificial neurons), including an input layer, hidden layers, and an output layer. Each node has a ‘weight’ associated with it mathematically using algorithms:

If the output of any individual node is above the threshold value, that node is activated, sending data to the next layer. Otherwise, no data is passed along to the next layer of the network.¹⁰⁶

For regulatory purposes, the exact functioning of a neural network for any individual output, such as a decision or recommendation, is practically impossible to discern.

Deep Learning

Closely related to neural networks is ‘deep learning.’ Whereas machine learning may use a few levels of simple neural networks, deep learning uses hundreds or thousands of layers to train the model.¹⁰⁷

Deep learning algorithms are incredibly complex, and there are different types of neural networks to address specific problems or datasets ... One potential weakness across them all

the-inner-workings-of-ais-are-hidden-203888>.

¹⁰⁴ Hannah Yee-Fen Lim, ‘Regulatory Compliance,’ p 86 in Charles Kerrigan (ed), *Artificial Intelligence: Law and Regulation*, Cheltenham, 2022, 86

¹⁰⁵ Fangfang Lee, IBM, ‘What is a Neural Network?’ 10 September 2025. See <<https://www.ibm.com/topics/neural-networks>>. (My emphasis).

¹⁰⁶ Ibid.

¹⁰⁷ Jim Holdsworth and Mark Scapicchio, IBM, ‘What is Deep Learning?’, 17 June 2024. See <<https://www.ibm.com/topics/deep-learning>>.

is that deep learning models are often “black boxes” making it difficult to understand their inner workings and posing interpretability challenges.¹⁰⁸

Deep learning is the source technology on which the current wave of AI is based, including large language models such as ChatGPT.

Black Boxes

Many types of machine learning, including those that utilise deep learning and neural networks are, as Frank Pasquale has labelled them, ‘black boxes.’¹⁰⁹ AI black boxes refer to AI systems with internal workings that are invisible to, or cannot be discerned by, their developer or user. ‘You can feed them input and get output, but you cannot examine the system’s code or the logic that produced the output.’¹¹⁰

Neither the developers nor users understand:

... the basis on which the algorithms function, partly because they themselves do not know or understand how the sub-algorithms function and have simply used algorithms ‘off-the-shelf;’ as such, these would be very difficult to establish compliance.¹¹¹

Black boxes pose significant, perhaps insurmountable, problems for AI regulation. If it is impossible to know how AI produced an outcome, it is equally impossible to know how and why it produced it. Fault-based accountability for foreseeable errors and mistakes under conventional doctrines of the law of negligence becomes unachievable. It is for this reason that regulatory interventions that delineate between *ex ante* and *ex post* transparency break down. The dynamic interplay between data, algorithms and models are of such complexity and occur at such velocity and scale that it is impossible for conventional regulatory ideas, tools and legal assumptions about causation, knowability and transparency to have any real meaning: there simply may not be a discernible record of what has happened when machine learning provides a recommendation, decision or other outcome.

Data

Data is recognised as an indispensable component to power AI systems, serving as the fuel that drives machine learning algorithms and decision-making processes. It is not merely the quantity of data that matters, but also its quality and relevance to the local context. High-quality, contextualised data is essential so that AI systems are tailored to local needs.¹¹²

¹⁰⁸ Ibid.

¹⁰⁹ Frank Pasquale, *The Black Box Society*, Cambridge, 2015.

¹¹⁰ Hannah Yee-Fen Lim (Above, n 104).

¹¹¹ Ibid.

¹¹² Forum of Small States and Digital Forum of Small States, *AI Playbook for Small States*, 22 September 2024, p 21. See <<https://www.imda.gov.sg/-/media/imda/files/news-and-events/media-room/media-releases/2024/09/ai-playbook-for-small-states/imda-ai-playbook-for-small-states.pdf>>.

The machine learning techniques that underpin narrow AI rely on the availability of large data sets. These were largely unavailable to the developers of weak AI as most data during that period was locked within government, corporate or other institutional databases that were not widely shared. The advent of the internet and cloud computing made it possible to share more data but to generate more data derived from activities facilitated by networks of computers, internet-based platforms such as social media and e-commerce as well as internet-enabled devices through the Internet of Things (IoT). The ready availability of large amounts of data is a fundamental resource for machine learning without which it would not be able to function.

Training Data

There are a variety of types of datasets that are used for machine learning.

Training (data) sets, then, are the foundation on which contemporary machine-learning systems are built. They are central to how AI systems recognize and interpret the world. These datasets shape the epistemic boundaries governing how AI systems operate, and thus are an essential part of understanding socially significant questions about AI.¹¹³

Supervised learning uses 'labeled datasets to train algorithms to classify data or predict outcomes accurately.'¹¹⁴ The accuracy and efficacy of machine learning -based supervised learning is highly reliant on the labels used in the training data sets. These are applied by humans and involve attributing meaning to images or other content. In an analysis of the ImageNet database, used in a variety of AI systems, it was found that:

... [t]he classifications of humans within it take a sharp and dark turn. There are categories for Bad Person, Call Girl, Drug Addict, Closet Queen, Convict, Crazy, Failure, Flop, Fucker, Hypocrite, Jezebel, Kleptomaniac, Loser, Melancholic, Nonperson, Pervert, Prima Donna, Schizophrenic, Second-Rater, Spinster, Streetwalker, Stud, Tosser, Unskilled Person, Wanton, Waverer, and Wimp. There are many racist slurs and misogynistic terms.¹¹⁵

The result is that the images analysed by the AI system have meanings assigned to them that are inaccurate and biased, leading AI to produce erroneous outputs.

Unsupervised ML 'uses machine learning algorithms to analyze and cluster unlabelled datasets (subsets called clusters). These algorithms discover hidden patterns or data

¹¹³ Kate Crawford and Trevor Paglen, 'Excavating AI: the Politics of Images in Machine Learning Training Sets,' (2021) 36(1) AI and Society 1105, 1106.

¹¹⁴ Jim Holdsworth and Mark Scapicchio, (Above, n 107)

¹¹⁵ Kate Crawford and Trevor Paglen, (Above n 113) p 1110.

groupings without the need for human intervention.¹¹⁶ Semi-supervised learning is a mix of both supervised and unsupervised ML where AI:

[d]uring training ... uses a smaller labelled dataset to guide classification and feature extraction from a larger, unlabelled data set. Semi-supervised learning can solve the problem of not having enough labelled data for a supervised learning algorithm. It helps too if it's too costly to label enough data.¹¹⁷

Algorithms

Algorithms 'have been around since the beginning of time.'¹¹⁸ An algorithm is:

a specific set of instructions for carrying out a procedure or solving a problem, usually with the requirement that the procedure terminate at some point. Specific algorithms sometimes also go by the name method, procedure, or technique ... The process of applying an algorithm to an input to obtain an output is called a computation.¹¹⁹

Algorithms can 'either be conceived of as mathematical constructs, or as the implementation of such mathematical constructs into a technology, or as a configuration of the technology designed for a particular task.'¹²⁰

As with any set of instructions, some algorithmic processes are more effective than others and some computations are more accurate than others. Their effectiveness and accuracy are largely determined by the data selected for analysis, the assumptions and limitations built-in to the algorithm and the questions posed of the data.

Algorithms can operate autonomously or semi-autonomously. They process the data that they receive in accordance with the rules built-in to them. As such, they involve two related but different forms of automation. The first involves analysing data in ways that would be impossible to do manually. The outcomes then help to automate decision-making – these are algorithms that can augment or replace analysis and decision-making by humans. The second is that they can respond and evolve in response to the way they 'learn' from the data used to train them.

One of the most notorious algorithms is the Black-Scholes algorithm, sometimes blamed for the global financial crisis of 2007. Invented in 1973, it is used for financial trading in derivatives. Essentially, it prices a derivative (in this case an option) in relation to four qualities: time; the price of the asset that secures the option; the risk-free interest rate; and

¹¹⁶ Ibid.

¹¹⁷ Ibid.

¹¹⁸ Jean-Luc Chabert (ed), *A History of Algorithms: from the Pebble to the Microchip*, Berlin, 1999, p 1.

¹¹⁹ Wolfram Mathworld, 'Algorithm', (January 2020). See <<https://mathworld.wolfram.com/Algorithm.htm>>.

¹²⁰ Ugo Pagallo, 'Algo-Rhythms and the Beat of the Legal Drum', (2018) 31 *Philosophy and Technology*, p 507 at p 508.

market volatility. The equation assumes that long run volatility remains the same over the life of the option. However, this final assumption – stable long run volatility – cannot be measured precisely:

Was an equation to blame for the financial crash, then? Yes and no. Black-Scholes may have contributed to the crash, but only because it was abused ... Despite its supposed expertise, the financial sector performs no better than random guesswork ... The system is too complex to be run on error-strewn hunches and gut feelings, *but current mathematical models don't represent reality adequately.*¹²¹

Algorithms can take a number of forms. Central to narrow AI are machine learning, natural language processing (for Large Language Models) and Generative AI algorithms. Again, as with other challenges to the definitional certainty and clarity of AI-related terms that have been noted earlier, these concepts may be used interchangeably or in conjunction with each other.

Large Language Models (LLMs)

Large language models became prominent with the deployment of forms of AI such as ChatGPT. They are designed to process, understand and generate text and other content based on their training data and processed using deep learning techniques.

They can 'infer from content, generate coherent and contextually relevant responses, translate to languages other than English, summarize text, answer questions (general conversation and FAQs) and even assist in creative writing or code generation tasks.'¹²²

One way of considering LLMs in a textual context is that they are trained to predict the next word in a sentence based on the preceding words:

The model does this through attributing a probability score to the recurrence of words that have been tokenized – broken down into smaller sequences of characters. These tokens are then transformed into embeddings, which are numeric representations of this context.¹²³

Essentially LLM-based AI is a complex probability machine that uses training data to predict, on a statistical basis that is not usually apparent or discernible, what comes next. What comes next is widely acknowledged to include potential harm such as bias, hate speech, misinformation and error – phenomena also referred to as 'hallucinations.' Importantly, '[a]

¹²¹ Ian Stewart, 'The Mathematical Equation that Caused the Banks to Crash,' The Guardian, 11 February 2012. See <<https://www.theguardian.com/science/2012/feb/12/black-scholes-equation-credit-crunch>>.

¹²² Cole Stryker, IBM, 'What are large language models (LLMs)?', 10 September 2025. See <<https://www.ibm.com/topics/large-language-models>>.

¹²³ Ibid.

surprising fact about modern large language models is that nobody really knows how they work internally.’¹²⁴

Generative AI

Like LLMs, Generative AI (GAI) relies on deep learning to

generate statistically probable outputs when prompted ... generative models encode a simplified representation of their training data and draw from it to create a new work that’s similar, but not identical, to the original data.¹²⁵

The key difference between the outputs produced by LLMs compared to GAI is that the former is focused on language-based tasks while the latter can create content over wider domains such as images and audio.

Agentic AI

Agentic AI is another form of second wave AI – based around GAI and LLMs – that functions with a high degree of autonomy:

Unlike traditional AI models, which operate within predefined constraints and require human intervention, agentic AI exhibits autonomy, goal-driven behavior and adaptability. The term “agentic” refers to these models’ agency, or, their capacity to act independently and purposefully.¹²⁶

The characteristics of agentic AI include autonomy such that they do not require constant human oversight; proactive such that they ‘can “think” and “do” in a more human-like fashion;’¹²⁷ specialised so as to solve not only single but multiple problems; adaptable to an extent that they ‘can learn from experiences, take in feedback and adjust their behaviour’¹²⁸ and continuously improve; and intuitive – for example users can engage with the agentic AI using voice commands rather than having to use traditional software interfaces. An example of agentic AI is its use in autonomous vehicles where it can continuously process data in real time such as GPS and sensor data to drive a vehicle without a human being involved. Agentic AI ‘can bring serious consequences if (they) go “off the rails.” The usual AI risks apply, but can be magnified in agentic systems.’¹²⁹ For example a financial trading AI agent designed to maximise profit may engage ‘in risky or unethical trading practices, triggering market

¹²⁴ Anthropic, ‘Anthropic’s Interpretability Research’, August 2025. See <<https://transformer-circuits.pub>>.

¹²⁵ Kim Martineau, IBM, ‘What is generative AI?’, 20 April 2023. See <<https://research.ibm.com/blog/what-is-generative-ai>>.

¹²⁶ Cole Stryker, IBM, ‘What is agentic AI?’, 21 October 2024. See <<https://www.ibm.com/think/topics/agentic-ai>>.

¹²⁷ Ibid.

¹²⁸ Ibid.

¹²⁹ Ibid.

instability.¹³⁰ Agentic AI that does not comply with legal and regulatory norms as part of its design and deployment risks prioritising other, potentially unlawful conduct and will pursue those unlawful objectives automatically and at scale.

Narrow AI: some conclusions

Machine learning-based AI essentially ‘applies statistical and probabilistic methods to large data sets to create generalized representations that can be applied to future samples.’¹³¹

Without diminishing its scale, complexity and importance, it is not incorrect to conceptualise narrow AI as a vast probability machine with all of the limitations that statistical and probabilistic methods involve. Second wave AI has many limitations that can give rise to error. There is a distinct lack of recognition of these in AI discourse but they are highly significant to regulatory initiatives.

Disciplines and Skills

A distinguishing feature of AI that causes difficulties for regulatory activity is its complexity. In attempting to emulate human activity it draws on a wide range of disparate skills and disciplines that lawyers and policy makers do not often encounter in combination. These consist of:

- 1) Philosophy – the development of rules for thinking that presuppose that these can be replicated in a way that they can be emulated by a machine;
- 2) Mathematics – in which concepts such as logic, probability and statistics were ‘mathematised’ into a new branch of mathematics: computation;
- 3) Economics – regarding individual decision-making according to preferences was first investigated together with concepts of utility, decision theory, game theory and satisficing (making decisions that are “good enough”) were first developed;
- 4) Neuroscience – understanding how the brain operates and processes information;
- 5) Psychology – concerning the study of how humans think and act, including sub-disciplines such as cognitive psychology;
- 6) Computer Engineering – developments in computer engineering that have seen the utility and speed of digital processing are fundamental to AI;

¹³⁰ Ibid.

¹³¹ Defence Advanced Research Projects Agency, (above n 98)

- 7) Control theory and cybernetics – under which an object or artifact, such as a robot, can operate under its own control; and
- 8) Linguistics – an essential element in “translating” text, including its subject matter and context, into a form usable by computing technology. Including natural language processing.¹³²

For any given instance of AI, any number of these disciplines are relevant to the way in which it produces outcomes. Where those outcomes cause harm or involve risks, the underlying cause might be obscured beneath layers of assumptions and analysis produced by their application (or misapplication). This complexity produces significant difficulties for regulators or for litigants who wish to seek the assistance of the legal process – for example, the law of negligence – to address and compensate for loss and damage attributable to AI.

Definitions of AI

How does one define AI? And how does one clearly distinguish between AI and non-AI machine-based systems? To date, the answer is as elusive as the question is simple. Many of the world’s smartest people have tried to come up with a solid definition that everyone agrees upon but never seem to find consensus. There is no clear red line but a continuum of features characterising what we think of as artificial intelligence and where the “magic” happens.¹³³

As shown by this chapter, defining AI is complex: ‘There is no straightforward, consensus definition of artificial intelligence.’¹³⁴ Despite these challenges a definition is essential in a regulatory context as ‘[a]ny AI regulatory regime must define what exactly it is that the regime regulates; in other words, it must define artificial intelligence.’¹³⁵

The problem that lies at the heart of defining AI is in the conceptual difficulty of defining intelligence.

Definitions of intelligence ... vary widely and focus on myriad interconnected human characteristics that are themselves difficult to define, including consciousness, self-awareness, language use, the ability to learn, the ability to abstract, the ability to adapt, and the ability to reason.¹³⁶

Noting that many of these characteristics are themselves subject to debate, it is unsurprising that definitional clarity for AI is elusive.

¹³² Stuart Russell and Peter Norvig, (above n 85) p 23.

¹³³ Marko Grobelnik, Karine Perset, Stuart Russell, OECD, ‘What is AI? Can you make a clear distinction between AI and non-AI systems?’, 6 March 2024. See <<https://oecd.ai/en/wonk/definition>>.

¹³⁴ Ryan Calo, *Artificial Intelligence Policy: A Primer and Roadmap*, (2017) 51 U.C.D.L. Rev. 399, 404.

¹³⁵ Matthew Scherer, (above n 66) p 359.

¹³⁶ *Ibid*, p 360.

Although many international organisations, governments and researchers have proposed definitions of AI, the OECD's work in this space is highly influential. Its approach has been to work towards defining AI as an 'AI system.' It first published a definition in 2019:

An AI system is a machine-based system that can, for a given set of human-defined objectives, make predictions, recommendations, or decisions influencing real or virtual environments. AI systems are designed to operate with varying levels of autonomy.¹³⁷

This definition was revised and significantly changed in 2024. The current¹³⁸ definition is:

An AI system is a machine-based system that, for explicit or implicit objectives, infers, from the input it receives, how to generate outputs such as predictions, content, recommendations, or decisions that can influence physical or virtual environments. Different AI systems vary in their levels of autonomy and adaptiveness after deployment.¹³⁹

The differences between the 2019 definition and the 2024 definition are noteworthy. The following shows the 2019 definition as updated by the 2024 text:

An AI system is a machine-based system that , for explicit or implicit objectives, infers, from the input it receives, how to generate outputs such as, predictions, content, recommendations, or decisions that can influence physical or virtual environments. Different AI systems vary in their levels of autonomy and adaptiveness after deployment.

The Australian government adopted the OECD's 2024 definition of AI system for the purposes of its policy work in its regulatory proposals paper for AI, *Safe and Responsible AI in Australia: Proposals Paper for Introducing Mandatory Guardrails for AI in high-risk Settings*¹⁴⁰ (2024 Proposals Paper).

When comparing the OECD's definitions, the key point to note is that the 2024 definition is more restrictive than the 2019 definition. The addition of 'infers, from the input it receives' describes second wave AI, not first wave AI. First wave AI does not infer anything from the input it receives: its intelligence is already built in. The effect of the updated 2024 definition is to excise a significant category of AI from the scope of regulation. Although the development and use of weak AI and its use for some decades has clearly not attracted the attention of lawyers, regulators and policy makers to the same extent as second wave AI,

¹³⁷ OECD, *Scoping the OECD AI Principles: Deliberations of the Expert Group on Artificial Intelligence at the OECD*, OECD Digital Economy Papers, No 291, November 2019, p 7. See <<https://www.oecd-ilibrary.org/docserver/d62f618a-en.pdf?expires=1731028740&id=id&accname=guest&checksum=52860BEC7CB7EC0935B2B4E719E38DB1>>.

¹³⁸ As at 30 June 2025.

¹³⁹ OECD, *Explanatory Memorandum on the Updated OECD Definition of an AI System*, OECD Artificial Intelligence Papers, No 8, March 2024, p 4. See <<https://www.oecd-ilibrary.org/docserver/623da898-en.pdf?expires=1731026306&id=id&accname=guest&checksum=6BF135B77E16058D47EDB67C73BB9CA0>>.

¹⁴⁰ Australian Government, Department of Industry, Science and Resources, *Safe and Responsible AI in Australia: Proposals Paper for Introducing Mandatory Guardrails for AI in high-risk Settings*, September 2024, p 8. See <https://storage.googleapis.com/converlens-au-industry/industry/p/prj2f6f02ebfe6a8190c7bdc/page/proposals_paper_for_introducing_mandatory_guardrails_for_ai_in_high_risk_settings.pdf>.

Australia's Robodebt scheme clearly demonstrates the widespread damage that simple first wave AI can cause and the practical difficulties that those affected experienced in seeking redress.

Other approaches

While most of the advanced world economies have settled using definitions of AI that are either the same as or consistent with the OECD's 2024 definition to anchor their regulatory approaches, alternatives have been developed which have immediate regulatory impact in real-world settings and are worth consideration.

One of the most significant of these in Australia is the Supreme Court of Victoria's *Guidelines for Litigants: Responsible Use of Artificial Intelligence in Litigation* (VSC AI Guidelines).¹⁴¹

These 'have been developed to assist those conducting litigation in the Supreme Court of Victoria'¹⁴² and define AI as '[a] term describing a range of technologies and techniques used to computationally generate outputs that typically require human intelligence to produce.'¹⁴³

This definition underpins a range of principles for the use of AI by litigants and guidance about how to apply the principles. It is capable of covering a wide spectrum of AI, including both weak and narrow AI. It is also significantly more straightforward than either of the OECD's 2019 or 2024 definitions. This thesis adopts the Victorian Supreme Court definition.

Closing observations

AI is not 'intelligent.' First wave AI simply embodies existing human knowledge and uses it to solve domain-limited problems. Its advantage is that it does so faster and more accurately than humans. Second wave AI is also not intelligent. It is a vast, complex, statistical and probability machine that uses data and algorithms to make decisions, analyse patterns and make informed decisions. Both *appear* to be 'intelligent' because their functionality emulates what humans can do. The Victorian Supreme Court definition captures both forms of AI.

The effectiveness and accuracy of any form of AI is determined by how closely it emulates human intelligence or agency. If training data used in second wave AI is inaccurate or

¹⁴¹ Supreme Court of Victoria, *Guidelines for Litigants: Responsible Use of Artificial Intelligence in Litigation*, May 2024. See <<https://www.supremecourt.vic.gov.au/sites/default/files/2024-05/AI%20Guidelines%20SCV.pdf>>.

¹⁴² *Ibid*, p 1

¹⁴³ *ibid*, p 5

insufficient its ability to do so becomes compromised and it produces errors. If the algorithms used do not model reality correctly errors also occur.

Compounding these issues, second wave AI is a black box type of intelligence:

Understanding the nature of this intelligence is a profound scientific challenge, which has the potential to reshape our conception of what it means to *think*. The stakes of this scientific endeavor are high; as AI models exert increasing influence on how we live and work, we must understand them well enough to ensure their impact is positive.¹⁴⁴

Despite ongoing research, our deficits in understanding second-wave types of AI, and our lack of reliable pathways to explain *how* it produces outcomes remain an intractable regulatory problem. This thesis argues that one of the goals of any AI regulatory system should be to create the incentives to address this issue. In the meantime, these deficits should not be permitted to enable AI from being held to account.

¹⁴⁴ See Jack Lindsay and Ors, 'On the Biology of a Large Language Model,' Transformer Circuits Thread, 27 March 2025. See <<https://transformer-circuits.pub/2025/attribution-graphs/biology.html#discussion>>.

Chapter Three – The foundational risks posed by AI

Background

Existing laws are not well adapted to regulating second wave AI because it has unique characteristics that mean it is not easily amenable to the application of those existing laws.

Second wave AI relies on data inputs that involve the use of large quantities of personal data and other material that is protected by copyright. These *input* issues raise difficulties in applying data protection and copyright laws. Processing those inputs using algorithms and other complex computing techniques produces *outputs* in a manner that is not readily understandable or explainable. This presents significant problems in applying traditional fault-based accountability mechanisms such as the law of negligence.

Approach and scope

A multitude of AI-related risks and harms have been identified. There is a significant body of literature that describes and analyses those risks and harms – both actual and potential. As discussed in Chapter 1, more than seventy have been identified by Arie Freiberg alone.¹⁴⁵ As with much of the discourse surrounding AI risks, the extent to which these risks and harms might be offset by other positive impacts is debated.

AI threatens to displace workers who undertake routine tasks through increased mechanisation as well as those employed in non-routine cognitive tasks such as the creative occupations,¹⁴⁶ in jobs where research and the ability to assemble facts and arguments is a key part of their role, or where tasks involve some form of decision-making. These impacts may not be uniform. They may have a larger impact on low-skilled workers or in replacing entry level positions, for example in the professions, administration and computer software coding.¹⁴⁷ Against these risks there are also claims that AI will boost productivity and create new employment opportunities.¹⁴⁸

¹⁴⁵ Arie Freiberg, (above n 9) 2025, p 460 – 462.

¹⁴⁶ For example, through machine-generated images and sound recordings. See World Economic Forum, 'This is how AI is impacting – and shaping – the creative industries, according to Experts at Davos', 28 February 2024. See <<https://www.weforum.org/stories/2024/02/ai-creative-industries-davos/>>.

¹⁴⁷ Steve Lohr, 'A.I. Is Prompting an Evolution, Not Extinction, for Coders,' *The New York Times*, 20 February 2025. See <<https://www.nytimes.com/2025/02/20/business/ai-coding-software-engineers.html>>.

¹⁴⁸ These debates are discussed in detail in Australian Government, Senate Select Committee on Adopting Artificial Intelligence, *Final Report*, November 2024, chapter 4. See <[https://parlinfo.aph.gov.au/parlInfo/download/committees/reportsen/RB000470/toc_pdf/SelectCommitteeonAdoptingArtificialIntelligence\(AI\).pdf](https://parlinfo.aph.gov.au/parlInfo/download/committees/reportsen/RB000470/toc_pdf/SelectCommitteeonAdoptingArtificialIntelligence(AI).pdf)>

At the same time, various concerns have been expressed about how the use of AI in the workplace can diminish workers' skills and abilities. Although workers may be able to 'perform knowledge-intensive tasks significantly more efficiently when given access to generative AI',¹⁴⁹ their ability 'to know beforehand whether generative AI can solve a task correctly, [means that] they must vigilantly check its outputs and fix any issues present. However, if they do not learn the underlying skills, they may lack the expertise required to do so.'¹⁵⁰ Furthermore, while generative AI can make some tasks easier, that advantage is accompanied by 'the risk of deteriorating our ability to effectively learn some of the skills required to solve these tasks.'¹⁵¹ One consequence is that predicted productivity gains may not materialise. Another is that human oversight of AI as a risk mitigation measure, having a 'human in the loop'¹⁵² who can review AI outputs and check them for errors, may well be ineffective. Emerging, related risks include reports that the use of AI by students is leading to a decline in cognitive development¹⁵³ and threatens critical thinking skills.¹⁵⁴

Environmental risks are also frequently raised as AI-related harms.¹⁵⁵ The data centres that power AI require significant amounts of electricity that is often produced from fossil fuels and large amounts of water for cooling in a context where water is in increasingly short supply. Complex AI componentry requires the supply of minerals and rare earths that may be mined unsustainably and, at the end of their lifespan, leave significant e-waste that is difficult to recycle economically.¹⁵⁶

The 2024 Proposals Paper categorises AI-based risks. One of these risk is bias, either as a result of embedding existing human biases or amplifying them in new ways attributable to 'inaccurate, insufficient, unrepresentative or outdated data'¹⁵⁷ or from the problematic 'design, configuration of or deployment of the AI system itself.'¹⁵⁸

¹⁴⁹ Hamsa Bastani, Osbert Bastani, Alp Sungu, Haosen Ge, Özge Kabakçı and Rei Mariman, *Generative AI Can Harm Learning* (July 15, 2024). The Wharton School Research Paper, 15 July 2024, p 2. See SSRN: <<https://ssrn.com/abstract=4895486>> or <<http://dx.doi.org/10.2139/ssrn.4895486>>.

¹⁵⁰ Ibid, pp 2-3.

¹⁵¹ Ibid, p 16.

¹⁵² Xiao-Le Meng, 'Data Science and Engineering with Human in the Loop, Behind the Loop and Above the Loop,' (2023) 5(2) *Harvard Data Science Review*, 28 April 2023, p 1.

¹⁵³ Chris Westfall, 'The Dark Side of AI: Tracking the Decline of Human Cognitive Skills,' *Forbes Magazine*, 18 December 2024. See <<https://www.forbes.com/sites/chriswestfall/2024/12/18/the-dark-side-of-ai-tracking-the-decline-of-human-cognitive-skills/>>.

¹⁵⁴ Jill Barshay, *Kids who use ChatGPT as a Study Assistant do Worse on Tests*, The Hechinger Report, 2 September 2024. See <<https://hechingerreport.org/kids-chatgpt-worse-on-tests/>>.

¹⁵⁵ Adam Zwe, 'Explained: Generative AI's Environmental Impact,' MIT News, 17 January 2025. See <<https://news.mit.edu/2025/explained-generative-ai-environmental-impact-0117>>.

¹⁵⁶ Josey Walden, 'The New Cold War: Rare Earths, AI, and Strategic Competition with China,' American Security Project, 2 July 2025. See <<https://www.americansecurityproject.org/the-new-cold-war-rare-earths-ai-and-strategic-competition-with-china/>>.

¹⁵⁷ 2024 Proposals Paper p 12

¹⁵⁸ Ibid.

Another category involves harms:

- from AI misuse and failure that result in individual harm (either physical or psychological) from privacy breaches or exclusion from access to opportunities or services;
- collective harms such as discrimination based on protected attributes under relevant discrimination laws;
- organisational harms such as through the introduction or identification of security vulnerabilities into enterprise software or cyber-attacks exposing training data; and
- societal harms such as growing inequality, misinformation, extreme content and the erosion of social cohesion.¹⁵⁹

National security harms are also pinpointed, including the creation and dissemination of foreign influence manipulation content;¹⁶⁰ disinformation that erodes public trust in democratic institutions and damages social cohesion;¹⁶¹ and lowering entry barriers for actors to engage in malicious cyber activity through enhanced levels of cybercrime, scams and threats to critical infrastructure.¹⁶²

As noted in Chapter One, a comprehensive analysis of AI-associated legal risks and their potential mitigation would require a separate thesis for each. This chapter focuses on the problems of applying existing laws to key input and output risks considered to be foundational, namely legal risks that are fundamental to the production of downstream harms.

Foundational risks

Many AI risks do not present within neat compartments or categories of traditional legal and regulatory governance. In many cases they extend across one or more of them, thus complicating their analysis and remediation.

The key foundational risks addressed in this chapter are data scraping, breaches of data protection law, copyright infringement and liability for AI-produced harms. The chapter also contains two case studies. The first, focusing on AI in the public sector, illustrates the

¹⁵⁹ Ibid.

¹⁶⁰ Ibid.

¹⁶¹ Ibid.

¹⁶² Ibid.

difficulties in using administrative law to regulate AI by reference to the Robodebt programme and the Commonwealth Ombudsman's Report into Automation in the Targeted Compliance Framework.¹⁶³ The second is an example of discrimination involving the deployment of AI in Australian employment screening settings.

The risks posed by AI challenge the application of existing laws that predate contemporary AI. Machine learning-based AI involves the acquisition of large volumes of data for training. This raises a threshold question – where does that data come from?

Frequently it is obtained by scraping personal data and copyright content from the internet. Data protection law, with its focus on providing individuals with rights that enable them to exercise control over their personal data, is not well adapted to this environment. Likewise, the use of copyright works and other subject matter for AI training involves the reproduction of that content without the copyright owner's permission. No Australian court has yet produced a binding decision as to whether this constitutes copyright infringement, but it is argued that it is highly likely that it is.

Another source is data collected on the basis of frequently impenetrable organisational privacy policies and adhesion terms and conditions of service that are presented on a 'take it or leave it' basis, where the legal authority for, and individual consent to, personal data being shared and used so broadly is, legally questionable. The use of these privacy policies and terms and conditions is now a 'business as usual practice.' Although these documents might technically comply with data protection laws, they fail to provide individuals with any real control over their personal data. Even if individuals have the capacity to understand and analyse them, they can have little understanding of how their personal data is propagated across various organisations and jurisdictional boundaries by an ever-multiplying ensemble of partners, affiliates and service providers'.

The unique characteristics of AI also mean that it is not easily amenable to traditional fault-based liability doctrines such as the law of negligence. This stems primarily from the black box nature of many forms of AI, which means that it is not readily understandable or contestable: if it is impossible to understand how AI reaches an output, fault-based liability is a practical impossibility. In addition, there are other impediments, including the difficulty of

¹⁶³ Commonwealth Ombudsman, 'Automation in the Targeted Compliance Framework: when the Law is Changed but the System Isn't,' August 2025. See <https://www.ombudsman.gov.au/_data/assets/pdf_file/0017/320750/Automation-in-the-Targeted-Compliance-Framework.pdf>.

applying doctrines of foreseeability and causation to AI across what is known as the ‘AI value chain.’ GPAI and Agentic AI, with their characteristic high levels of autonomy and adaptability tend to amplify these risks.

Even comparatively simple first wave AI that theoretically should be amenable to existing regulation has proven to be difficult to oversee in practice, as the Robodebt programme has demonstrated. Administrative law was ultimately used to unravel its illegality and unfairness, but it applies only to the use of AI by government, not the private sector.¹⁶⁴ Even then it took a Royal Commission with its broad investigatory powers to uncover the root cause of the error and to highlight repeated failures of regulators to properly undertake their oversight of it. Despite this process, it appears that the errors made by government in respect of the Robodebt programme have been replicated.¹⁶⁵

Data scraping

Machine learning-based AI would not function without data. ‘Data is a key component for all AI systems – to date, the most significant improvements in AI systems have been tied to access to very large amounts of training data.’¹⁶⁶

One of the main methods used to acquire training data is through data scraping. Data scraping is a widespread practice that involves the automated extraction of publicly available information from online sources using a variety of techniques such as web scraping,¹⁶⁷ web crawling¹⁶⁸ and screen scraping,¹⁶⁹ each of which has different technical parameters and applications. Online data sources used for scraping can include social media platforms, web

¹⁶⁴ Except to the extent that outsourcing is involved.

¹⁶⁵ Commonwealth Ombudsman, (above n 163).

¹⁶⁶ Jennifer King and Caroline Meinhardt, (above, n 64) p 5

¹⁶⁷ Web scraping involves extracting data from third-party websites or social media platforms, using specialised software like crawlers, bots and scrapers. The process typically entails sending a request to a website, parsing the website’s HTML content, and extracting the desired data from such a website. See Organisation for Economic Cooperation and Development, *Intellectual Property Issues in Artificial Intelligence Trained on Scraped Data*, February 2025, p17. See <https://www.oecd.org/content/dam/oecd/en/publications/reports/2025/02/intellectual-property-issues-in-artificial-intelligence-trained-on-scraped-data_a07f010b/d5241a23-en.pdf>.

¹⁶⁸ Web crawling ‘refers to the automated navigation and indexing of web pages via hyperlinks. This technique uses web crawlers, also known as spiders or bots, to systematically collect data from web pages, databases or other digital sources, often by crawling through the content and retrieving specific information. See Organisation for Economic Cooperation and Development, *Intellectual Property Issues in Artificial Intelligence Trained on Scraped Data*, February 2025, p17. See <https://www.oecd.org/content/dam/oecd/en/publications/reports/2025/02/intellectual-property-issues-in-artificial-intelligence-trained-on-scraped-data_a07f010b/d5241a23-en.pdf>.

¹⁶⁹ Screen scraping involves the automatic extraction of data that is visually displayed on a screen, such as text, images and videos, PDFs and other documents (sic) formats. Screen scraping captures and processes visual output that is rendered as it would be on a user’s device. See Organisation for Economic Cooperation and Development, *Intellectual Property Issues in Artificial Intelligence Trained on Scraped Data*, February 2025, p 17. See <https://www.oecd.org/content/dam/oecd/en/publications/reports/2025/02/intellectual-property-issues-in-artificial-intelligence-trained-on-scraped-data_a07f010b/d5241a23-en.pdf>.

sites, databases, online forums and e-commerce sites. ‘Scraping is a ubiquitous practice, and it is increasing.’¹⁷⁰ An estimated quarter of all web traffic is attributable to it.¹⁷¹

There is no settled, consensus definition of data scraping. The OECD suggests that an appropriate working definition is ‘the automated extraction of AI training data from the web, online databases and from other sources using automated software tools or scripts.’¹⁷² It states that data scraping is usually characterised by the following three features.

1. **Automation:** the use of software tools or scripts designed to quickly and efficiently harvest or otherwise aggregate data with minimal human intervention.’
2. **Scalability:** ‘often used to collect or make accessible large amounts of data that would be impractical to aggregate manually’ and ‘the tools and techniques employed can be scaled up to extract data from numerous sources simultaneously;’ and
3. **Lack of coordination:** ‘often done without coordination between the data scraper and the entity hosting the data.’¹⁷³

Data scraping is different to data mining, which ‘refers to computational processes for identifying patterns, trends, and correlations.’¹⁷⁴ The distinction is important as some legal systems, but not Australia’s, permit data mining as an exception to copyright.¹⁷⁵

The legality or otherwise of data scraping is a cross-cutting issue. Although data scrapers often contend that ‘scraping is legal if you scrape data that is publicly available on the internet,’¹⁷⁶ this position is more consistent with US than Australian, law. Two of the key legal issues concern the relationship between data scraping, data protection and copyright law. Although both are discussed in detail later in this chapter, a brief account of the issues is necessary at this point to make sense of data scraping issues.

¹⁷⁰ Woodrow Hartzog and Daniel Solove, *The Great Scrape: The Clash Between Scraping and Privacy*, [2024] Boston University School of Law, Scholarly Commons at Boston University Law School 1 at p 7.

¹⁷¹ Igal Zeifman, *Bot Traffic Report 2016*, 24 January 2017. See <<https://www.imperva.com/blog/archive/bot-traffic-report-2016/?redirect=Incapsula>>. Statistics on web scraping are difficult to obtain. This report, which was prepared in 2016 in a pre-generative AI era could constitute an underestimate of the current extent of web scraping.

¹⁷² Organisation for Economic Cooperation and Development, *Intellectual Property Issues in Artificial Intelligence Trained on Scraped Data*, February 2025, p 16. See <https://www.oecd.org/content/dam/oecd/en/publications/reports/2025/02/intellectual-property-issues-in-artificial-intelligence-trained-on-scraped-data_a07f010b/d5241a23-en.pdf>.

¹⁷³ Ibid.

¹⁷⁴ Ibid p 9.

¹⁷⁵ These are often referred to as ‘Text and Data Mining’ (TDM) exemptions. For example, EU Directive 2019/790 on copyright and related rights in the Digital Single Market, see <<https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32019L0790>> and *Copyright, Designs and Patents Act 1988* (UK), s 29A.

¹⁷⁶ Ondra Urban, ‘Is Web Scraping Legal?’, Apify, 26 May 2025. See <<https://blog.apify.com/is-web-scraping-legal/>>.

Data scraping and Data Protection

Data scraping may involve personal information. In Australia, the collection and other handling of personal information engages regulatory obligations under the *Privacy Act 1988* (Cth) (Privacy Act). These include transparency, collection limitation, purpose limitation, notification, use and disclosure limitation principles and obligations to ensure personal data is accurate and up to date. Data scraping does not generally comply with these requirements.

The Clearview case

This issue arose in the course of the joint UK Information Commissioner's Office (ICO) and Office of the Australian Information Commissioner's (OAIC's) investigation into the personal information handling practices of Clearview AI, Inc (Clearview).¹⁷⁷ Clearview provided a facial recognition tool that enabled users to upload a digital image of a face and undertake a search for it against a database of more than three billion images in order to display likely matches and thus identify images uploaded by the user. It functioned by collecting images of individuals' faces from publicly available internet sources and comparing them against the image uploaded by the user. The program determined a match, and other 'close' matches, using a variety of technical tools. Clearview's service was provided to public sector organisations for law enforcement and national security purposes. Additional functionalities included enabling agencies to learn more about the individuals matched through linking to additional information about them gleaned from business, relationship and dating information sources. This was designed to allow them, for example, to 'accurately dispense social benefits and reduce fraud'.¹⁷⁸

In Australia, the investigation determined that multiple breaches of the Australian Privacy Principles (APPs) had occurred. The interferences with individuals' privacy included collecting their sensitive information without consent; failing to collect personal information by lawful and fair means; failing to notify individuals that their personal information had been collected; and not ensuring that the personal information Clearview used or disclosed was accurate, up-to-date, complete and relevant.¹⁷⁹ The OAIC made declarations under s

¹⁷⁷ Commissioner Initiated Investigation into Clearview AI, Inc, [2021] AICmr 54, 14 October 2021. See <https://www.oaic.gov.au/__data/assets/pdf_file/0016/11284/Commissioner-initiated-investigation-into-Clearview-AI,-Inc.-Privacy-2021-AICmr-54-14-October-2021.pdf>.

¹⁷⁸ Ibid, p 3.

¹⁷⁹ Ibid, pp 1-2.

52(1A) of the Privacy Act in the nature of orders, including orders that Clearview ‘not repeat or continue the acts or practices ... found [to have been] an interference with the privacy of one or more individuals’¹⁸⁰ and requiring Clearview to confirm within ninety days that it had ceased the acts and practices that were the subject of the declarations.¹⁸¹

Clearview subsequently appealed against the OAIC decision to the Administrative Appeals Tribunal,¹⁸² arguing that it was not bound by the Privacy Act. Its appeal was based on two grounds: that as a Delaware corporation with no staff located in Australia, Clearview was not bound by the Privacy Act; and that, even if the Privacy Act applied, Clearview was exempt as a small business operator.¹⁸³

The Privacy Act has extra-territorial application where an act or practice undertaken outside Australia has an ‘Australian link.’¹⁸⁴ At the time, the relevant consideration was whether Clearview, by continuing to collect images of Australians using web crawler data scraping techniques amounted to carrying on business in Australia.¹⁸⁵ The Tribunal found that it did, relying on the Full Court of the Federal Court’s decision in *Facebook Inc v Australian Information Commissioner*,¹⁸⁶ in particular Allsop CJ’s formulation that:

acts or activity in Australia need not be intrinsically commercial in themselves if they involve acts within the territory that amount to or are ancillary to, transactions that make up and support the business.¹⁸⁷

Clearview’s small business operator argument was dismissed for the reasons that Clearview’s evidence of its turnover was unsatisfactory,¹⁸⁸ and Clearview disclosed personal information for a benefit and thus was not covered by the exemption.¹⁸⁹

Despite these regulatory interventions, it appears that Clearview continued to collect and process Australians’ images well after it was ordered to stop, raising questions about the enforceability of Australian data protection legal decisions against overseas companies. Clearview has continued to assert that it collects Australians’ images from servers not

¹⁸⁰ Ibid, p 2

¹⁸¹ Ibid.

¹⁸² Clearview AI Inc and Australian Information Commissioner [2023] AATA 1069.

¹⁸³ The small business operator exemption in the Privacy Act 1988 (Cth) (*‘Privacy Act’*) is discussed in greater detail below.

¹⁸⁴ *Privacy Act*, s 5B.

¹⁸⁵ *Privacy Act*, s 5B(3)(b).

¹⁸⁶ [2022] FCAFC 9.

¹⁸⁷ Ibid, [10].

¹⁸⁸ See n 38, [116].

¹⁸⁹ See n 38, [107].

located in Australia and that it is unable to determine whether an image it collects is an image of an Australian.¹⁹⁰

Data Scraping and Copyright

In the case of copyright, data scraping involves the reproduction of works or other subject matter without the authority of the copyright owner through the process of acquiring text, images, sound recordings and video. It thus gives rise to copyright infringement unless the scraped content is not protected by copyright law, the copying is not substantial, or some other exemption applies. The specific outcome is jurisdiction dependent.

There have been some examples of large organisations that own and produce a significant volume of copyright material entering into copyright licensing agreements with AI developers. The assumption that underpins these arrangements is that the use of their copyright material for AI training would constitute infringement in the absence of the licensing arrangement. An example is the global partnership between News Corporation and OpenAI¹⁹¹ under which 'OpenAI has permission to display content from News Corp mastheads in response to user questions and to enhance its products.'¹⁹² In Australia, the publisher Black Inc sought permission from its published authors to use their works to train AI, requesting that they grant to it the 'right to reproduce or use, adapt, and exploit the work in connection with the development of any software program, including, without limitation, training, testing, validation and deployment of machine learning or generative artificial intelligence system.'¹⁹³

Some have argued that the use of copyright material for AI training falls within the exemption for 'fair dealing.'¹⁹⁴ In Australia, there are fair dealings exemptions in respect of works for research and study;¹⁹⁵ criticism or review;¹⁹⁶ parody or satire;¹⁹⁷ and news

¹⁹⁰ Cam Wilson, 'Clearview AI is Still Collecting Photos of Australians for its Facial Recognition Database,' Crikey, 8 February 2024. See <<https://www.crikey.com.au/2024/02/08/clearview-ai-australia-facial-recognition-data/>>.

¹⁹¹ News Corp and OpenAI Sign Landmark Multi-Year Global Partnership, 22 May 2024. See <<https://investors.newscorp.com/node/15946/pdf>>.

¹⁹² Ibid.

¹⁹³ Hannah Amoud, 'Death Warrant': Australian Publisher Slammed over Author AI Request,' 5 March 2025, *The Sydney Morning Herald*. See <<https://www.smh.com.au/culture/books/death-warrant-australian-publisher-slammed-over-author-ai-request-20250305-p5lh0u.html>>.

¹⁹⁴ *Copyright Act 1968* (Cth) Part IV Division 3 (works) and Part IVA Division 6 (other subject matter). (*Copyright Act*)

¹⁹⁵ *Copyright Act*, s 40.

¹⁹⁶ *Copyright Act*, s 41.

¹⁹⁷ *Copyright Act*, s 42.

reporting¹⁹⁸. Fair dealing provisions also cover similar categories in respect of other subject matter.¹⁹⁹

For both works and other subject matter, the most relevant of the fair dealing limitations on copyright is research or study. The matters to be taken into account in determining whether a fair dealing has taken place include (but are not limited to) the purpose and character of the dealing; the nature of the copyright work; the possibility of obtaining the work within a reasonable time at an ordinary commercial price;²⁰⁰ the effect of the dealing on the potential market for or value of the work; and, if only a part of the work is reproduced, the amount and substantiality of the copy. Similar considerations apply to the copyright in other subject matter.²⁰¹

The law regarding fair dealing is imprecise and ‘amorphous.’²⁰² In considering fair dealing in *Hubbard v Vosper*, Lord Denning MR stated that:

It is impossible to define what is “fair dealing”. It must be a question of degree. You must consider first the number and extent of the quotations and extracts. Are they altogether too many and too long to be fair? Then you must consider the use made of them. If they are used as a basis for comment, criticism or review, that may be fair dealing. If they are used to convey the same information as the author, for a rival purpose, that may be unfair. Next, you must consider the proportions. To take long extracts and attach short comments may be unfair. But, short extracts and long comments may be fair. Other considerations may come to mind also. But, after all is said and done, it must be a matter of impression.²⁰³

It is difficult to see how the fair dealing test so interpreted could be satisfied by data scraping given the widespread, indiscriminate data scraping of the entirety of copyright works and other subject matter that data scraping entails and the difficulty in characterising it as ‘research’ or ‘study’ within the normal meaning of those words. However, whether data scraping constitutes fair use for research or study in Australia has yet to be determined by the courts.

Data scraping may be permissible in other jurisdictions; for example, under the ‘US fair use doctrine, the ‘text and data mining’ exception in the EU, the Copyright Act in Japan, the fair use provision in the Israeli copyright law and both the fair use and text and data mining

¹⁹⁸ Copyright Act, s 43.

¹⁹⁹ The relevant provisions are s 103A (criticism or review); s 103AA (parody or satire); s 103B (news reporting); and s 103C (research or study).

²⁰⁰ Copyright Act, ss 40(2)(c).

²⁰¹ Copyright Act, ss 103C(2)(a)-(e).

²⁰² AGL Energy Ltd v Greenpeace Australia Pacific Ltd [2021] FCA 625 [45].

²⁰³ [1972] 2 QB 84 at p 94.

provisions in Singapore's Copyright Act.²⁰⁴ It is generally considered that the US fair use doctrine is significantly broader than Australia's. The variety of these exemptions adds an additional layer of jurisdictional complexity to copyright and data scraping issues, including the fact that data scraped may be legally used in one jurisdiction but may constitute a copyright infringement where used in another. As the OECD states, '[t]his raises complex questions of private international law, particularly in determining where the infringement, if any, occurred and what laws should be applied.'²⁰⁵

Data Scraping: A wicked problem

Wicked problems are discussed in Chapters One and Four. The legal issues relating to data scraping meet this description. There is an obvious need for large amounts of data and content to ensure that training data is comprehensive and up-to-date, and thus that the AI on which it is based performs optimally. But this need creates a significant tension with the rights of copyright owners whose content has been scraped. As noted by Hartzog and Solove, data scraping prohibitions:

can lead to many socially detrimental consequences, including depriving ... researchers of important tools to keep industry and government accountable. Market forces might compel some companies to restrict third party scraping in an effort to protect what they view as their proprietary data, but this, too would be highly undesirable, leading to an internet more akin to a series of walled gardens. A regulatory intervention must be made, but both encouraging and discouraging scraping comes with huge costs, resulting in a choice between Scylla and Charybdis.²⁰⁶

The arguments for and against data scraping and the application of fair use limitations are often cast in either/or propositions that involve false binaries. For example:

[t]he stakes are high and the consequences are often described in existential terms. Some warn that requiring AI companies to license copyrighted works would throttle a transformative technology, because it is not practically possible to obtain licences for the volume and diversity of content necessary to power cutting-edge systems. Others fear that unlicensed training will corrode the creative ecosystem, with artists' entire bodies of works used against their will to produce content that competes with them in the marketplace.²⁰⁷

²⁰⁴ Organisation for Economic Cooperation and Development, *Intellectual Property Issues in Artificial Intelligence Trained on Scraped Data*, February 2025, pp 23-24. See <https://www.oecd.org/content/dam/oecd/en/publications/reports/2025/02/intellectual-property-issues-in-artificial-intelligence-trained-on-scraped-data_a07f010b/d5241a23-en.pdf>.

²⁰⁵ *Ibid* p 26

²⁰⁶ Hartzog and Solove (above, n 170) p 6.

²⁰⁷ Gilbert and Tobin, 'Is AI Training Fair Use of Copyright Works?', 26 May 2025. See <<https://www.gtlaw.com.au/insights/is-ai-training-fair-use-of-copyright-works>>.

In Australia, this argument overlooks the existence of well-established potential solutions that involve the use of existing statutory compulsory licensing and voluntary licensing concepts in the Copyright Act.

Although data scraping, data protection and copyright are ‘in desperate need of reconciliation, ... [as scraping] is in conflict with nearly all core privacy principles’²⁰⁸ and copyright protections ... a ban on scraping is ‘untenable’,²⁰⁹ the problems raised are not, at least for copyright, unprecedented or novel. However, the data protection issues are more conceptually complex and intractable. These challenges are discussed in more detail in Chapter Six.

Data Protection

Data protection, also commonly known as information privacy, bestows rights designed to provide individuals with a measure of control over their personal data. They accordingly constitute a subset of Daniel Solove’s taxonomy of privacy rights: privacy as control over information about oneself.²¹⁰

Part of the training data used by machine learning-based AI is personal data. That data is obtained through either data scraping or through privacy policies or terms and conditions of service. In many cases individuals do not know that their personal data has been collected, used or disclosed for these purposes and their consent to collection (in the case of sensitive information) has not been given freely, fairly or on an informed basis.

In addition, individuals are generally unaware that their personal data has been scraped, by whom it has been scraped and the identity of the potential multitude of actors spread across jurisdictions who may be involved in the sharing of their personal data. This lack of knowledge of potential wrong-doing means, in practice, that they are not in a position to exercise their data protection rights. In any event, in Australia and a variety of other jurisdictions, individuals cannot directly enforce their data protection rights. They are required to do so through a regulator-administered complaints process that functions as an enforcement gatekeeper thus raising questions about whether regulators are willing to pursue, and are resourced, to pursue highly complex and often cross-jurisdictional

²⁰⁸ Hartzog and Solove (above n 170) p 64.

²⁰⁹ Ibid.

²¹⁰ Daniel Solove, ‘Conceptualizing Privacy,’ (2002) 90 *California Law Review* 1087.

enforcement actions. Even if individuals had sufficient knowledge that their data protection rights have been interfered with and they were entitled to exercise those rights personally, the significant power imbalances between them and the (often large, well-resourced and overseas-based) companies would make enforcement impractical and prohibitively expensive.

The widespread use of personal data for AI training highlights conceptual problems with data protection rights in an era of AI, as well as the limitations of Australia's data protection laws. This section addresses these legislative limitations and then turns to the limitations of the concept of data protection rights in the context of AI.

Australian Information Privacy Law: Background

All data protection laws, including the GDPR and Australian Commonwealth, state and territory information privacy laws are derived from OECD Guidelines,²¹¹ which, in turn, were built on a set of *Fair Information Practice Principles* developed to govern dealings in personal information by US federal agencies.²¹² The rights they establish are not absolute.

Those principles continue to form the foundation of the current *Australian Privacy Principles* (APPs)²¹³ and, together with the *International Covenant on Civil and Political Rights*, are part of the Constitutional underpinnings of the Privacy Act.²¹⁴

In general terms, the Privacy Act is designed to regulate information privacy in the Commonwealth public sector and the private sector on a national basis. State and territory laws are primarily designed to regulate information privacy in the relevant state or territory public sector.²¹⁵

The Privacy Act

The Privacy Act provides individuals with a *measure of control* over their personal information via a set of APPs that govern the collection and handling of personal

²¹¹ Organisation for Economic Cooperation and Development, *Recommendation of the Council concerning Guidelines Governing the Protection of Privacy and Transborder Flows of Personal Data*, 23 September 1980, amended 11 July 2013. See <<https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0188>>.

²¹² These continue to be overseen by the Federal Privacy Council. The Principles are set out and discussed at <<https://www.fpc.gov/resources/fipps/>>.

²¹³ These are set out in Schedule 1 of the *Privacy Act 1988* (Cth) ('*Privacy Act*').

²¹⁴ These expressly engage the Commonwealth's external affairs power to extend the scope and application of the *Privacy Act* beyond the Commonwealth's other legislative powers under s 51 of the Australian Constitution. They are referred to in the preface to the *Privacy Act*. For example, they extend the application of the legislation to individuals and non-incorporated entities.

²¹⁵ Given their limited application to the public sectors of the states and territories, these information privacy laws are not discussed. However, as they are based around the OECD Principles, most of the comments in this section that relate to the principles are generally applicable to them.

information. Under the Privacy Act, information privacy obligations are not designed to constitute absolute protections of personal information. Section 2A of the Privacy Act makes this intent explicit, stating that its objects include ‘to recognise that the protection of the privacy of individuals is balanced with the interests of entities in carrying out their functions or activities’²¹⁶ and ‘to facilitate the free flow of information across national borders while ensuring that the privacy of individuals is respected.’²¹⁷

The Privacy Act applies to ‘APP entities,’²¹⁸ a term that encompasses agencies (i.e., Commonwealth public sector entities) and organisations (i.e., in broad terms private sector organisations, including corporations,²¹⁹ individuals,²²⁰ partnerships,²²¹ unincorporated associations,²²² and trusts²²³).

Scope limitations

There are two significant gaps in the Privacy Act’s private sector coverage. The first, which is unique to Australia, is the exclusion of small business operators from the definition of ‘organisation.’²²⁴ A business qualifies for exemption as ‘small business operator’ if it has an annual turnover of \$3 million or less.²²⁵ This definition potentially covers the vast majority of Australian businesses,²²⁶ although it excludes businesses that provide health services and hold health information, credit reporting entities, operators of residential tenancy databases and organisations that trade in personal information.²²⁷

The second is the exemption for employee records, which applies to acts or practices directly related to a current or former employment relationship between an employer and an individual and an employee record held by an organisation.²²⁸

²¹⁶ *Privacy Act*, ss 2A(b).

²¹⁷ *Privacy Act*, ss 2A(f).

²¹⁸ That is, the Commonwealth agencies and private sector organisations to which the *Privacy Act* applies.

²¹⁹ *Privacy Act*, s 6C(b).

²²⁰ *Privacy Act*, s 6C(a).

²²¹ *Privacy Act*, s 6C(c).

²²² *Privacy Act*, s 6C(d).

²²³ *Privacy Act*, s 6C(e).

²²⁴ *Privacy Act*, s 6C(1).

²²⁵ *Privacy Act*, s 6D.

²²⁶ According to the Australian Small Business and Family Enterprise Ombudsman, about ninety-two per cent of Australian Businesses had a turnover of less than \$2 million as at 30 June 2023. As this calculation is based on GST registrations and returns which do not have to be filed unless a business has a turnover of more than \$75,000.00 this, combined with the unavailability of precise statistics in relation to the \$3 million threshold, means that these calculations are an underestimate. See Australian Government, Australian Small Business and Family Enterprise Ombudsman, *Number of small businesses in Australia*, August 2024. See <https://www.asbfeo.gov.au/sites/default/files/2024-09/Number%20of%20small%20businesses%20in%20Australia_Aug%202024.pdf>.

²²⁷ *Privacy Act*, s 6D(4).

²²⁸ *Privacy Act*, s 7B(3).

Another significant limitation is that, with the exception of the recent enactment of a privacy tort,²²⁹ the Privacy Act does not directly confer legal rights on individuals. Breaches of the Privacy Act are dealt with through a complaints process administered by the Privacy Commissioner.²³⁰ The rights and protections established under the APPs are thus mediated rights that can only be enforced through the complaints process. Navigating this process involves the Commissioner first deciding whether to entertain the complaint²³¹ and, if so, undertaking a conciliation and investigation process.²³² Where a complaint is substantiated the Commissioner has several remedies available, including the payment of compensation.²³³

Limitations of the Australian Privacy Principles

The Privacy Act is designed to be technology-neutral, principles-based regulation. This is achieved through the application of the APPs,²³⁴ which set out a series of rules that govern the collection and handling of personal information.

The APPs require that personal information is managed openly and transparently, primarily through establishing practices and procedures that ensure that an entity complies with the APPs. The primary transparency obligation is to ‘have a clearly expressed and up-to-date privacy policy.’²³⁵ As will be seen, this requirement provides scant substantive protection other than to ensure that a privacy policy is provided and made freely available.

The requirements for the collection of personal information vary, depending on the entity collecting the information and whether the information is personal or sensitive information. For personal information, an agency must not collect personal information ‘unless the information is reasonably necessary for, or directly related to, one of its functions or activities.’²³⁶ Collection limitation requirements are considered to constitute data minimisation requirements, which ensure that only necessary personal information is collected. However, there are many instances where data collection is maximised rather than minimised and the concept of ‘reasonably necessary’ is interpreted and applied elastically.²³⁷

²²⁹ This is discussed below.

²³⁰ *Privacy Act* Part V, Division 1.

²³¹ *Privacy Act*, s 41.

²³² *Privacy Act*, ss 40A and 41.

²³³ *Privacy Act*, ss 52(1), 52(1A) and 52(1AB).

²³⁴ The Australian Privacy Principles (APPs) are set out in Schedule 1 of the *Privacy Act*.

²³⁵ APP 1.3.

²³⁶ APP 3.1, my emphasis. Neither the term ‘reasonably’ nor ‘reasonably necessary’ are defined. The concept of ‘reasonableness’ is used frequently in the APPs. It appears in each of the thirteen APPs.

²³⁷ See, for example, the discussion of the ChatGPT privacy policy below.

Private sector organisations may collect personal information only if it is reasonably necessary for ‘one or more [their] functions or activities.’²³⁸ But organisations can engage in any activity, and these can change over time. In other words, what is intended to limit the collection of personal information does not have that effect in practice provided the organisation has a privacy policy that nominates the purposes of collection: the purposes of collection can be at large.

For personal information that is sensitive information, the limitations on collection of personal information are narrower as it is generally collectable only with consent.²³⁹

Sensitive information is defined to include information or opinion about an individual’s racial or ethnic origin, political opinions, religious beliefs or affiliations, health information, genetic information, biometric information to be used for automated verification or identification, and biometric templates.²⁴⁰

However, the restrictions intended to flow from the additional consent requirement are somewhat theoretical. ‘Consent’ is defined broadly and can be given expressly or implicitly.²⁴¹ The requirements for giving a valid consent are that it be informed, voluntary, current and specific and that the individual giving consent has the capacity to do so.²⁴²

However, the existence of implied consent means that a failure to object to the collection of personal information can be taken to constitute consent.²⁴³ Furthermore, it is typically obtained through adhesion-type privacy policies and terms and conditions of service that are non-negotiable.

These difficulties are not ameliorated by requirements, under APP 5, to provide individuals with notification of the collection of their personal information. The notification requirement is that before, or at the time of, ‘or if that is not practicable, as soon as practicable after’²⁴⁴ collection the relevant entity must ‘take such steps (if any) as are reasonable in the circumstances’²⁴⁵ to notify the individual that their personal information has been collected. The notification is required to, amongst other things, let individuals know if the entity

²³⁸ APP 3.2.

²³⁹ APP 3.3(a).

²⁴⁰ *Privacy Act*, s 6(1)(e).

²⁴¹ *Privacy Act*, s 6(1).

²⁴² Office of the Australian Information Commissioner, *Australian Privacy Principles Guidelines: Key Concepts*. See <<https://www.oaic.gov.au/privacy/australian-privacy-principles/australian-privacy-principles-guidelines/chapter-b-key-concepts#ftnref35>>.

²⁴³ David Lindsay, *Some Problems of ‘Consent’ Under Australian Data Privacy Law*, (2022) 3(3) *Global Privacy Law Review*, p 116.

²⁴⁴ APP 5.1

²⁴⁵ *Ibid*.

collects the individual's personal information from someone else;²⁴⁶ the purposes of collection²⁴⁷ and the main consequences if the information is not collected.²⁴⁸ In practice the collection notification is often bundled with the entity's privacy policy and is subject to the associated limitations discussed earlier.

The other key provisions that relate to the use of personal information for AI training data are those that cover the use and disclosure of personal information. These terms are often collectively referred to as 'sharing.'

Neither 'use' nor 'disclosure' are defined in the Privacy Act. 'Use' is considered to be the handling and management of personal information within an entity's effective control: it is concerned with how personal information is handled internally.²⁴⁹ 'Disclosure' occurs where an entity makes personal information available 'to others outside the entity and releases the subsequent handling of the personal information from its effective control.'²⁵⁰ In general, entities are permitted to use and disclose personal information for the purposes for which the information was collected. The information cannot be used or disclosed for another purpose (a secondary purpose) unless the individual consents²⁵¹, where the individual would reasonably expect the entity to 'to use or disclose the information for the secondary purpose,'²⁵² or another permitted secondary purpose is available. If the information is sensitive information, the secondary purpose must be directly related to the primary purpose,²⁵³ If not, it must be related to the secondary purpose.²⁵⁴ The Privacy Act does not define 'related' or 'directly related.' The effect of both terms depends on the context and circumstances. As the NSW Court of Appeals observed when considering a similar requirement in NSW privacy legislation:

'[i]t is not necessary to express any concluded view on the metes and bounds of "directly related," and it is probably impossible to do so in the abstract. In any particular case, it will turn on the size and nature of the public sector agency, and the purpose of collection. In each case a judgement needs to be made whether there is a direct, as opposed to indirect or tangential or incidental, relationship between a lawful purpose and one of the agencies activities or functions.'²⁵⁵

²⁴⁶ APP 5.2(b)(i) and (ii)

²⁴⁷ APP 5.2(d)

²⁴⁸ APP 5.2(e)

²⁴⁹ *Australian Privacy Principles Guidelines: Key Concepts* (above n 242)

²⁵⁰ *Ibid.*

²⁵¹ APP 6.1(a)

²⁵² APP 6.2(a)

²⁵³ APP 6.2(a)(ii).

²⁵⁴ APP 6.2(a)(i).

²⁵⁵ *Norkin v University of New England*, [2023] NSWCA 194, [56].

In either case, the scope for using and disclosing personal information is broad and the restrictions, based around concepts such as ‘consent,’ reasonable expectations and ‘directly related or ‘related’ purposes are fluid and inchoate, susceptible to being shaped and manipulated by broadly drafted privacy policies and click wrap adhesion terms and conditions of service. The consequence is that personal information can easily find its way into training data for large language models (LLMs).

Finally, there are limitations on the period for which entities can retain personal information. If personal information is no longer needed ‘for any purpose for which the information may be used or disclosed’²⁵⁶ the entity concerned ‘must take such steps as are reasonable in the circumstances to destroy the information or to ensure that the information is de-identified.’²⁵⁷ The requirement provides weak protection as the entity that has collected it and has specified, under its privacy policy, how it will be used and disclosed can retain it for prolonged periods provided those purposes remain.

The role of contractual liability limitations and privacy policies

The technology industry and digital platforms routinely limit their liability to end users through the use of adhesion contracts that disclaim all responsibility for the use of the AI system to the fullest extent permitted by law. ‘Online ‘contracts’ such as terms-of-service or terms-of-use agreements are also referred to as ‘click wrap’ because, as a great deal of research shows, most people get wrapped in these oppressive contract terms by simply clicking on the box that says ‘I agree’ without ever reading the agreement.’²⁵⁸ In many cases the mere use of the service is taken to constitute the establishment of a contractual relationship that excludes all liability that is not otherwise excludable under law.

Likewise, the effectiveness of data protection laws and privacy policies, which are mandated by the Privacy Act,²⁵⁹ are significantly circumscribed by adhesion-type constraints. APP entities, i.e., Commonwealth public sector agencies and private sector organisations,²⁶⁰ are required to collect personal information only if it is reasonably necessary for, or directly related to, their functions or activities. Consent to collection is not required unless the

²⁵⁶ APP 11.2(b).

²⁵⁷ APP 11.2.

²⁵⁸ Shoshana Zuboff, *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power*, New York, 2019, p 48.

²⁵⁹ These are required under APP Principle 1.3.

²⁶⁰ In general, an ‘agency’ is a Commonwealth public sector organisation under s 6(1) and an organisation is a private sector organisation under Privacy Act, s 6C.

personal information is also sensitive information,²⁶¹ although the collection must be by lawful and fair means.²⁶² Once collected, personal information can be used or disclosed for a secondary purpose where consent has been given or the individual would reasonably expect the use or disclosure for the secondary purpose and, if the personal information is not sensitive information, the secondary purpose is related to the primary collection purpose or, if it is, that it is directly related to the secondary purpose.

The use and disclosure provisions are frequently used in such a way that users of digital services are taken to consent to the widespread use and disclosure of their personal information through broadly drafted privacy notices and policies that enable the widespread 'sharing' of their personal information for purposes they would not have anticipated.

Personal data constitutes a significant proportion of the data that finds its way into training data for generative AI. Unbeknownst to most individuals, they consent to such a use on a 'take it or leave it' basis. If an individual fails to click the 'I agree' button, they cannot use the service, such as the commonly used digital platforms that have become an essential part of their lives. Alternatively, consent is implied through the continued use of the digital service.

Despite data protection law reform in Australia, these fundamental problems concerning the loosely regulated collection, use and disclosure of personal data that are then included in LLMs continues without any meaningful solutions in sight. Although consent requirements under the GDPR are stricter and provide a higher degree of protection for EU residents, their effectiveness is still limited. Many digital services and platforms that harvest personal information that is later used to train AI have become essential to contemporary forms of social and commercial interaction and are designed to be addictive.²⁶³ Their users willingly trade off their personal data for the 'free' service they receive in an example of the aphorism that 'if something is free, you're not the customer; you're the product.'²⁶⁴

Applying the Privacy Act to OpenAI's ChatGPT Privacy Policy

An example of the limitations of the Privacy Act can be seen in how it relates to the ChatGPT privacy policy.²⁶⁵ The personal data that is collected by ChatGPT is extensive. Apart from

²⁶¹ APP 3.

²⁶² APP 3.5.

²⁶³ Christine Montag, Bernd Lachman, Marc Herrlich, Katharina Zweig, 'Addictive Features of Social Media/Messenger Platforms and Freemium Games against the Background of Psychological and Economic Theories' [2019] *Int. J. of Environmental Research and Public Health*, p 1.

²⁶⁴ Bruce Schneier, *Data and Goliath: The Hidden Battles to Collect your Data and Control Your World*, New York, 2025, p 35.

²⁶⁵ OpenAI, *Privacy Policy*, 27 June 2025. See <<https://openai.com/policies/row-privacy-policy/>>.

typical categories such as account information; communication information;²⁶⁶ and other information,²⁶⁷ it also encompasses ‘user content’ which is defined as:

personal data that you provide in the input to our Services (“Content”), including your prompts and other content you upload, such as files, images, and audio, depending on the features you use.²⁶⁸

The effect is that all content a user inputs into ChatGPT is collected. The content is at large and can include any material, whether or not the user is legally entitled to upload it; for example, images or audio or video files where the user has no legal right to reproduce, or authorise the reproduction of, the content under copyright law.

The breadth of OpenAI’s stated use of that data is also extensive. ChatGPT may use personal data to ‘provide, analyze and maintain services;’ and to ‘improve and develop our Services.’²⁶⁹ This anodyne description begets many layers of meaning. Putting aside the undefined nature of the ‘Services,’ the use of personal data is extensive in that it can contribute to providing services, making them more effective and extending them. As personal data is the raw material used to deliver and augment ChatGPT, this means that any personal data inputted into it is processed and consumed by it to enhance its capabilities. It also means that the personal information can function to enable ChatGPT to personalise its responses to the user in ways that are opaque and unpredictable.

Disclosure is also open-ended. The Privacy Policy specifies that disclosure can encompass the transfer of personal information to ‘Vendors and Service Providers’ to Open AI;²⁷⁰ for what are referred to as ‘Business Transfers;’²⁷¹ ‘Government Authorities or Other Third

²⁶⁶ OpenAI, *Privacy Policy*, 27 June 2025, Clause 1. This is defined to include information a user communicates to OpenAI through email or ‘our pages on social media sites.’

²⁶⁷ Defined as ‘other information you provide to us’ such as when participating in surveys or information provided to establish identity or age.

²⁶⁸ *Ibid.*

²⁶⁹ OpenAI, *Privacy Policy*, 27 June 2025, cl 2.

²⁷⁰ These include but are expressly not limited to ‘providers of hosting services, customer service vendors, cloud services, content delivery services, support and safety monitoring services, email communication software, web analytics services, payment and transaction processors, and other information technology providers. Pursuant to our instructions, these parties will access, process, or store Personal Data only in the course of performing their duties to us.’

²⁷¹ These are defined as ‘strategic transactions, reorganization, bankruptcy, receivership, or transition of service to another provider (collectively, a “Transaction”), your Personal Data may be disclosed in the diligence process with counterparties and others assisting with the Transaction and transferred to a successor or affiliate as part of that Transaction along with other assets.’

Parties;²⁷² ‘Affiliates’;²⁷³ ‘Business Account Administrators’;²⁷⁴ and ‘Other Users and Third Parties You Interact or Share Information With.’²⁷⁵ It is difficult to conceive of a disclosure of personal information that is not covered by one of these categories. Moreover, when ‘content’ is inputted into ChatGPT and then disclosed (shared) with one or more of the nominated categories of disclosees, the personal information is collected by them and becomes subject to the disclosee’s privacy policy. This cycle of collection, use and disclosure (sharing) cascades across entity after entity in a manner where it is a practical impossibility for the data subject to know or understand where and with whom their personal data is shared, thus undermining the central assumptions on which information privacy law is based: privacy self-management and personal control.

The conceptual break-down of data protection in an era of AI

Information privacy law was conceived and designed ‘long before the emergence of the commercial internet, let alone social media platforms and generative AI tools.’²⁷⁶ The privacy principles that have underpinned, and continue to underpin, Australia’s information privacy laws do not provide adequate protections in a technological environment where ‘continued AI development will continue to increase developers’ hunger for data – the foundation of AI systems.’²⁷⁷ There are a number of specific reasons why data protection laws are failing in the wake of a data onslaught from AI.

The first is that the current principles-based approach was developed with dual objectives: providing individuals with a measure of control over their personal information whilst supporting the free flow of personal information. This has always involved a compromise, frequently referred to as a ‘balancing’ of these dual interests. In Australia this balancing has,

²⁷² This disclosure is described as ‘We may share your Personal Data, including information about your interaction with our Services, with government authorities, industry peers, or other third parties in compliance with the law (i) if required to do so to comply with a legal obligation, or in the good faith belief that such action is necessary to comply with a legal obligation, (ii) to protect and defend our rights or property, (iii) if we determine, in our sole discretion, that there is a violation of our terms, policies, or the law; (iv) to detect or prevent fraud or other illegal activity; (v) to protect the safety, security, and integrity of our products, employees, users, or the public, or (vi) to protect against legal liability.’

²⁷³ This disclosure is described as ‘We may disclose Personal Data to our affiliates, meaning an entity that controls, is controlled by, or is under common control with OpenAI. Our affiliates may use this Personal Data in a manner consistent with this Privacy Policy.’

²⁷⁴ This disclosure primarily relates to users who have paid ChatGPT subscriptions and is described as ‘When you join a ChatGPT Enterprise or business account, the administrators of that account may access and control your OpenAI account, including being able to access your Content. In addition, if you create an account using an email address belonging to your employer or another organization, we may share the fact that you have an account and certain account information, such as your email address, with your employer or organization to, for example, enable you to be added to their business account.’

²⁷⁵ This disclosure is described as ‘Certain features allow you to interact or share information with other users or third parties. For example, you can share ChatGPT conversations with other users via [shared links \(opens in a new window\)](#). You can also send information to third-party applications, such as via custom actions for GPTs, or for searching the web to help answer questions that benefit from more recent information. Information you share with third parties is governed by their own terms and privacy policies, and you should make sure you understand those terms and policies before sharing information with them.’

²⁷⁶ Jennifer King and Caroline Meinhardt, (above, n 64) p 37.

²⁷⁷ *Ibid*, p 17.

as discussed earlier, involved many trade-offs that result in broad exemptions from the scope of the Privacy Act, such as the small business and employee records exemptions. As so much Australian personal information is simply not covered by any legislative protections, there is a high risk that it is not protected by any information privacy rights at all.

Secondly, the rights conferred under the Privacy Act are largely indirect, mediated rights that can only be exercised through a gatekeeper, the Privacy Commissioner. Although some individuals may find it helpful to be able to enlist the resources and investigatory powers of a government-appointed regulator, the exercise of rights is at the discretion of the regulator.

Thirdly, *even if* information privacy rights vested in individuals:

[m]any rights are not practical for individuals to exercise. Rights put too much onus on individuals to fight a war they can't win. Attempting to use privacy rights as a primary way to protect privacy is akin to arming an individual with a dagger to fight an entire army. People can't exercise their rights in the kind of systematic way necessary to have a meaningful impact.²⁷⁸

Privacy rights involve individuals undertaking privacy self-management. But 'people lack the expertise to make meaningful choices about their data'²⁷⁹ and to make appropriate assessments regarding the collection, use or disclosure of their personal data. Even when they may have the necessary skills to do so, they are faced with an information landscape stretching across the world that is dauntingly difficult to navigate and risk manage: the latitude given by the APPs means that personal data could be held anywhere without the individual's knowledge. Where AI is involved, particularly generative AI that can make inferences about individuals that constitute personal information, it is virtually meaningless to vest self-management responsibilities on the individual concerned.

Finally,

privacy can't be protected at the level of the atomistic individual. Individuals make privacy choices that have effects not just for themselves but for many others. For example, sharing one's genetic data shares the genetic data of one's family members. In today's world of machine learning, the personal data of everyone in a data set has an impact on the decisions that the system makes.²⁸⁰

AI's functionality enables it to *create* personal data in ways that were not envisaged when information privacy laws were first developed, for example by processing personal data

²⁷⁸ Daniel Solove, *The Limitations of Privacy Rights*, (2023) 98(3) Notre Dame Law Review 975, 978.

²⁷⁹ Ibid.

²⁸⁰ Ibid pp 978-979.

taken from a variety of sources to draw inferences, or make predictions, about individuals, thus undermining the assumptions about how the personal data information lifecycle functions.

Reform of the Privacy Act

The Australian Competition and Consumer Commission's (ACCC) Digital Platforms Inquiry (DPI Inquiry) in 2019²⁸¹ made recommendations to reform the Privacy Act to address a range of new and emerging information privacy risks.²⁸² It made specific recommendations,²⁸³ as well as more general, recommendations for broader reform of Australian privacy law 'to effectively protect consumers' personal information in light of the increasing volume and scope of data collection in the digital economy.'²⁸⁴ Although the DPI Inquiry touched on AI-associated issues, in a pre-ChatGPT environment the main thrust of its privacy-related recommendations were not specifically directed at AI data protection risks.

The outcome of the ACCC's general recommendations was to pave the way for a review of the Privacy Act. A Privacy Act Review Report was produced in 2022²⁸⁵ that covered three main themes: the scope and application of the Privacy Act; the protections conferred by it; and uplifting regulatory and enforcement measures. It produced one hundred and sixteen proposals. A government response published in 2023²⁸⁶ was designed to set out the reform pathway. Of the Review Report proposals, government agreed to thirty-eight; agreed in principle to sixty-eight²⁸⁷ and noted an additional ten.

The Commonwealth government has indicated that information privacy reform would be undertaken incrementally. To this end 'Tranche 1' reforms were enacted in November 2024.²⁸⁸ Most came into effect on 11 December 2024. The most relevant reform measure for AI was the addition to APP 1 of a requirement to include in privacy policies information about automated decision-making. This does not come into effect until 10 December 2026.

²⁸¹ Australian Competition and Consumer Commission, *Digital Platforms Inquiry Final Report*, June 2019. See <<https://www.accc.gov.au/system/files/Digital%20platforms%20inquiry%20-%20final%20report.pdf>>.

²⁸² These are set out in recommendations 16-19 at pp 34-37 of the Report.

²⁸³ These are contained in recommendation 16 and include updating the definition of 'personal information' to capture technical data such as IP addresses.

²⁸⁴ Recommendation 17.

²⁸⁵ Australian Government, *Privacy Act Review Report 2022*. See <https://www.ag.gov.au/sites/default/files/2023-02/privacy-act-review-report_0.pdf>.

²⁸⁶ Australian Government, *Government Response: Privacy Act Review Report*, 28 September 2023. See <<https://www.ag.gov.au/sites/default/files/2023-09/government-response-privacy-act-review-report.PDF>>.

²⁸⁷ Note that "Agrees in-principle" indicates that further engagement with entities and a comprehensive impact analysis is needed before the Government makes a final decision on implementation of a proposal.' See Fact Sheet: Government Response to the Privacy Act Review Report, p2. See <<https://www.ag.gov.au/sites/default/files/2023-09/fact-sheet-government-response-privacy-act-review-report.PDF>>.

²⁸⁸ *Privacy and Other Legislation Amendment Act 2024* (Cth).

The requirement applies where an entity arranges for a computer program to ‘make or do a thing that is substantially and directly related to making a decision;’ ‘the decision could reasonably be expected to significantly affect the rights or interests of an individual;’ and ‘personal information about the individual is used in the operation of the computer program to make the decision or do the thing that is substantially and directly related to making the decision.’²⁸⁹ The privacy policy must contain information about the kinds of personal information used in the computer programs; the kinds of decisions made solely by the computer programs; and the kinds of decisions made.²⁹⁰ Positive features of this reform measure are that it extends beyond decisions based *solely* on automation and thus is broader than the equivalent GDPR protection²⁹¹ and that the use of ‘computer program’ in the requirement is not limited to machine learning AI but could also apply to weak AI. That said, the reform simply adds to existing transparency requirement and does not confer a right *not to be* subject to automated decision-making. This additional disclosure obligation is unlikely to provide individuals with substantive protections from AI risks: they will just know more about them.

Copyright

There are a variety of risks that AI poses to the owners of intellectual property. These arise over a range of modalities. One of the most important is in relation to copyright, but risks also arise for other categories of intellectual property rights such as trademarks, moral rights, performers’ rights and trade secrets or confidential information.²⁹²

All of these laws were designed before concerns emerged about the use of intellectual property to develop, train and validate AI that uses vast quantities of data. This raises complex issues about how these laws apply to the use of intellectual property to this new use of copyright material. Does this type of use infringe existing intellectual property rights?

Copyright background

The key copyright risks posed by AI is that the training data assembled by AI developers includes copyright material that has been copied without permission of the copyright owner.

²⁸⁹ *Privacy and Other Legislation Amendment Act 2024* (Cth), s 88, APP 1.7.

²⁹⁰ *Privacy and Other Legislation Amendment Act 2024* (Cth), s 88, APP 1.8.

²⁹¹ GDPR, Art 22(1).

²⁹² Australia has not yet developed a separate category of intellectual property protection for database rights. The EU has developed a *sui generis* database right under EU Directive 96/9/EC. See <<https://eur-lex.europa.eu/eli/dir/1996/9/oj/eng>>.

Copyright confers a bundle of rights on the authors of works and those who make subject matter other than works. In general terms, a copyright owner has a time-limited monopoly to exercise control over a variety of forms of exploitation of copyright-protected works and other subject matter.

Under the *Copyright Act 1968* (Cth), (Copyright Act) the rights of copyright owners in literary, dramatic and musical works include the right to reproduce the work in a material form;²⁹³ to publish the work;²⁹⁴ to perform the work in public;²⁹⁵ to communicate the work to the public;²⁹⁶ and to make an adaptation of the work.²⁹⁷ For artistic works, copyright owners have the right to reproduce;²⁹⁸ to publish;²⁹⁹ and to communicate the artistic work to the public.³⁰⁰

Copyright in subject matter other than works is more varied. For copyright in sound recordings, films and videos, it includes the right to make a copy;³⁰¹ to cause it to be heard or seen in public;³⁰² and to communicate it to the public.³⁰³

Copyright is infringed if a person does, or authorises the doing of, one or more of the copyright owners' exclusive rights with permission or authorisation³⁰⁴ in relation to a substantial part of the work or other subject matter.³⁰⁵

The substantiality test is qualitative, not quantitative. Four criteria are used to determine whether copying is substantial:

First, the volume of the material taken, bearing in mind that quality is more important than quantity; secondly, how much of the material is the subject-matter of copyright and how much is not; thirdly, whether there has been an *animus furandi* on the part of the defendant; this was treated by Page-Wood V.C. in *Jarrold v Houlston* (1875) 3K. and J. 708 as equivalent to an intention on the part of the defendant to take for the purpose of saving himself labour; fourthly, the extent to which the plaintiff's and the defendant's books are competing works.³⁰⁶

For the purposes of AI training, only substantial copies of a work or other subject matter will infringe copyright. However, given that the dominant purpose of assembling training data is

²⁹³ Copyright Act, 1968 (Cth) (Copyright Act) s 31(1)(a)(i).

²⁹⁴ Copyright Act, s 31(1)(a)(ii).

²⁹⁵ Copyright Act, s 31(1)(a)(iii).

²⁹⁶ Copyright Act, s 31(1)(a)(iv).

²⁹⁷ Copyright Act, s 31(1)(a)(v).

²⁹⁸ Copyright Act, s 31(1)(b)(i).

²⁹⁹ Copyright Act, s 31(1)(b)(ii).

³⁰⁰ Copyright Act, s 31(1)(b)(iii).

³⁰¹ Copyright Act, ss 85(1)(a) and s 86(a).

³⁰² Copyright Act, ss 85(1)(b) and 86(b).

³⁰³ Copyright Act, ss 85(1)9C0 and 86(c).

³⁰⁴ For works, Division 2, and in particular, s 36(1) and for other subject matter, Division 6 and, in particular, s101(1).

³⁰⁵ Copyright Act, s 14.

³⁰⁶ *Ravenscroft v Herbert and New English Library* [1980] RPC 193, 203.

to capture as much data as possible, questions regarding substantiality are, overall, likely to be of little significance.

AI that relies on access to large datasets in order to train, test and validate AI models, in particular generative AI, involves accessing and using data, typically in the form of text, images and video. AI training data ‘refers to datasets used for AI system development, including for training, testing, evaluation, verification and validation. They include both structured and unstructured data, such as numerical, text, image or audio data, collected through human or machine means.’³⁰⁷ These can be obtained from sources such as digitized books, social media platforms, online forums, online databases newsletters, blogs and similar. In many cases the material obtained from these sources is protected by copyright and permission to reproduce or use copyright works or other subject matter is not sought and obtained from copyright owners.

These techniques involve a high risk that the copying of content occurs without the authority of the owner of the site or service in which the data is embodied. In many cases the owner of the site or service may not be the owner of the copyright in the data taken. For example, a social media platform that publishes an image posted by one of its users, is not the owner of the copyright in the image: the user is the copyright owner. Moreover, neither the source of the scraped data nor the underlying owner of the copyright in the scraped data may have any knowledge that the copyright work or other subject matter has been reproduced.

Copyright reform?

Australian copyright law contains a number of exceptions that enable temporary reproductions of works or audio-visual items as part of a technical process to facilitate their use in a digital environment.³⁰⁸ However, they do not cover the long-term copying of material that is used in training data.

³⁰⁷ Organisation for Economic Cooperation and Development, *Intellectual Property Issues in Artificial Intelligence Trained on Scraped Data*, OECD Artificial Intelligence Papers, No. 33, February 2025, p 16. See <https://www.oecd.org/en/publications/intellectual-property-issues-in-artificial-intelligence-trained-on-scraped-data_d5241a23-en.html>.

³⁰⁸ These cover a range of technical uses such as for purposes associated with making or receiving a communication (*Copyright Act 1968 (Cth)*, ss 43A and 111A); reproductions incidentally made as a necessary part of a technical process (*Copyright Act 1968 (Cth)* ss 43B and 111B); and automated caching (*Copyright Act 1968 (Cth)*, s 200A).

Several jurisdictions have developed text and data mining (TDM) exceptions to copyright. In the EU text and data mining for the purposes of scientific research is an exception to the rights of copyright owners³⁰⁹.

More important is the issue of whether data scraping that involves the reproduction of copyright material is a fair dealing. These issues are discussed earlier in this chapter.

Fault-based liability: negligence

The law of negligence is the 'paradigmatic common law approach to attributing fault for accidents.'³¹⁰

Proof of a cause of action in negligence involves establishing a duty of care; breach of that duty; and causation of damage.³¹¹ Liability for harm only occurs where a duty of care exists to prevent it. If such a duty of care exists a plaintiff must prove on the balance of probabilities that the defendant was at fault.³¹² A defendant 'breaches the duty of care owed to the plaintiff by failing to take reasonable steps to avoid reasonably foreseeable not insubstantial risk of injury.'³¹³ Duties of care are:

not owed in the abstract. Rather, they are obligations of a particular scope, and that scope may be more or less expansive depending on the relationship in question. Secondly, whatever their scope, all duties of care are to be discharged by the exercise of reasonable care. They do not impose a more stringent or onerous burden.³¹⁴

Statutory provisions have been enacted in each of the Australian states that specify three conditions to a finding of breach of duty. In Victoria these relevantly state that:

A person is not negligent in failing to take precautions against a risk of harm unless—
(a) the risk was foreseeable (that is, it is a risk of which the person knew or ought to have known); and
(b) the risk was not insignificant; and
(c) in the circumstances, a reasonable person in the person's position would have taken those precautions.³¹⁵

These requirements are conjunctive. Each must be satisfied before the standard of care can be considered in any particular claim. Although there are a number of elements that must

³⁰⁹ A 3, EU Directive 2019/790 of 17 April 2019 on copyright and related rights in the Digital Single Market.

³¹⁰ Henry Fraser and Nicolas Suzor, 'Locating Fault for AI Harms: A Systems Theory of Foreseeability, Reasonable Care and Causal Responsibility in the AI Value Chain,' (2025) 17(1) *Law, Innovation and Technology* 103.

³¹¹ *Donoghue v Stevenson* [1932] AC 562. See also *Gala v Preston* (1991) 172 CLR 243, in particular Brennan J at pp 257-259.

³¹² *TNT Management Pty Limited v Brooks* (1979) ALJR 267.

³¹³ Danuta Mendelson, *The New Law of Torts*, edn 3, Oxford University Press, 2014, p 362.

³¹⁴ *Roads and Traffic Authority of NSW v Dederer* [2007] HCA 42 [43].

³¹⁵ *Wrongs Act 1958 (Vic)*, s 48(1).

be established to successfully prosecute a claim in negligence, foreseeability of harm and the capacity to prevent that harm by taking reasonable care are fundamental to the success of such a claim. The law of negligence 'will not attribute fault unless harm is a foreseeable consequence of a person's failure to exercise reasonable care, and it can be shown that reasonable care would have prevented the harm.'³¹⁶

In principle, the law of negligence is sufficiently elastic to adequately address these issues in the case of first-wave AI:

Although the law has dealt with IT-based systems for some time, the ones it has encountered have hitherto normally been on a fairly low level of sophistication. As a result, they have required no particular adaptation or alteration of traditional legal reasoning ... There is precious little of AI, as against simple electronic logic, about most of these applications. The computers that carry them out are in essence just super-efficient machines used by humans, and the traditional scheme of liability for negligence or breach of contract is generally adequate.³¹⁷

However, the ability of negligence law to address second wave, machine learning based AI becomes significantly more problematic. Proving the essential elements of negligence in those circumstances to the civil standard of the balance of probabilities is highly challenging:

AI poses serious challenges for negligence law's ability to continue compensating the injured. By inserting a layer of inscrutable, unintuitive, and statistically derived code in between a human decisionmaker and the consequence of her decisions, AI disrupts our typical understanding of responsibility for choices gone wrong.³¹⁸

Such difficulties, which are inherent in the nature of these types of AI, disengage human responsibility and traditional forms of fault attribution, from AI instantiated loss and damage:

Unfortunately, once we turn to AI,³¹⁹ the process immediately becomes much more problematical. The workings of machines equipped with AI are not intended to be readily understandable, even on a general level, to people; indeed, many such devices are designed precisely to operate autonomously and independently of human supervision. In such a situation, the general rule in negligence, that there is no liability unless a human being ought to have spotted that something was wrong, gives seriously skewed results if applied unaltered. First, it means that in most cases of damage done by AI, a claimant will have either to go uncompensated or try to invoke some head of non-fault-based liability, such as product liability or liability under *Rylands v Fletcher*.³²⁰ And second, we now have the serious

³¹⁶ Henry Fraser and Nicolas Suzor, (above, n 310), p 111.

³¹⁷ Baris Soyer and Andrew Tettenborn, 'Artificial Intelligence and Civil Liability – Do we need a New Regime,?' (2022) 30 *International Journal of Law and Information Technology* 385, 386.

³¹⁸ Andrew Selbst, 'Negligence and AI's Human Issue,' (2020) 100 *Boston University Law Review* 1315.

³¹⁹ The authors are referring here to machine learning forms of AI.

³²⁰ In Australia, the rule in *Rylands v Fletcher* (1886) LR 1 Ex 265 was abolished by the High Court of Australia in *Burnie Port Authority v General Jones Pty Ltd* (1994) 179 CLR 520.

possibility that even if an error would give rise to damages when committed by a human, the self-same error having the self-same effect will create no liability if committed by a smart computer or an algorithm. This cannot be right.³²¹

These problems multiply when it is understood that AI is generally the product of a ‘value chain’ that consists of many actors, each of which may have contributed to loss of damage caused by AI. For example, the Productivity Commission identifies five main components to the AI value chain consisting of computer hardware (including the advanced computer chips required to power AI); cloud platforms that provide access to computing resources and data storage; general purpose AI models on which downstream AI applications can be built; model hubs and machine-learning operations, including tools to curate, fine-tune and manage AI models; and applications and services that utilise AI models ‘as is’ or that are fine tuned for specific use cases.³²² The multiplicity of the actors involved creates significant barriers for negligence-based accountability. Moreover, the dynamic and unpredictable way in which this type of AI functions accompanied by interactions with end users, for example through prompts or text and audio-based questions and answers, means that the same AI may produce outputs that vary significantly from user to user. For traditional theories of negligence to apply to second wave AI there needs to be either a reliable, consistent and practical way in which to ‘open’ the black box or readily available analytic methods that can overcome the forensic challenges of understanding *what* AI has done and *how* it has done it. Some suggested approaches to these issues are discussed in Chapter Six.

Administrative law and AI illegality

The Robodebt Programme

One of Australia’s highest profile and damaging instances of a very basic form of automated decision making – weak AI – going wrong was the Robodebt programme. It serves as an example of both automation that ‘twisted or ignored the requirements of the governing legislation’³²³ to pursue political objectives as well as the repeated failure of regulatory oversight to address unlawful automated decision-making.

³²¹ Baris Soyer and Andrew Tettenborn, (above, n 317) p 387.

³²² Productivity Commission, *Making the most of the AI Opportunity, Research Paper 1, AI Uptake, Productivity, and the Role of Government*, January 2024, p 6. See <<https://www.pc.gov.au/research/completed/making-the-most-of-the-ai-opportunity/ai-paper1-productivity.pdf>>.

³²³ Peter Hanks, *Administrative Law and Welfare Rights: a 40-year Story from Green v Daniels to “Robo debt recovery,”* (2017) 89 AIAL Forum 1.

The Robodebt programme was an initiative designed by the Australian government to reduce welfare fraud. Beginning in 2016, Centrelink, a portfolio agency of the Department of Human Services (DHS, now Services Australia) that was (and remains) responsible for administering welfare payments, introduced a new method for assessing and recovering what it referred to as debts, i.e., payments made to welfare recipients to which they were allegedly not wholly or partially entitled.³²⁴

This new method involved the use of an automated process to calculate welfare overpayments, that replaced a more labour intensive, manual assessment process. This was known as the online compliance intervention (OCI) system for raising and recovering debts. The Acting Ombudsman described the operation of the OCI system as follows.

The OCI matches the earnings recorded on a customer's Centrelink record with historical employer-reported income data from the Australian Taxation Office (ATO). Parts of the debt raising process previously done manually by compliance officers within DHS are now done using this automated process. Customers are asked to confirm or update their income using the online system. If the customer does not engage with DHS either online or in person, or if there are gaps in the information provided by the customer, the system will fill the gaps with a fortnightly income figure derived from the ATO income data for the relevant employment period ('averaged' data).³²⁵

The statutory framework that governed the establishment of an entitlement to a welfare payment took account of a person's eligibility to receive it, how it was to be calculated and the method by which any overpayment was to be recovered is the *Social Security Act 1991* (Cth) (SSA).

Under the SSA there are several methods applicable to the calculation of different welfare benefits.³²⁶ Relevantly, the statutory income test for Newstart (now the Jobseeker benefit, the primary unemployment welfare payment) was Module G in s 1068 of the SSA which imposed an income test to determine eligibility and the quantum of the benefit paid. The income test was based on the individual's averaged income, as defined in s 8, during the relevant fortnight for which the benefit was paid.

³²⁴ The full detail of the Robodebt scandal is set out in the *Report of the Royal Commission into the Robodebt Scheme, 2023*. (Robodebt Report). See <<https://robodebt.royalcommission.gov.au/system/files/2023-09/rrc-accessible-full-report.PDF>>.

³²⁵ Commonwealth Ombudsman, *Centrelink's Automated Debt Raising and Recovery System: A Report about the Department of Human Services' Online Compliance Intervention System for Debt Raising and Recovery*, April 2017, p 1. See <https://www.ombudsman.gov.au/__data/assets/pdf_file/0022/43528/Report-Centrelinks-automated-debt-raising-and-recovery-system-April-2017.pdf>.

³²⁶ See, for example, Chapter 3 of the SSA which sets out the general provisions relating to payability and rates. There is a general rate calculator (Part 3.1) as well as other methodologies for different benefits.

Instead of following this legal prescription, the OCI substituted a different form of averaging. It used income data taken from the Australian Taxation Office (ATO) spread evenly over the period of an individual's employment, over a period of up to twelve months, 'despite the social security income test using income received in each fortnight and despite the ATO earnings information using an income definition that differs from the definition in the Social Security Act.'³²⁷ The SSA required that entitlements be determined each two weeks, rather than on an averaged basis where the data was derived from the ATO. '[J]ust as rainfall in a given fortnight cannot be calculated by dividing the 'annual' figure by 26, nor can an 'average' from the ATO records of earnings over many weeks or months speak to what was earned in particular fortnights.'³²⁸ The simple algorithm that failed to observe the correct income test also struggled to account for other variations in data, including different spellings of employer names in the relevant databases and being unable to properly calculate individuals' work patterns such as casual work.³²⁹ This threshold illegality led to a cascade of other unlawful activity.

One of the most serious of these was the process established to recover the welfare overpayments wrongly assessed under the averaging process. By s 1222A(a) of the SSA an amount paid as a social security payment is 'a debt due to the Commonwealth if, and only if, ... a provision of this Act ... expressly provides that it is ...' Read in conjunction with s 1223(1), the effect of which is that where a welfare payment is made and the person who received it was not entitled to it 'the amount of the payment is a debt due to the Commonwealth by the person and the debt is taken to arise when the person obtains the benefit of the payment,' the existence of a debt must be established by the Commonwealth. However, the Robodebt procedure did not conform with this requirement. If the alleged overpayment was 'established' using ATO data and the individual did not update their data, the ATO income data was assumed to apply. This process 'was built on the premise that the onus lay on the social security recipient to prove that Centrelink's assumption as to the recipient's income was not correct – despite the SSA making the existence of the debt dependent on the Commonwealth establishing the receipt of amounts to which the recipient was not

³²⁷ Peter Hanks, (above, n 323) pp 15-16

³²⁸ Terry Carney, *Robo-Debt Illegality: A Failure of Rule of Law Protections?*, AUSPUBLAW, 30 April 2018. See <<https://www.auspublaw.org/blog/2018/04/robo-debt-illegality>>.

³²⁹ Tapani Rinta-Kahila, [Ida Someh](#), Nicole Gillespie, Marta Indulska, Shirley Gregor, *Algorithmic Decision-making and System Destructiveness: A case of Automatic Debt Recovery* (2022) 31 *European Journal of Information Systems* 313.

entitled.³³⁰ The notices sent to welfare recipients also used 'language indicating that a response was required, not just requested ("You need to tell us...") but they were not said to be notices from the secretary of DHS requiring information under any of the provisions of the *Social Security (Administration) Act* which enabled information to be obtained in that way.'³³¹ Finally a ten percent 'recovery fee' was added to debts raised against welfare recipients under the scheme. This, too, was unlawful as there was no lawful debt raised in the first place.³³² In the Federal Court's decision in the first Robodebt class action brought against the Commonwealth, Murphy J commented that:

[i]n the course of the proceeding the Commonwealth admitted that it did not have a proper legal basis to raise, demand or recover asserted debts which were based on income averaging from ATO data. The evidence shows that the Commonwealth unlawfully asserted such debts totalling at least \$1.763 billion against 433,000 Australians. Then, including through private debt collection agencies, the Commonwealth pursued people to repay these wrongly asserted debts, and recovered approximately \$751 million from about 381,000 of them.

The proceeding has exposed a shameful chapter in the administration of the Commonwealth social security system and a massive failure of public administration.³³³

As the Robodebt Royal Commission concluded:

Robodebt was a crude and cruel mechanism, neither fair nor legal, and it made many people feel like criminals. In essence, people were traumatised on the off-chance they might owe money. It was a costly failure of public administration, in both human and economic terms.³³⁴

One of the main sources of concern regarding the continuation of Robodebt was extensive failures within the executive to challenge it. As the Royal Commissioner stated in the preface to her final report:

At the outset of my inquiry, I had anticipated that the Commission would uncover how such a patently unreliable methodology as income averaging, without other evidence, to determine entitlement to benefit could become part of an Australian Government debt raising and recovery scheme. What has been startling in the Commission's investigation of the Robodebt scheme has been the myriad of other ways in which it failed the public interest.

It is remarkable how little interest there seems to have been in ensuring the Scheme's legality, how rushed its implementation was, how little thought was given to how it would affect welfare recipients and the lengths to which public servants were prepared to go to oblige ministers on a quest for savings. Truly dismaying was the revelation of dishonesty and collusion to prevent the Scheme's lack of legal foundation coming to light. Equally disheartening was the ineffectiveness of what one might consider institutional checks and

³³⁰ Peter Hanks, (above, n 323) p 15.

³³¹ Robodebt Report, (above n 324) p xxv.

³³² Ibid, p xxvi.

³³³ *Prygodicz v Commonwealth (No 2)* [2021] FCA 634, paras 4 and 5.

³³⁴ Ibid p xxix

balances – the Commonwealth Ombudsman’s Office, the Office of Legal Services Coordination, the Office of the Australian Information Commissioner and the Administrative Appeals Tribunal – in presenting any hindrance to the Scheme’s continuance.³³⁵

Robodebt represents a myriad of failures. Primarily it involved the design of a very basic form of AI – first-wave AI – that did not comply with the legislation it was required to apply. There was a lack of human oversight: the automated system independently assessed overpayments and sent debt notifications. The debts were applied to welfare recipients, people who were least able to challenge or question them. Equally concerning was that oversight agencies, such as the Ombudsman, failed to identify its basic legal flaws. These flaws were not remedied until a class action lawsuit was settled by the Commonwealth, some six years after the system was initiated. It took a Royal Commission to uncover the full extent of the errors and to expose the misconduct of the executive to attempt to avoid accountability.

Despite these problems, the illegalities of Robodebt were only uncovered and remedied, albeit after six years, because of administrative law rights and remedies and the fact that a proper analysis of the system’s departures from the applicable legal framework, could be undertaken once the facts were known. That would not have been possible in the case of machine learning-based AI that cannot be readily understood or scrutinised. Many examples of AI developed and deployed by the private sector are not subject to these safeguards.

The Targeted Compliance Framework

Despite the failures of the Robodebt programme and its controversial aftermath, the Commonwealth has continued to recover welfare payments illegally using automated, first-wave AI systems.

The Targeted Compliance Framework (TCF) was introduced in 2018 and designed ‘to make job seekers complete specific activities to receive income support from the Commonwealth Government.’³³⁶ It did so by requiring recipients to comply with mutual obligation conditions using an automated decision-making system that did not comply with the TCF’s legal requirements for discretionary decision-making. Following an investigation, the Ombudsman found that ‘cancellation of income support ... was not lawful;’³³⁷ that ‘[a]gencies did not

³³⁵ Robodebt Report(above, n 324), p iii.

³³⁶ Commonwealth Ombudsman, *Automation in the Targeted Compliance Framework: when the Law is Changed but the System Isn’t*, August 2025. See <https://www.ombudsman.gov.au/__data/assets/pdf_file/0017/320750/Automation-in-the-Targeted-Compliance-Framework.pdf>.

³³⁷ Ibid, p 5.

ensure their respective processes and computer systems were complying with the amended legislation;³³⁸ and that the Secretary of the Department responsible for the TCF ‘has not complied with the statutory requirement to establish a Digital Protections Framework.’³³⁹

Robodebt and the Targeted Compliance Framework: concluding observations

Existing administrative law mechanisms appear to be sufficiently robust to address issues where first-wave AI has been used by government unlawfully. However, remedies have not been produced in a timely manner. In the case of the Robodebt programme these delays were exacerbated by the failures of oversight bodies to undertake their functions effectively: institutional checks and balances broke down at multiple points. More broadly, failures within the executive to ensure that the initiatives were lawful point to cultural deficits within the public service and the political process:

[w]hether a public service can be developed with sufficient robustness to ensure that something of the like of the Robodebt scheme could not occur again will depend on the will of the government of the day, because culture is set from the top down.³⁴⁰

These failures suggest that the use of second wave AI by governments should not be undertaken without significant reforms within government. They also point to the need for enhanced levels of government accountability where it deploys that type of AI.

Recommendations to address these issues are set out in Chapter Six.

Discrimination

The data sets that are used to train AI systems are often incomplete, under-representative or, as explained in Chapter Two, consist of labelling errors. These shortcomings, for example, ‘[i]f particular groups are over or underrepresented in the training sets, including along racial and ethnic lines, algorithmic bias can result. Similarly, if the training sets include already biased data, they can produce biased outcomes.’³⁴¹ Historic biases:

[c]an affect the data themselves. A core element of machine learning is making predictions about the future on the basis of data from the past. However, if past data are biased against certain groups, including along racial and ethnic lines, the computer models can reproduce and amplify those biases. The use of biased or flawed data to inform real life decisions can further target and harm marginalized racial and ethnic groups because the use of those data in the context of artificial intelligence creates more data, which are then used to inform

³³⁸ Ibid.

³³⁹ Ibid.

³⁴⁰ *Report of the Royal Commission into the Robodebt Scheme* (above n 204), p iii.

³⁴¹ K.P. Aswini, *Report of the Special Rapporteur on Contemporary Forms of Racism, Racial Discrimination, Xenophobia and Related Intolerance*, UN Doc A/HRC/56/68 (3 June 2024), para 13.

future decisions. Such self-reinforcing systems can replicate and deepen existing disparities.³⁴²

Bias and discrimination go together.

Like data protection laws, Australia's legal protections against discrimination consist of a mixture of Commonwealth, state and territory laws. Although they cover similar subject matter, there are differences between them. It is unnecessary for the purposes of this thesis to undertake a comprehensive analysis of their scope, overlaps and varying methods of enforcement. Instead, the focus is on the Commonwealth legislation that applies across Australia, in particular, those laws that potentially affect AI.

There are four main laws that proscribe certain forms of discrimination. These are the *Racial Discrimination Act 1975 (Cth)*, the *Sex Discrimination Act 1984 (Cth)*, the *Disability Discrimination Act 1992 (Cth)*, and the *Age Discrimination Act 2004 (Cth)*. Added to these is the *Australian Human Rights Commission Act 1986 (Cth)*, which establishes the Commission and confers functions and powers on it, including jurisdictional oversight of the discrimination legislation. Under these pieces of legislation, discrimination against a person based on certain protected attributes (such as race, sex, disability and age) when engaging in public life is prohibited. Discrimination can occur directly or indirectly. Direct discrimination occurs where an individual is treated less favourably because of a protected attribute. Indirect discrimination occurs where a general rule or requirement disproportionately disadvantages a person with a protected attribute.³⁴³ The statutory complaints mechanism that enables individuals to seek a review of conduct that falls within the statutory proscriptions enables discriminatory conduct to be contested.

The potential for AI discrimination in the area of employment in Australia has recently been documented in research undertaken by Natalie Sheard that considers whether the 'use of automated reasoning to aid or replace human decision-making at the screening stage in recruitment [algorithmic hiring systems or AHSs] are contestable under federal anti-discrimination laws.'³⁴⁴

³⁴² Ibid, para 15.

³⁴³ Victorian Equal Opportunity and Human Rights Commission, 'Discrimination,' See <<https://www.humanrights.vic.gov.au/for-individuals/discrimination/>>.

³⁴⁴ Natalie Sheard, 'No Notice and No Explanation: The Incontestability of Hiring Discrimination by Algorithm,' (2022) 35 *Australian Journal of Labour Law* p 119.

Increasingly, AHSs are used during processes used to hire employees. They use AI-based automated decision-making to either replace or assist recruitment decisions that would otherwise be undertaken by a human.

An Australian example is Sapia.ai '[t]he AI hiring agent for unbelievably effective hiring'³⁴⁵ that 'keeps you in the drivers seat, streamlining your entire hiring process so you deliver hiring outcomes that are faster, fairer and backed by AI-driven insights.'³⁴⁶ Sapia lists its users as including Qantas, David Jones, Berry Street, Woolworths and Tennis Australia, among others. It deploys a 'competency framework' that is claimed to be a 'data-driven approach for the AI era'³⁴⁷ that identifies 'enduring job-agnostic competencies that reflect what drives success in the modern workplace – capabilities like adaptability, resilience, learning agility, and customer orientation.'³⁴⁸ The competency framework involves using a 'rich dataset of over 37,000 job descriptions'³⁴⁹ and uses 'large language models (LLMs) and advanced NLP techniques'³⁵⁰ from which were extracted 'over 200,000 behavioural descriptors'³⁵¹ distilled into a 'list of 25 human-centric competencies'³⁵² that are used to assess job applicants. Candidates are first asked to undertake a text-based chat interview. The answers are compared to the competency framework. A video interview follows for 'the top people'³⁵³ and then a recommendation is made about the suitability of the candidate.

To provide additional information, Sapia offers a range of resources, including blog posts, case studies and whitepapers, that cover topics such as 'Responsible use of Gen AI in hiring,' 'How AI technology helps you hire for equality'; and 'Building trust in ethical AI through transparency.'³⁵⁴ However, full access to these is limited to those who provide a work email, thus apparently limiting access to those who are currently employed. The reason for requiring this personal information is that '[w]e need to store and process your personal data to provide you with the requested content and send relevant follow-up communications.'³⁵⁵ It is highly questionable that there is any such 'need' and the requirement tends to undermine Sapia's claims to transparency. Noting that these additional

³⁴⁵ See <<https://sapia.ai>>.

³⁴⁶ Ibid.

³⁴⁷ See <<https://sapia.ai/resources/blog/reinventing-the-competency-framework-a-data-driven-approach-for-the-ai-era/>>.

³⁴⁸ Ibid.

³⁴⁹ Ibid.

³⁵⁰ Ibid. Presumably NLP refers to natural language processing.

³⁵¹ Ibid.

³⁵² Ibid.

³⁵³ See <<https://sapia.ai/products/video-interview/>>.

³⁵⁴ See <<https://sapia.ai/resources/>>.

³⁵⁵ Ibid. The author chose not to sign up to receive the resources.

resources may contain a discussion about the algorithmic basis of Sapia.ai's decision making, employment applicants are not provided with a clear description of how the system makes recommendations or decisions.

There are no laws in Australia that specifically address the risks of AHSs. In their absence, the impact of these systems is covered by existing laws, primarily discrimination laws and, to a lesser extent, data protection laws.

In order to bring a discrimination case under Commonwealth discrimination law, a complainant must first lodge a written complaint with the Australian Human Rights Commission (AHRC). There is a standard complaint form³⁵⁶ that requires the complainant to specify *who* the complaint is about,³⁵⁷ *what* the complaint is about,³⁵⁸ *when* the discrimination occurred³⁵⁹ and *what* happened.³⁶⁰ Once a complaint is established, a range of remedies are available.

In order to contest AHS discrimination, the complainant bears the onus of proof. It is difficult to see how that onus could be discharged where an AHS is used, or more generally, for other instances of AI-enabled employment decisions. '[T]here is no legislative requirement that a job applicant be provided with notice or an explanation of the operation of an AHS in the recruitment or hiring process. It will therefore be inherently difficult for a complainant to assess whether wrongdoing has occurred and marshal evidence to prove discriminatory conduct.'³⁶¹

AI-based employment screening is an example of the incontestability of AI that can have profound impacts on the ability of an individual to obtain employment. In particular, its use presents as an example where AI-based processes are less accountable, and amenable, to discrimination law than if screening decisions were undertaken by a human, raising the prospect that it evades the rule of law in ways that human decision-making cannot.

Conclusion

Foundational input AI legal risks enable many downstream harms. Foundational output legal risks tend to enable second wave AI to escape liability for the harms that arise from their

³⁵⁶ See <<https://humanrights.gov.au/sites/default/files/2024-10/AHRC%20Complaint%20form%2020241024.pdf>>.

³⁵⁷ See Part B of the Complaint Form.

³⁵⁸ See Part C of the Complaint Form.

³⁵⁹ Ibid.

³⁶⁰ Ibid.

³⁶¹ Natalie Sheard, 'Employment Discrimination by Algorithm: Can Anyone be held Accountable?' (2022) 45(2) *UNSW Law Journal* 617, 646.

use. Although administrative law remedies appear to be sufficiently robust to address harms arising out of government use of first-wave AI, in practice effective remedies are not available in a timely manner and, to date in Australia, the misapplication of that form of AI by government has primarily been directed at uncovering welfare fraud. It has been directed at those who are disadvantaged and who are least likely to have access to the resources to contest government automated decision-making. Cultural issues within government and the repetition of the errors of the Robodebt programme in the TCF programme do not signify that lessons have been learnt:

Anti-welfare rhetoric is easy populism, useful for campaign purposes. It is not recent, nor is it confined to one side of politics, as some of the quoted material in this report demonstrates. It may be that the evidence in this Royal Commission has gone some way to changing public perceptions. But largely, those attitudes are set by politicians, who need to abandon for good (in every sense) the narrative of taxpayer versus welfare recipient.³⁶²

Administrative law is likely, however, to struggle to address the use of second wave types of AI by government. The problems associated with government implementing first-wave AI do not inspire confidence in its ability to deploy second wave types of AI in the public interest.

Although many jurisdictions have developed or are developing ways in which to address AI-related legal issues their efforts are hampered by a range of factors that inhibit normal public policy approaches, most prominently the fear that regulating AI will impede innovation and diminish the perceived economic and social benefits associated with it. The background to, and an explanation of, these impediments is discussed in Chapter Four.

³⁶² Robodebt Report (above n 324) p iii

Chapter Four- AI regulation and the issues it raises

Introduction

There are a variety of approaches to AI policy and regulation, but each shares a common feature that governments, as the drivers of public policy and regulation, seek to address: ‘to introduce policies and regulatory arrangements that effectively mitigate [the] risks of AI while fostering its vast potential economic and social benefits.’³⁶³ A consistent theme that emerges from the literature and public debate is how to strike an appropriate balance between the two.

Despite agreement about these key objectives, public policy approaches to the design of a regulatory architecture differ significantly. As a prelude to the discussion of the main regulatory approaches that have been developed in Chapter five, this Chapter provides an account of the public policy issues that underpin and inform the different regulatory architectures that have emerged and a critical analysis of key regulatory approaches.

AI regulation as a public policy issue

Public policy is a product of the political process and is overwhelmingly the province of government. It is ‘a course of action by government designed to attain certain results’³⁶⁴ that is intended to produce ‘an authoritative response to a public issue or problem.’³⁶⁵

That response is frequently designed to reduce some form of perceived threat or harm. As noted recently by Frieberg:

Most obviously, governments regulate to prevent harms to individuals, to property, to the environment and to the state itself. Much regulation is the result of some harm occurring, or being recognised, followed by a call to action and a government response.³⁶⁶

In order to deal with this issue government selects one or more policy instruments.

There are five main types of public policy instruments that are used by all Australian governments. These consist of public policy through advocacy; networking, i.e., using relationships to influence behaviours and outcomes; funding, such as grants, financial

³⁶³ Australian Senate, Select Committee on Adopting Artificial Intelligence, *Select Committee on Adopting Artificial Intelligence Final Report*, November 2024, p 44. See <[https://parlinfo.aph.gov.au/parlInfo/download/committees/reportsen/RB000470/toc_pdf/SelectCommitteeonAdoptingArtificialIntelligence\(AI\).pdf](https://parlinfo.aph.gov.au/parlInfo/download/committees/reportsen/RB000470/toc_pdf/SelectCommitteeonAdoptingArtificialIntelligence(AI).pdf)>.

³⁶⁴ Catherine Althus, Peter Bridgeman and Glyn Davis, (above, n 60) p 6.

³⁶⁵ Ibid p 8.

³⁶⁶ Arie Freiberg, (above n 9) p 46.

incentives and concessions; direct intervention, such as delivering programmes and services; and through the law, using legislation and regulatory instruments to coerce, require or prohibit certain behaviours. Of each of these possible instruments, policy implementation through the law, i.e., regulation, is ‘the traditional instrument of government policy, and the final guarantee that policy intent can be translated into action.’³⁶⁷ The choice of policy instruments matters. ‘It is the link between an objective and its attainment.’³⁶⁸

The choice of regulatory instrument involves an assessment that considers its suitability and fitness for purpose. Generally, the criteria used to guide this process are:

- 1) Appropriateness – is the proposed initiative a reasonable way of proceeding in this particular area of policy?
- 2) Efficiency – will the initiative be cost-effective?
- 3) Equity – are the likely consequences fair?
- 4) Suitability – will there be a conflict with existing policies and initiatives?
- 5) Workability – is the initiative coherent and implementable?³⁶⁹

These considerations are relevant to all public policy initiatives, and thus are important in assessing how to go about regulating AI.

AI, innovation and disruption

In a public policy setting, ‘[i]nnovation’ is generally defined as the development of improved products, services and processes, the creation of new markets, and the use of new products.’³⁷⁰ It is commonly considered to be a positive development overall that drives economic growth, prosperity and social wellbeing. It can take various forms, ranging from gradual and evolutionary change to radical disruption that challenges existing regulatory and governance frameworks. AI can be considered to exhibit elements of each. Its early development was gradual and, periodically, faltering. It became disruptive over the last decade when the theoretical technologies of machine learning could be put into practice. Its explosion since 2022 has highlighted a range of regulatory risks:

³⁶⁷ Ibid, p 94.

³⁶⁸ Catherine Althus, Peter Bridgeman and Glyn Davis, (above, n 60) p 94.

³⁶⁹ Ibid, pp 95, 96.

³⁷⁰ New South Wales, Treasury, *NSW Government Guide to Better Regulation*, January 2019, p 12. See <https://www.productivity.nsw.gov.au/sites/default/files/2022-05/TPP19-01_Guide-to-Better-Regulation.pdf>.

Innovation creates uncertainty about the speed of innovation and its adoption, its effect on current regulatory practices, how it may be used or abused, and the community might react both to the innovation and any regulatory responses. However, uncertainty is antithetical to the concept of the rule of law, which requires predictability and stability.³⁷¹

One of the key difficulties with regulation in this environment is that technology develops much faster than regulation, leaving policy makers, legislators and regulators in a continual cycle of playing catch up. The speed of AI technological change also exacerbates problems for policy makers and regulators who may lack sufficient information to understand the nature of the disruption and thus how to gauge the nature and extent of its risks, understand community concerns and expectations, and to either modify existing regulation or develop new regulatory responses. In the case of technology regulation, and AI regulation in particular, these dilemmas are compounded by a range of additional factors.

Regulation and wicked problems

However skilfully policy instruments are chosen and evaluated they have inherent limitations, particularly in a disruptive environment. There are some intractable public policy problems, known as ‘wicked problems’ for which there is no optimal policy solution. Wicked problems ‘refer to those dilemmas that either cannot be defined or, at best are not open to easy formulation.’³⁷² In addition, ‘[s]ome issues are simply not open to solution however well-structured the problem. Historical factors, competing interests or sunk costs can make all sides to a dispute unwilling to compromise.’³⁷³

Other factors also contribute to the ‘wickedness’ of a public policy problem. These include:

- structural complexity – where the technical dimensions of a problem are intractable;
- its lack of knowability – where there is a lack of sufficient knowledge to fully understand the problem;
- knowledge fragmentation – where the availability of knowledge about a problem is fragmented among multiple stakeholders each of whom may have some but not all of the information needed to address it;
- knowledge framing – where some knowledge about a problem receives too much or too little attention because of the way it is framed;

³⁷¹ Arie Freiberg (above, n 9) p 638.

³⁷² Ibid, p 54.

³⁷³ Catherine Althus, Peter Bridgeman and Glyn Davis (above, n 60) p 53.

- interest differentiation – where interested stakeholders having interests or values that are substantially in conflict; and
- power imbalances – where there is a dysfunctional distribution of power among stakeholders and powerful actors overwhelm the less powerful.³⁷⁴

Using this framework, AI presents as a wicked problem. Its technologies are complex and knowledge about their functions and operations are concentrated in the organisations that develop them and which are often hostile to regulatory interventions. Many are powerful and well-resourced international enterprises whose key objective is to commercialise their technologies without restriction. Frequently, the harm that AI can cause is fully appreciated only well after it has become available and widely used. AI is not homogenous; it takes a variety of forms each of which gives rise to similar, but not necessarily the same, risks. It also consists of a variety of components each of which requires careful analysis to understand their individual contribution to risk and how to mitigate them. Many AI enterprises are, or are owned by, some of the wealthiest and well-resourced global companies. Their ability to influence the public policy process at national and international levels enables them to overpower the voices of other stakeholders.

The lengthy gestation of AI: how it impedes AI regulation

Further issues arise due to the lengthy gestation of AI and its evolution from weak forms that were less powerful and more easily comprehensible, to modern forms based on machine learning and neural networks. The lack of specific regulation in the past has led to expectations on the part of those who profit from and use it that this state of affairs would continue and to confusion as to why change may be required after all this time.

So called ‘weak AI’ is based on predefined algorithms such as the one used in the Robodebt programme as discussed at in Chapter Three. The fact that the algorithm is predefined means that it is more comprehensible, therefore raising less complex policy and regulatory issues. Modern AI has developed in the context of Big Data, where it is now possible to accumulate large quantities of data, including personal data, at low cost, and new analytics have been developed to facilitate their harvesting. The latter include the machine learning based algorithms that not only process but also learn from the vast troves of personal and

³⁷⁴ John Alford and Brian W. Head, ‘Wicked and Less Wicked Problems: A Typology and a Contingency Framework’ (2017) 36(2) *Policy and Society*, 397.

other data available for processing. The fact that these algorithms change as they learn makes them inherently more opaque, giving rise to the black box problem.

The regulatory challenges posed by weak AI were such that these could be absorbed comparatively readily into and overseen by traditional forms of regulation such as the law of torts, contracts and product liability and through the elasticity of the common law. While some of the later pre-machine learning AI have raised issues of lack of transparency (where it was based on algorithms that are complex and not readily capable of simple explanation) and potential issues of unfairness and due process, as illustrated in the Robodebt case study, they pose less challenges to issues of intent and causation that are fundamental to traditional models of fault-based liability. Likewise, they do not engage the foundational risks such as data protection and copyright issues discussed in Chapter Three. Where 'new' risks did manifest themselves in particular domains, these could generally be addressed within existing specific regulatory regimes, sometimes without any regulatory changes or with minimal amendment (for example, in the case of medical devices).³⁷⁵ Regulatory sufficiency could overall be achieved through the applying 'traditional subjects such as tort law, contract law, property law and so forth.'³⁷⁶

Regulation and the broader tech sector

Many of the difficulties experienced in attempting to regulate AI are a residue of long unresolved debates about how the internet should be regulated:

The experience of internet regulation from 1995 to today serves as a warning which is applicable to all emerging technologies. By failing to take early steps to structurally regulate the internet and instead focusing on individual harms, governments failed to appreciate that they had left markets to control a communications technology which relies on network effects and can create system wide risks and impacts.³⁷⁷

The technology sector has occupied a privileged position when it comes to regulation. The default position for many governments, particularly the US which has, until recently, been the jurisdiction that leads technology and AI development, has been to resist regulation in favour of a 'light touch' approach for fear of stifling innovation and to allow the market to

³⁷⁵ Medical devices are regulated under the *Therapeutic Goods Act 1989 (Cth)*. Regulations designed to address risks associated with AI enabled or embodied medical devices, the *Therapeutic Goods (Medical Devices) Regulations 2002* were amended commencing in 2021 to clarify regulatory obligations for software-enabled medical devices.

³⁷⁶ Lyria Bennett Moses, *How to Think about Law, Regulation and Technology: Problems with 'Technology' as a Regulatory Target*, (2013) 5(1) *Law, Innovation and Technology* 1, 15.

³⁷⁷ Julia Black and Andrew Murray, 'Regulating AI and Machine Learning: Setting the Regulatory Agenda' (2019) 10 (3) *European Journal of Law and Technology* 1.

regulate itself: overarching regulatory strategies were avoided. Any regulation involved a patchwork of individual laws targeting specific, rather than systemic, harms.

Governments seemed paralysed, with no government (at least no Western government) wanting to be the first to be seen to be regulating this innovative, creative space where freedom seemed to create both economic and civic benefits. Governments interacted with this space in a piecemeal fashion: some laws on copyright here, some on hate and harmful speech there, but no coherent strategy for regulation of the space emerged.³⁷⁸

Market-led regulation

The prevailing ‘hands off’ regulatory orthodoxy for technology dates to the dawn of the internet when Vice President Al Gore gave his Information Superhighways Speech in 1994.³⁷⁹ Gore articulated a plan based on five principles. Principle three addressed internet regulation:

[W]e aspire to build our information highways according to a set of principles ... Our plan is based on five principles: ... Third, create a flexible regulatory framework that can keep pace with rapid technological and market changes ...³⁸⁰

In elaborating on principle three he stated that:

Hand in hand with the need for private investment and competition is *the necessity of appropriate and flexible regulations developed by an authoritative regulatory body*. In order for the private sector to invest and for initiatives opening a market for competition to be successful, it is necessary to create a regulatory environment that fosters and protects competition and private sector investments, *while at the same time protecting consumers’ interests*.³⁸¹

An authoritative regulatory body for the internet did not emerge. Instead, a range of internet-related functions were added to the Federal Trade Commission’s portfolio and several non-government bodies were established in the US to coordinate standards, domain names and engineering issues.³⁸² There was no overarching regulation to protect consumer or other stakeholder interests.

In 1994 it was difficult to foresee many of the harms that the ‘information superhighway’ would cause, particularly what the rise of AI would mean in the context of a globalised and connected market for technology, data and information. Gore’s speech set the scene for what became known as ‘light touch’ regulation as a default regulatory posture. Australia followed this approach, which was identified as being able to ‘give us an innovative and

³⁷⁸ Ibid, p 6.

³⁷⁹ Vice President Al Gore, *Information Superhighways Speech*, International Telecommunications Conference, 21 March 1994. See <<http://vlib.iue.it/history/internet/algore speech.html>>. The speech is often referred to as the ‘Al Gore Invents the Internet’ speech.

³⁸⁰ Ibid.

³⁸¹ Ibid. (My emphasis.)

³⁸² These included the World Wide Web Consortium (W3C), the Internet Corporation for Assigned Names and Numbers (ICANN), and the Internet Engineering Taskforce (IETF).

competitive future ... [a]n important feature (of which) will be the so-called “information superhighways” which transform all our lives over the coming years.’³⁸³

Neoliberalism and de-regulation

The 1980s and 1990s were a period in which ‘de-regulation’ became the norm. Regulation was characterised as an impediment to economic growth and as imposing unnecessary, inefficient and costly burdens that did not confer corresponding economic or societal benefits. De-regulation was considered to increase competition in markets by removing barriers to market entry and to reduce compliance burdens, thus promoting innovation and market flexibility:

Driven by Chicago School economics, regulation was framed as economically harmful and blocking efficient resource allocation, demonstrated by its costs to business and the community, its effect on investment, employment and innovation, decreased productivity and declining government revenue ... In the 1980s the dominant theme was that of ‘deregulation’ ...³⁸⁴

Accordingly, regulation of sectors such as finance and banking, transport, telecommunications and resources, as well licensing requirements for commercial activities and for professional services were relaxed. In many cases government agencies that owned and operated enterprises in these sectors were sold into the private sector. This policy stance became known as ‘neo-liberalism.’

Cyber exceptionalism

In parallel many internet advocates made a series of ideological claims that challenged the legitimacy of *any* internet regulation at all. One of the most influential of the so-called cyberlibertarians was John Perry Barlow whose Declaration of the Independence of Cyberspace³⁸⁵ at the World Economic Forum in 1996 stated that:

Governments of the Industrial World ... I come from Cyberspace, the new home of Mind. On behalf of the future, I ask you of the past to leave us alone. You are not welcome among us. You have no sovereignty where we gather ... we are building to be naturally independent of the tyrannies you seek to impose on us. You have no moral right to rule us nor do you possess any methods of enforcement we have true reason to fear.³⁸⁶

³⁸³ Paul Keating, ‘Statement by the Prime Minister, Hon. P.J. Keating, A National Strategy for Information and Communications Services and Technologies’, Media Release, 6 April 1994. See <<https://pmtranscripts.pmc.gov.au/release/transcript-9541>>.

³⁸⁴ Arie Freiberg, (above, n 9) p 70

³⁸⁵ John Perry Barlow, ‘A Declaration of the Independence of Cyberspace’, Davos, 8 February 1996. See <<https://www.eff.org/cyberspace-independence>>.

³⁸⁶ Ibid.

The legal foundations for this approach denied the legitimacy of national governments to exercise sovereignty over a new, claimed *supra* national cyber territory. These arguments contended that state jurisdiction could only legitimately be exercised over persons within a physical territory or jurisdiction; that regulatory legitimacy is derived from the consent of those governed and the citizens of cyberspace had not consented; and that no state could legitimately exercise control over the claims of another state in cyberspace.³⁸⁷

The problematic nature of these arguments is that they misconceive the internet as existing separately and disconnected from its operation and effect within each jurisdiction: any state is entitled to regulate conduct in its own territory. Although ‘in truth little of the cyberlibertarian argument was true,’³⁸⁸ it continues to reverberate.

In an environment where de-regulatory, neo-liberal and cyberlibertarian arguments coalesced at about the same time as the internet became publicly available, the trend of opinion was that regulatory interventions in it were inappropriate and unjustifiable. The counterpart of these arguments in the context of AI regulation — and in particular, the AI debate in Australia — is that innovation and regulation are seen as binary choices.

The innovation v regulation debate

One of the strongest and most pervasive arguments that has been deployed to avoid or limit AI regulation is that regulation stifles innovation. The Productivity Commission warns ‘of regulations having an unintended chilling effect on innovation’³⁸⁹ in support of an argument that ‘calls for new AI-specific regulations are largely misguided.’³⁹⁰

There is a widely held view that ‘regulation of the digital economy compromises innovation and undermines technological progress. Regulation is ... a burden that diverts resources away from firms’ innovative activities.’³⁹¹ The primary way in which this argument is advanced is to compare the success of US-based technology companies, which have developed, grown and in many cases dominated international digital innovation, against the lack of success of European-based technology companies.

³⁸⁷ David R. Johnson and David Post, *Law and Borders—The Rise of Law in Cyberspace* (1996) 48(5) *Stanford Law Review* 1367.

³⁸⁸ Black and Murray (above, n 377) p 6.

³⁸⁹ Productivity Commission, *Harnessing Data and Technology: Interim Report*, August 2025. See <<https://www.pc.gov.au/inquiries/current/data-digital/interim/data-digital-interim.pdf>>

³⁹⁰ Productivity Commission, Stephen King, ‘The best way to regulate AI might be not to specifically regulate AI. This is why.’ 20 September 2024. See <<https://www.pc.gov.au/media-speeches/articles/best-way-to-regulate-ai>>.

³⁹¹ Anu Bradford, *The False Choice between Digital Regulation and Innovation* (2024) 19 *NW. U. L. Rev* 377, 378.

Until recently, most governments have refrained from regulating the tech industry precisely because of their fear that attempts to interfere with tech companies' operations would undermine their innovative capacity. The United States has led this charge against regulation, insisting on the primacy of free markets, free speech, and free internet as foundations of the digital economy. The American regulatory approach reflects a view that the country's technology leadership derives from an unregulated marketplace. The US regulatory regime, with its deeply embedded techno-libertarian ethos, consists of weakly enforced antitrust laws, the absence of a federal data privacy law, and permissive content-moderation rules that shield tech companies from liability.³⁹²

Although there are a variety of references that can be used to support this analysis, a salient example is the submission made by OpenAI, the developers of ChatGPT, to the US consultation process on the development of a US AI Action Plan, established under section 4 of the Presidential Executive Order No 14179 of 23 January 2025, *Removing Barriers to American Leadership in Artificial Intelligence*.³⁹³

In its submission 'Preemption: Ensuring the Freedom to Innovate,' OpenAI states that:

[w]e propose creating a tightly-scoped framework for *voluntary partnership* between the federal government and the private sector *to protect and strengthen American national security*. This framework would extend the tradition of government receiving learnings and access, where appropriate, in exchange for providing the private sector relief from the 781 and counting proposed AI-related bills already introduced this year in US states. This patchwork of regulations *risks bogging down innovation* and, in the case of AI, *undermining America's leadership position*.³⁹⁴

OpenAI's arguments warrant close reading. They reiterate a number of familiar, anti-regulation sub-arguments that often accompany the innovation debate, particularly in the US. The first is the desirability of *voluntary* participation in a public/private sector collaboration. The second is the linkage made between that voluntary collaboration and strengthening America's national security. The third is the highly contingent and limited nature of the trade-offs that are suggested, being the provision of 'learnings' to government *where appropriate* in exchange for immunity from proposed AI regulations being developed by the US states.³⁹⁵ The fourth is that state-based regulation is a risk to innovation and will undermine America's position as the world leader in the development of AI. Although there is probably merit in avoiding a complex patchwork of state-by-state AI regulation in the US, it

³⁹² Ibid, pp 379-380.

³⁹³ See <<https://www.federalregister.gov/documents/2025/01/31/2025-02172/removing-barriers-to-american-leadership-in-artificial-intelligence>>.

³⁹⁴ See <<https://cdn.openai.com/global-affairs/ostp-rfi/ec680b75-d539-4653-b297-8bcf6e5f7686/openai-response-ostp-nsf-rfi-notice-request-for-information-on-the-development-of-an-artificial-intelligence-ai-action-plan.pdf>>. The italicised emphasis in this quotation is mine. The underlined emphasis is OpenAI's.

³⁹⁵ OpenAI does not provide any detail about these claims.

is by no means clear that the other arguments offered have significant merit or are supported by an evidence base that relates to innovation.

A problematic dimension of the ‘regulation stifles innovation debate’ is that it presents a complex, multi-faceted, wicked public policy problem in an overly simplistic form that privileges innovation over any other issue. As analysis of the foundational risks posed by AI demonstrates, AI risks are significant and structural. AI can produce harm at scale that is difficult to identify and link to precise causes.

Innovation and *ex ante* regulation

Regulation of AI can take the form of either *ex ante* or *ex post* regulation. *Ex ante* regulation seeks to limit the risks of AI before they occur, such as through requiring AI providers to undertake risk assessments and to comply with measures that enhance the transparency of AI models *before* they are placed on the market. They are often referred to as precautionary measures.³⁹⁶ These are also resisted on the basis that they adversely affect innovation, the key arguments being that *ex ante* regulation, as a precautionary principle:

- 1) slows and makes the development of AI more expensive. ‘For example, if all fifty US states had laws such as New York’s, which requires autonomous vehicle firms to perform road testing under the paid supervision of police, testing such vehicles would be more expensive;’³⁹⁷
- 2) stifles innovation;
- 3) leads to lower quality AI if the regulation requires explainability. ‘Developing an AI system capable of explaining itself or justifying its decisions is an incredibly challenging technical feat, so much so that the U.S. Defense Advanced Research Projects Agency (DARPA) devoted \$75 million in 2017 to research how to achieve it ... it is legitimate for companies such as IBM, to create internal requirements for AI explainability. Requiring all firms to meet such a standard, however, would create a barrier to adopting AI, because not all AI systems are alike and not all businesses have a similar level of expertise;’³⁹⁸ and

³⁹⁶ Daniel Castro and Michael McLaughlin, *Ten Ways the Precautionary Principle Undermines Progress in Artificial Intelligence*, Information Technology and Innovation Foundation, 4 February 2019. See <<https://itif.org/publications/2019/02/04/ten-ways-precautionary-principle-undermines-progress-artificial-intelligence/>>.

³⁹⁷ Ibid, p 13.

³⁹⁸ Ibid, p 14.

4) leads to less AI adoption.

Once again, these arguments should not be accepted at face value.

EU Regulation and the Innovation Debate

The EU AI Act is frequently cited as having stifled AI innovation in the EU but it did not commence operation until 1 August 2024 and does not fully come into effect until mid-2026. Although the EU AI Act had been under development for several years before its enactment, it is nevertheless puzzling that such a recently developed and partially implemented regulatory framework is claimed to have had such an effect.

The relationship between regulation and innovation in the EU is significantly more complex. There are important differences between the EU and the US that explain more comprehensively why the development of technology, and AI in particular, in the EU lags behind the US:

A closer examination of the US-EU differences suggests that the EU's inability to cultivate an equally successful tech industry can be traced to various other factors. These include (1) the fragmented digital single market that limits the scaling of innovations within the EU, (2) underdeveloped capital markets that limit tech companies' ability to grow in the EU, (3) Europe's punitive bankruptcy laws and cultural attitudes that deter risk-taking, and (4) the absence of a proactive immigration that would allow the EU to harness foreign tech talent.³⁹⁹

Chinese AI Regulation and the AI Debate

The argument also overlooks the fact that AI has developed significantly in a regulated market – China. The Chinese regulatory model is discussed in greater detail in Chapter Five, but for present purposes it is sufficient to note that Chinese AI regulation contains detailed rules that are designed to limit AI risks. These are expressly designed to apply the rule of law to AI and include requirements to register algorithms and disclose training data as well as rules that require compliance with data protection and copyright laws. Chinese AI regulation has been iterative and has foreshadowed the shape of regulatory initiatives before they come into effect thus providing market participants with the trajectory of AI regulation in advance. Western debates over AI regulation have largely overlooked Chinese developments, focusing instead on choosing between US and EU approaches, and there is no reference to them in Australian government policy papers on AI regulation. However, the Chinese experience provides contemporary and compelling evidence that regulation and

³⁹⁹ Bradford (above n 391) p 382.

innovation are not binary choices. Moreover, it shows that proponents of innovation over regulation are engaged in a form of ‘special pleading’, i.e., ignoring successful AI innovation in a large, carefully regulated market.

The Rationale for the regulation stifles innovation argument: humans make mistakes too

Another argument deployed against AI regulation that is relatively novel and specific to AI is that human decision-making is also flawed, subject to failure and non-transparent. Decision-making, particularly at a governmental level:

is already driven by algorithms of arguably greater complexity and potential for abuse: the algorithms implicit in human decision-making. The human brain operates algorithmically through complex neural networks. And when humans make collective decisions, they operate via algorithms too – those reflected in legislative, judicial, and administrative processes. Yet these human algorithms undeniably fail and are far from transparent. On an individual level, human decision-making suffers from memory limitations, fatigue, cognitive biases, and racial prejudices, among other problems.⁴⁰⁰

In conclusion, if the problems associated with AI decision-making are shared by human decision-making, why then is it necessary to have additional regulation of AI?

The problematic nature of this argument is that AI decision-making and predictions occur *at scale*, in other words at a rate and intensity that vastly exceeds that of human decision-making. While a person may make erroneous decisions, they will not replicate the error thousands or millions of times. The reach and implications of human decision-making is bounded and confined. There are also well-established review mechanisms, including administrative law requirements that apply to public sector decision-making, that serve to provide avenues for review and correction of decisions made in error, just as there are judicial processes and remedies that apply to regulate commercial decisions or advice that errs. Although the principles that govern these review processes can be transposed into the digital domain, their application is limited by the fact that black box AI cannot be adequately scrutinised. But even when simple, non-machine learning automated decision-making such as was used in the Robodebt programme creates major issues of public concern, there are significant impediments to achieving redress for those adversely affected.

Learning from Experience

As Julia Black and Andrew Murray argue, the experience of internet regulation influences all emerging technologies, including current debates about AI regulation:

⁴⁰⁰ Cary Coglianese and Alicia Lai, *Algorithm vs Algorithm* (2022) 72 Duke Law Journal 181.

By failing to take the early steps to structurally regulate the internet and instead focusing on individual harms, governments failed to appreciate that they had left markets to control a communications technology which relies on network effects and can create system wide risks and impacts ... Today parallels may be seen between the internet regulation debate of the 1990s and the debate surrounding the regulation of AI and machine learning.⁴⁰¹

Although some technology-related regulatory reform has emerged, most of it has evolved in a piecemeal way to address individual risks, like copyright, some data protection laws and hate speech.

Against this background is a lack of focus in the regulatory debate about AI on key questions about what, exactly, are the goals and objectives that AI regulation should seek to achieve and a calibrated discussion of systemic AI risks, such as black box AI, as well as individual risks. In particular, there is a puzzling lack of discussion in Australia about the *liability* of those who develop and deploy AI for the damage it causes.

AI Ethics

“AI Ethics” is another term that is repeatedly used in relation to AI regulation. It has been one of the most dominant responses proposed by the technology industry to address concerns about AI risks. There are countless examples⁴⁰² and, although there are some similarities, there are also differences.

The Commonwealth published Australia’s AI Ethics Principles⁴⁰³ in 2019 as one of its first responses to AI.⁴⁰⁴ They ‘are designed to ensure AI is safe, secure and reliable’ and to help ‘achieve safer, more reliable and fairer outcomes for all Australians; reduce the risk of negative impact on those affected by AI applications; and businesses and governments to practice the highest ethical standards when designing, developing and implementing AI.’⁴⁰⁵ The ‘principles are entirely voluntary’⁴⁰⁶ and are designed to ‘be aspirational and complement, not substitute, existing AI regulations and practices.’⁴⁰⁷

⁴⁰¹ Black and Murray, (above n 377) p 6.

⁴⁰² Nicholas Kluge Correa, Camila Galvao, James William Santos, Luiza Galvao, Edmund Terem and Nythamar de Oliveira, *Worldwide AI Ethics: A Review of 200 Guidelines and Recommendations for AI Governance*, (2023) Patterns 4, 13 October 2023. See <[Worldwide AI ethics: A review of 200 guidelines and recommendations for AI governance - PMC](#)>.

⁴⁰³ Australian Government, Department of Industry, Science and Resources, *Australia’s AI Ethics Principles*. See <<https://www.industry.gov.au/publications/australias-artificial-intelligence-ethics-principles/australias-ai-ethics-principles>>.

⁴⁰⁴ The principles were subsequently updated in October 2024.

⁴⁰⁵ Australian Government, Department of Industry, Science and Resources, *Australia’s AI Ethics Principles*, (above n 403).

⁴⁰⁶ *Ibid.*

⁴⁰⁷ *Ibid.*

There are eight principles:

- 1) Human, societal and environmental wellbeing: AI systems should benefit individuals, society and the environment.
- 2) Human-centred values: AI systems should respect human rights, diversity, and the autonomy of individuals.
- 3) Fairness: AI systems should be inclusive and accessible, and should not involve or result in unfair discrimination against individuals, communities or groups.
- 4) Privacy protection and security: AI systems should respect and uphold privacy rights and data protection, and ensure the security of data.
- 5) Reliability and safety: AI systems should reliably operate in accordance with their intended purpose.
- 6) Transparency and explainability: There should be transparency and responsible disclosure so people can understand when they are being significantly impacted by AI, and can find out when an AI system is engaging with them.
- 7) Contestability: When an AI system significantly impacts a person, community, group or environment, there should be a timely process to allow people to challenge the use or outcomes of the AI system.
- 8) Accountability: People responsible for the different phases of the AI system lifecycle should be identifiable and accountable for the outcomes of the AI systems, and human oversight of AI systems should be enabled.⁴⁰⁸

Although no one would disagree with these objectives, they are nevertheless:

vague, high-level principles and value statements that promise to be action-guiding, but in practice provide few specific recommendations and fail to address fundamental normative and political tensions embedded in key concepts (for example, fairness, privacy).⁴⁰⁹

AI ethical principles are abstractions that are too uncertain to provide an adequate framework in which to anchor AI regulation. At best they function as statements of intent that should guide more calibrated public policy thinking.

⁴⁰⁸ Ibid.

⁴⁰⁹ Brent Mittelstadt, *Principles Alone Cannot Guarantee Ethical AI* (2019) 1 *Nature Machine Learning* 501.

Aside from their abstract nature, there are no consequences for breaching AI Ethics principles. There are no enforcement mechanisms and no penalties for breaching them. Not only are they abstractions, they are toothless abstractions that are promoted to undertake a task for which they are not suitable or appropriate.⁴¹⁰ At worst, they enable AI developers to commit to ethics as a means of avoiding meaningful regulation and to provide governments and regulators with ‘a reason not to pursue new regulation.’⁴¹¹ They enable the technology industry to disingenuously engage in a process of virtue-signalling and ethics washing whereby claims are made that an AI system is developed and deployed according to ethical AI principles that are later abandoned without consequence, for example, Google’s unilateral abandonment of its commitment to ‘do no evil’ in 2018 and, another of its AI principles, not to use AI for potentially harmful applications such as weapons development and surveillance.’⁴¹²

Even when the shortcomings of AI Ethics are acknowledged, and supplementary activities are undertaken that are designed to operationalise them:

The translation from complex social concepts to technical rulesets is non-trivial. In a zero-sum world, the dominant turn to AI principles is not just fruitless but a dangerous distraction, diverting immense financial and human resources away from potentially more effective activity.⁴¹³

AI governance

Another confusing aspect of AI regulatory discourse is an emphasis upon ‘AI governance.’ Often appearing in conjunction with AI Ethics principles as a means to operationalise them, AI governance is frequently proposed as a solution to addressing and controlling AI risks and is referenced as an element of Australia’s regulatory response to AI: ‘Australia has a strong interest in working with international partners to mitigate the risks of AI through global governance mechanisms. A focus is the risks from frontier AI that is not being developed in Australia.’⁴¹⁴

Just like AI, the term ‘governance’ has proven to be notoriously difficult to define.

‘Governance’ is usually taken to refer to the processes and procedures that are employed by

⁴¹⁰ Luke Munn, ‘The Uselessness of AI Ethics’ (2023) 3 *AI and Ethics* 869.

⁴¹¹ Mittelstadt (above n 409) p 501.

⁴¹² The prior commitment can be found at the *Google AI Principles*, p 4. See <https://ai.googlestatic.com/documents/EN-AI-Principles.pdf>. It was abandoned in 2024. See <https://blog.google/technology/ai/responsible-ai-2024-report-ongoing-work/>.

⁴¹³ Munn (above n 410) p 869.

⁴¹⁴ Australian Government, Department of Industry, Science and Resources, *Safe and Responsible AI in Australia Consultation: Australian Government’s interim response*, 2024, p 23. See https://storage.googleapis.com/converlens-au-industry/industry/pri2452c8e24d7a400c72429/public_assets/safe-and-responsible-ai-in-australia-governments-interim-response.pdf.

organisations to establish strategy and to manage risk and compliance. But process and procedure should not be conflated with the regulation with which compliance is required. This important distinction is frequently blurred in AI governance discourse.

Most frequently used by reference to ‘corporate governance,’ governance has become an important component of private and public sector management practice. As one of the standard texts on corporate governance states, ‘[c]orporate governance is about the exercise of power over corporate entities.’⁴¹⁵ Another explanation is that:

[i]n general terms, corporate governance encompasses the arrangements by which the power of those who implement the strategy and direction of an organisation is both delegated and limited to ensure the organisation’s success, taking into account the environment in which the organisation is operating ... it is not surprising that there is no universally agreed definition of corporate governance, just as there is no universally accepted structures and practices that constitute good governance.⁴¹⁶

AI governance in turn is described as:

[t]he system by which an organisation manages its development and use of AI systems, and includes a wide range of governance structures, policies, skills, practices, and other mechanisms that guide AI use, monitoring, and management ... [It is] an emerging field that explores ways that organisations can evolve existing corporate governance approaches and introduce new strategies to realise the benefits of AI systems while addressing the risk of AI harms to the organisation, individuals, other organisations, and society at large. It encompasses organisation-wide governance approaches as well as project-specific and system-level interventions designed to ensure that AI systems are accurate, accountable, fair and fit-for-purpose.⁴¹⁷

The seven key elements said to constitute AI governance are: accountability, oversight, roles and responsibilities; governance structures; people, skills, values, and culture; principles and policies; practices, processes and controls; supporting infrastructure; and stakeholder engagement, co-design and impact assessments⁴¹⁸ Although all of these may be relevant to the way in which organisations oversee AI, they are merely procedural and process steps that can be utilised to assist with compliance with legal norms. They do not constitute law or regulation.

⁴¹⁵ Bob Tricker, *Corporate Governance Principles, Policies, and Practices*, Oxford, 2009, p 7.

⁴¹⁶ John Uhrig, *Review of the Corporate Governance of Statutory Authorities and Office Holders*, Commonwealth of Australia, 2003, at p 17. See <http://cdn.nsw.ipaa.org.au/docs/Course-Readings---Dip-Gov-Policy-Development/23.%20The_Uhrig_Report_July_2003.pdf>.

⁴¹⁷ University of Technology Sydney, Human Technology Institute, *Essential Components of AI Governance: AI Governance Snapshot #1*, Shaping our Future Symposium 2024, 31 January 2024, p 2. See <<https://www.uts.edu.au/globalassets/sites/default/files/2024-01/ai-governance-snapshot---essential-components-of-ai-governance.pdf>>.

⁴¹⁸ Ibid pp 2-4.

If AI governance is intended to operate as a subset of overall governance, it is misconceived to conflate the processes and procedures by which control over an undertaking or activity is undertaken with the norms and rules that the processes and procedures are designed to meet. It conflates the normative content of laws, rules and standards – the rules that must be complied with – with the processes and procedures required to apply them. It is self-evident that AI risks need to be governed but it is a fundamental misconception of the nature of governance to use it as a substitute for proper regulation. When considering how to regulate AI, AI governance is necessary but hardly sufficient.

Conclusion

No new and disruptive technology including AI manifests itself into a regulatory void. A range of regulation already applies to AI including tort, contract, data protection, intellectual property, employment, environmental and occupational health and safety laws. Those who resist AI regulation do not often argue that they should be exempt from these laws.⁴¹⁹ The innovation v regulation debate most frequently arises in the context of whether or not new regulations are needed to address new disruptive risks or whether existing laws should be modified to do so or both.

‘At its foundation, regulation is about creating the conditions of certainty and trust, and the management of risk.’⁴²⁰ It is ‘fundamentally concerned with achieving public policy objectives or creating public value.’⁴²¹ Inappropriate or badly designed regulation, complex and uncertain regulation, poorly articulated roles and responsibilities and lack of clarity can have adverse impacts on innovation, not the least of which is that regulatory uncertainty can deter investment. But these are not arguments against regulation *per se*; they are arguments against poorly designed regulation. The ‘regulation stifles innovation’ argument overlooks the positive and facilitative effects that consistent and well-designed regulatory schemes can have in promoting innovative technologies such as AI within an environment of relative legal certainty, including promoting competition by ensuring that incumbents do not use regulation as a shield to protect against themselves being disrupted. Arguments that assert that regulation stifles innovation are simplistic and misleading: they are a false binary that

⁴¹⁹ Although they frequently try to limit their application to new technology. Good examples of this are data protection and copyright laws.

⁴²⁰ Arie Freiberg (above n 9) p 642.

⁴²¹ Ibid, p72.

also ignores the ‘broader set of legal and institutional reforms that are necessary for tech companies to innovate and for digital economies to thrive.’⁴²²

Several jurisdictions have developed approaches to regulate AI. Although they differ in fundamental ways, they each move beyond debates about the need for, and the legitimacy of, regulatory initiatives. Implicit in each is that AI innovation and regulation can coexist and that AI ethics and AI governance alone are unsatisfactory methods by which to regulate AI. The next chapter provides an account of these.

⁴²² Bradford (above n 391).

Chapter Five- Insights from international initiatives

Introduction

There have been a number of international initiatives designed to address issues associated with AI, discussed below. These have primarily focused on non-binding declarations and statements of intent. Only one treaty has been developed.

At the international level three broad models of AI regulation have emerged: the Chinese state-driven regulatory model, the European Union's rights-driven, risk-based regulatory model and a US market-driven approach.⁴²³ To this list may be added the UK's approach, which focuses on using existing legislation.

In western liberal economies the regulatory debate has revolved mainly around the approaches of the EU and the US or variations of them. However, recent policy developments in the US require a recalibrated consideration of the trajectory of US AI policy, which has shifted towards explicitly prioritising US global AI dominance and the removal of any regulatory impediments that are not congruent with this overarching objective.

This chapter does not aim to provide an exhaustive account of each of the main regulatory models. Its focus is to highlight their main elements, assess their commonalities and draw some conclusions about how they may inform Australia's approach to regulating AI.

International law and related initiatives

The Vilnius Convention

The first, and to date only, binding international treaty on AI is the Council of Europe's *Framework Convention on Artificial Intelligence and Human Rights, Democracy and the Rule of Law*⁴²⁴ (Vilnius Convention) which was opened for signature on 5 September 2024.

Although a Council of Europe treaty, it is open to all countries to sign and ratify it.

The Convention aims 'to ensure that activities within the lifecycle of artificial intelligence systems are fully consistent with human rights, democracy and the rule of law.'⁴²⁵

⁴²³ Anu Bradford, *Digital Empires: The Global Battle to Regulate Technology*, Oxford, 2023.

⁴²⁴ Council of Europe, *Council of Europe Framework Convention on Artificial Intelligence and Human Rights, Democracy and the Rule of Law*, Council of Europe Treaty Series No 225, Vilnius, 5 November 2024

⁴²⁵ Article 1(1).

Article 2 of the Convention defines an ‘artificial intelligence system’ to mean:

A machine-based system that, for explicit or implicit objectives, infers, from the input it receives, how to generate outputs such as predictions, content, recommendations or decisions that may influence physical or virtual environments. Different artificial intelligence systems vary in their levels of autonomy and adaptiveness after employment.

The Convention’s reference to lifecycle is designed to instantiate ‘a comprehensive approach towards addressing AI related risks and adverse impacts on human rights, democracy and the rule of law by capturing all stages of activities relevant to artificial intelligence systems.’⁴²⁶

The Convention primarily applies to the public sector⁴²⁷ and the private sector acting on behalf of the public sector.⁴²⁸ So far as the private sector is covered, each treaty party is obliged to address ‘risks and impacts ... in a manner conforming with the object and purpose of this Convention.’⁴²⁹ There are exemptions for national security,⁴³⁰ defence⁴³¹ and for research and development purposes.⁴³²

It requires each state party to ‘adopt or maintain measures to ensure that the activities within the lifecycle of artificial intelligence systems are consistent with obligations to protect human rights, as enshrined in applicable international law and in its domestic law.’⁴³³

To date, seventeen states have signed the Convention, including the UK and the US⁴³⁴ as well as the EU. Australia has not signed the Convention.

Other international activities

A range of declarations, statements of intent and communiques have emerged from a series of international meetings concerning AI.

These include the inaugural AI Safety Summit, hosted by the UK government in 2023, which produced the *Bletchley Declaration by Countries attending the AI Safety Summit* (Bletchley

⁴²⁶ Council of Europe, *Explanatory Report to the Council of Europe Framework Convention on Artificial Intelligence and Human Rights, Democracy and the Rule of Law*, para 14. See <<https://rm.coe.int/1680afae67>>.

⁴²⁷ Article 3(1)(a).

⁴²⁸ Ibid.

⁴²⁹ Article 3(1)(b).

⁴³⁰ Article 3(2).

⁴³¹ Article 3(4).

⁴³² Article 3(3).

⁴³³ Article 4.

⁴³⁴ Although the US has signed the Vilnius Convention it has not ratified it. Ratification appears highly unlikely given the current US government’s anti-regulatory posture for AI. This is discussed later in this chapter.

Declaration).⁴³⁵ The Bletchley Declaration recognises the potential benefits and risks of AI stating that ‘AI should be designed, developed, deployed, and used, in a manner that is safe, in such a way as to be human-centric, trustworthy and responsible.’⁴³⁶ Essentially discursive in nature and avoiding imposing any concrete regulatory obligations or actions on the States that participated⁴³⁷ or on non-State actors, the main outcomes were to emphasise the need for international cooperation, the need for working together ‘in an inclusive manner to ensure human-centric, trustworthy and responsible AI that is safe, and supports the good of all through existing international fora and other relevant initiatives, to promote cooperation to address the broad range of risks posed by AI’ and identifying AI risks and building risk-based policies to ensure AI safety.⁴³⁸

The UK AI Safety Summit was followed by the AI Seoul Summit in 2024, co-hosted by the UK government and the Republic of South Korea,⁴³⁹ the outcome of which was a further Declaration⁴⁴⁰ and two other instruments.⁴⁴¹ These non-binding initiatives focused on AI safety, international cooperation, education and training and the need to develop appropriate regulation and mitigation measures for risky AI. The AI Seoul Summit was followed by the Paris AI Action Summit in 2025, co-hosted by France and India, and attended by over ninety countries.⁴⁴² The Paris Summit included themes that extended beyond earlier preoccupation with AI safety and discussed issues such as sovereign AI infrastructure; the need for open-source AI models; the increasing demand for data to train AI; regulatory models; and the use of AI for crisis management. It produced the *Paris Statement on*

⁴³⁵ Prime Minister’s Office, UK government, ‘AI Safety Summit 2023: The Bletchley Declaration’, 1 November 2023. See <<https://www.gov.uk/government/publications/ai-safety-summit-2023-the-bletchley-declaration>>. For a copy of the Bletchley Declaration, see <<https://www.gov.uk/government/publications/ai-safety-summit-2023-the-bletchley-declaration/the-bletchley-declaration-by-countries-attending-the-ai-safety-summit-1-2-november-2023>>.

⁴³⁶ Ibid.

⁴³⁷ Department of Industry, Science and Resources, Australian Government, ‘The Bletchley Declaration by Countries Attending the AI Safety Summit’, 1-2 November 2023. Twenty-eight States, including Australia, as well as the European Union, participated. <<https://www.industry.gov.au/publications/bletchley-declaration-countries-attending-ai-safety-summit-1-2-november-2023>>.

⁴³⁸ Ibid.

⁴³⁹ AI Seoul Summit 2024, 21-22 May 2024. See <<https://www.gov.uk/government/topical-events/ai-seoul-summit-2024/about>>; <<https://aiseoulsummit.kr/?ckattempt=1>>.

⁴⁴⁰ Prime Minister’s Office, UK government, *Seoul Declaration for safe, innovative and inclusive AI by participants attending the Leaders’ Session: AI Seoul Summit*, 21 May 2024. The signatories were Australia, Canada, the European Union, France, Germany, Italy, Japan, the Republic of Korea, the Republic of Singapore, the United Kingdom, and the United States of America. See <<https://www.gov.uk/government/publications/seoul-declaration-for-safe-innovative-and-inclusive-ai-ai-seoul-summit-2024/seoul-declaration-for-safe-innovative-and-inclusive-ai-by-participants-attending-the-leaders-session-ai-seoul-summit-21-may-2024>>.

⁴⁴¹ *Seoul Ministerial Statement for advancing AI safety, innovation and inclusivity: AI Seoul Summit 2024*, which was signed by twenty-seven countries and the European Union on 22 May 2024. See <<https://www.gov.uk/government/publications/seoul-ministerial-statement-for-advancing-ai-safety-innovation-and-inclusivity-ai-seoul-summit-2024/seoul-ministerial-statement-for-advancing-ai-safety-innovation-and-inclusivity-ai-seoul-summit-2024>>. See also, the *Seoul Statement of Intent toward International Cooperation on AI Safety Science*, AI Seoul Summit 2024 (Annex).

<<https://www.gov.uk/government/publications/seoul-declaration-for-safe-innovative-and-inclusive-ai-ai-seoul-summit-2024/seoul-statement-of-intent-toward-international-cooperation-on-ai-safety-science-ai-seoul-summit-2024-annex>>.

⁴⁴² AI Action Summit, Paris, 10-11 February 2025. See <<https://www.elysee.fr/en/sommet-pour-l-action-sur-l-ia>>.

Inclusive and Sustainable Artificial Intelligence for People and the Planet,⁴⁴³ which identified future activity, such as AI accessibility, inclusive, transparent and safe and the need for ongoing international cooperation.

Although each of these initiatives canvassed familiar themes, none has led to more concrete regulatory initiatives. These have remained the responsibility of individual states.

China

Overview

China's approach to AI regulation has been gradual and incremental. Various regulatory measures, guidelines, opinions, rules and laws have evolved over time.

The Chinese government's functions and activities in respect of AI are multifaceted and encompass policymaking, regulation and investment, as well as it being an AI supplier and customer.⁴⁴⁴ China began building a policy framework and a 'regulatory architecture to guide its development and applications'⁴⁴⁵ as early as 2017. Its foundational policy for AI, *The New Generation Artificial Intelligence Development Plan*, highlights AI as a 'new focus of international competition'⁴⁴⁶ and 'a strategic technology that will lead in the future.'⁴⁴⁷ In common with many other jurisdictions, its overall objective is to maximise the economic benefits of AI while controlling its risks.

While vigorously developing AI, we must attach great importance to the potential safety risks and challenges, strengthen the forward-looking prevention and guidance on restraint, minimize risk, and ensure the safe, reliable, and controllable development of AI.⁴⁴⁸

A comprehensive range of development initiatives have been identified, accompanied by recognition of the need to 'develop laws, regulations, and ethical norms that promote the development of AI.'⁴⁴⁹ The plan nominates the legal issues associated with the 'healthy development of AI' as including civil and criminal responsibility for AI, protection of privacy and property and information security.⁴⁵⁰ Significantly it identifies the need to '[e]stablish a

⁴⁴³ AI Action Summit, *Statement on Inclusive and Sustainable Artificial Intelligence for People and the Planet*, 11 February 2025. See <<https://www.elysee.fr/admin/upload/default/0001/17/0efc53b2025d5bc1f0fc68556c41370cee166e7f.pdf>>. Sixty countries, including Australia, signed the *Statement*.

⁴⁴⁴ *Ibid.*, p 2.

⁴⁴⁵ Rogier Creemers, *The Regulation of Generative AI in China*, p 1. See

<<https://academic.oup.com/edited-volume/59908/chapter/512470025?login=true>> in Philipp Hacker, and others (eds), *The Oxford Handbook of the Foundations and Regulation of Generative AI*, online ed, Oxford Academic, 22 Apr. 2025.

⁴⁴⁶ Graham Webster, Rogier Creemers, Elan Kania and Paul Triolo, *Full Translation: China's 'New Generation Artificial Intelligence Development Plan'* (2017), Stanford University, DigiChina, para 1. See <<https://digichina.stanford.edu/work/full-translation-chinas-new-generation-artificial-intelligence-development-plan-2017/>>.

⁴⁴⁷ *Ibid.*

⁴⁴⁸ *Ibid.*

⁴⁴⁹ *Ibid.*, Part V(1), Guarantee Measures.

⁴⁵⁰ *Ibid.*

traceability and accountability system, and clarify the main body of AI and related rights, obligations, and responsibilities.⁴⁵¹

Pausing at this point, there are some important issues to note. The first comprehensive Chinese AI policy document was published in 2017, pre-dating the work of other jurisdictions by a significant margin. For example, the EU's foundational work on AI regulation, the White Paper of Artificial Intelligence, which articulates the joint objectives of promoting the uptake of AI whilst addressing its risks, was not published until 2020.⁴⁵² In addition, the Chinese foundational policy work does not suggest that there is a trade-off between AI innovation and regulation. To the contrary, it clearly considers that the benefits of AI are better realised where it is developed in tandem with appropriate regulation. As discussed in Chapter Four, China, which is one of the AI 'powerhouse' jurisdictions, has at no time proposed or recognised a binary choice between AI innovation and regulation.

Subsequently in 2020, China published legal reforms for a 'Rule of Law-Based Society',⁴⁵³ which emphasised that '[c]yberspace is not outside the law'⁴⁵⁴ and mandated the development and improvement of laws for 'new media operations and for algorithmic recommendations, deep fakes, and other applications of new technology ... and (to) strengthen normative guidance on the application of new technologies such as big data, cloud computing, and artificial intelligence.'⁴⁵⁵ Expanding on these reforms, the Cyberspace Administration of China (CCA) issued 'Guiding Opinions' for managing algorithms in online information services.⁴⁵⁶ The document states that over a period of three years 'central authorities would draft a comprehensive set of policies and regulations for algorithms, establish an administrative monitoring and oversight system, and 'standardize' the algorithmic ecosystem.'⁴⁵⁷ The first implementation activity concerned Algorithmic Recommendations Provisions⁴⁵⁸ developed by the Cyberspace Administration of China (CAC).

⁴⁵¹ Ibid.

⁴⁵² European Commission, *White Paper on Artificial Intelligence – A European Approach to Excellence and Trust*, 19 February 2020. See <https://commission.europa.eu/document/download/d2ec4039-c5be-423a-81ef-b9e44e79825b_en?filename=commission-white-paper-artificial-intelligence-feb2020_en.pdf>.

⁴⁵³ Chinese Communist Party Central Committee, *Implementation Outline for the Establishment of a Rule of Law-Based Society (2020-2025)*. See <<https://www.chinalawtranslate.com/en/establishing-a-rule-of-law-society/>>.

⁴⁵⁴ Ibid, Part VI.

⁴⁵⁵ Ibid, Part VI(22).

⁴⁵⁶ Chinese State Internet Information Office and Ors, *Guiding Opinions on Strengthening the Overall Governance of Internet Information Service Algorithms*, 17 September 2021. See <<https://www.chinalawtranslate.com/en/algorithm-plan/>>.

⁴⁵⁷ Rogier Creemers (above, n 445), p 6.

⁴⁵⁸ Cyberspace Administration of China and Ors, *Provisions on the Management of Algorithmic Recommendations in Internet Information Services*, 31 December 2021. See <<https://www.chinalawtranslate.com/en/algorithms/>>.

Algorithm registration

These provisions cover a range of applications of algorithmic recommendation technology, meaning ‘the use of algorithm technologies types (sic) such as generation and synthesis, individualized pushing, sequence refinement, search filtering, schedule decision-making, and so forth to provide users with information.’⁴⁵⁹ One of the most significant regulatory requirements is the establishment of an algorithm register where, within ten days of providing services

algorithmic recommendation service providers that have that have public opinion properties or capacity for social mobilization shall provide the service provider's name, service form, field of application, type of algorithm, algorithm self-assessment report, content to be displayed, and other such information through the internet information services algorithm filing system, and perform filing procedures.⁴⁶⁰

These requirements came into effect on 1 March 2022.⁴⁶¹ A list of the filings published between August 2022 and July 2025 shows the extensive reach of the registration requirements: hundreds of service providers have registered their algorithms.⁴⁶² While the public-facing register shows only high-level descriptions of the nature and scope of the algorithms and self-assessments, it includes ‘a basic identification system for algorithms, the possibility of sanction for companies that did not register their algorithms, and an explicit ground for complaint if algorithms were used in excess of their stated purposes.’⁴⁶³

An analysis of the register shows that only part of the information filed is publicly facing. The filings:

usually consisted of a single page with six different short-response categories, including ‘Algorithm Fundamentals’ and ‘Algorithm Operating Mechanism’... But the actual descriptions of these filings were pitched at such a high level as to be almost devoid of meaningful detail ... If this were the full extent of information given to Chinese regulators, it would provide them with no meaningful insights into the algorithms, how they were trained, or how they might perform.⁴⁶⁴

Further examination of the disclosure requirements indicate that Chinese regulators receive much more detailed information:

⁴⁵⁹ Ibid, Art 2.

⁴⁶⁰ Ibid, Art 24.

⁴⁶¹ Ibid, Art 35.

⁴⁶² State Internet Information Office of the People's Republic of China and Anor, *Announcement of the State Internet Information Office on the Release of Internet Information Service Algorithm Record Information*. See <https://www.cac.gov.cn/2022-08/12/c_1661927474338504.htm>.

⁴⁶³ Rogier Creemers (above n 445), p 8.

⁴⁶⁴ Matt Sheehan and Sharon Du, *What China's Algorithm Registry Reveals about AI Governance*, Carnegie Endowment, 9 December 2022. See <<https://carnegieendowment.org/posts/2022/12/what-chinas-algorithm-registry-reveals-about-ai-governance?lang=en>>.

[A] closer look at the algorithm registry's landing page, which contains a downloadable user manual for entities registering their algorithms, provides a wider window into what information the CAC is actually gathering. A close read of that manual and examination of the screenshots within it show that only a portion of the information filed in the registry has been revealed.⁴⁶⁵

The exact reasons for the delineation between the publicly available information and the more detailed information provided to regulators is not stated. However, possible explanations are that the full public disclosure of algorithms could reveal sensitive, commercially valuable confidential information and that China's preoccupation is to ensure that its government and regulators are able to determine whether algorithms satisfy broader Chinese political and social objectives rather than providing individuals with granular AI transparency.

Generative AI and training data

The *Interim Measures for the Management of Generative AI Services* (Interim GAI Measures) were published in July 2023 and took effect on 15 August 2023. These include requirements that the provision and use of generative AI services must:

- take effective measures to prevent discrimination when designing algorithms, selecting training data, generating and optimizing models and providing services;⁴⁶⁶
- respect intellectual property rights;⁴⁶⁷
- not infringe privacy and personal information rights;⁴⁶⁸ and
- take effective measures to improve the transparency of generative AI services and improve the accuracy and reliability of generated content based on the characteristics of service types.⁴⁶⁹

There are also specific requirements applicable to training data that mandate providers of GAI 'carry out training data processing activities such as pre-training and optimization training in accordance with the law.'⁴⁷⁰ These stipulate that pre-training, training and optimisation processes must be undertaken in accordance with the law, including only using

⁴⁶⁵ Ibid.

⁴⁶⁶ State Internet Information Office of the People's Republic of China and anor, *Interim Measures for Generative Artificial Intelligence Service Management*, 13 July 2023, Article 4(2) See <https://www.cac.gov.cn/2023-07/13/c_1690898327029107.htm>.

⁴⁶⁷ Ibid, Article 4(3).

⁴⁶⁸ Ibid, Article 4(4).

⁴⁶⁹ Ibid, Article 4(5).

⁴⁷⁰ Ibid, Art7.

data and basic models that are otherwise legal;⁴⁷¹ not infringing intellectual property rights;⁴⁷² obtaining consent where personal information is used;⁴⁷³ and taking ‘effective measures to improve the quality of training data and enhance the authenticity, accuracy, objectivity and diversity of training data.’⁴⁷⁴ Article 8 contains extensive obligations concerning the labelling of training data including quality assessments sampling and accuracy verification and training for labelling personnel.

Supervision, Remedies and Jurisdiction

Chapter IV of the Interim GAI Measures sets out a framework for oversight, technical state support and the responsibilities of GAI providers.

In terms of supervision, a broad range of public and private institutions are required to ‘strengthen the management of generative artificial intelligence services in accordance with their respective responsibilities’⁴⁷⁵ Relevant government agencies are also required to ‘improve the scientific supervision methods suitable for innovation and development ... and formulate corresponding classification and grading supervision rules or guidelines.’⁴⁷⁶

A key feature of the regulation relates to the conduct of security assessments. Providers of generative AI services ‘with public opinion attributes or social mobilization ability’⁴⁷⁷ must conduct specified security assessments, as well as complying with requirements relating to algorithm filing. At the same time, ‘bodies and personnel participating in security assessments and oversight inspections of generative AI services are required to keep the confidentiality of state secrets, commercial secrets, personal privacy, and personal information that they learn of in performing their duties, and must not leak or unlawfully provide it to others.’⁴⁷⁸

More generally, providers must ‘cooperate in accordance with the law, explain the source, scale, type, labelling rules, algorithm mechanism, etc. of the training data as required, and provide necessary technical, data and other support and assistance.’⁴⁷⁹

⁴⁷¹ Ibid, Art 7(1).

⁴⁷² Ibid, Art7(2).

⁴⁷³ Ibid, Art 7(3).

⁴⁷⁴ Ibid, Art 7(4).

⁴⁷⁵ Ibid, Art 16.

⁴⁷⁶ Ibid.

⁴⁷⁷ Ibid, Art17.

⁴⁷⁸ Ibid, Art 19.

⁴⁷⁹ Ibid.

The regulation also gives end users who consider that GAI services do not comply with it ‘the right to complain and report to the relevant competent departments.’⁴⁸⁰ A final noteworthy feature concerns the scope of its application to providers of GAI services from outside China. It states that:

Where generative AI services provided from outside the [mainland] PRC do not meet the requirements of laws, administrative regulations, or these Measures, the state internet information department shall notify the relevant organs to employ technical measures and other necessary measures to address it.⁴⁸¹

Summary

China has taken a cautious and iterative approach to AI regulation. Although some of the rules it emphasises (such as core socialist values and associated ideological propositions)⁴⁸² are at odds with more familiar liberal western democratic frameworks, much of its regulatory work attempts to strike a balance between promoting the benefits of AI and ameliorating its risks and provide interesting perspectives worthy of close consideration. For the purposes of this thesis, Chinese perspectives on algorithmic registration and regulation and ensuring the accuracy and quality of training data while preserving data protection and intellectual property rights hold particular interest in relation to the Australian context. Although China’s AI industry has developed rapidly, it is important to understand that some of its objectives, such as China’s interest in regulating the dissemination of information and undertaking widespread surveillance of its citizens is not compatible with the values associated with western democracies. Those features are not recommended for an Australian approach to AI regulation. Nevertheless, there are a range of elements of Chinese regulation that are comparable to the *ex ante* initiatives adopted in the EU AI Act that can usefully the future of AI regulation in Australia.

⁴⁸⁰ Ibid, Art 18.

⁴⁸¹ Ibid, Art 20.

⁴⁸²For example, the “Guiding Thought” that underpins the *Implementation Outline for the Establishment of a Rule of Law-Based Society (2020-2025)* is to “Hold high the great banner of socialism with Chinese characteristics, adhere to the guidance of Marxism-Leninism, Mao Zedong Thought, Deng Xiaoping Theory, the important thinking of “Three Represents”, the scientific development concept, and Xi Jinping Thought on Socialism with Chinese characteristics for a New Era; and fully implement the spirit of the Party’s 19th National Congress and its Second, Third, Fourth and Fifth Plenary Sessions, fully implement Xi Jinping’s thought on the rule of law, strengthen the “four awarenesses”, solidify the “four self-confidences”, and achieve the “two preservasions”; unerringly follow the road of socialist rule of law with Chinese characteristics, persist in the integration of a nation, government, and society ruled by law; cultivate and practice the Core Socialist Values, carry forward the spirit of socialist rule of law spirit, build a socialist rule of law culture, and enhance the enthusiasm and initiative of the entire society to enforce the rule of law; promote the entire society’s compliance, study, and use of law; improve legal safeguards for social fairness and justice; safeguard people’s rights, raise the level at which social governance is brought under rule of law, and build a solid foundation of rule of law for the Chinese dream of fully building a modern socialist country and realizing the great rejuvenation of the Chinese people.”

Importantly China's AI industry has flourished in a regulated environment, the best-known example being the release of DeepSeek⁴⁸³ in January 2025 followed by other innovative AI models. The rise to prominence of Chinese AI that appears capable of matching and possibly exceeding those developed in the US has occurred despite — and perhaps because of the certainty provided by regulation — calling into question arguments that regulation stifles innovation.

European Union

Overview

The European Union's (EU) AI Act (EU AI Act)⁴⁸⁴ was the first piece of comprehensive AI regulation. It entered into force on 1 August 2024 and requires several implementation steps before becoming fully effective on 2 August 2026. It is complex legislation that is patterned on European product liability laws, taking a risk-based regulatory approach that regulates AI systems, delineates between levels of AI risk and establishes regulatory measures corresponding to each.

The EU AI Act seeks to address multiple objectives including the development of a 'uniform legal framework'⁴⁸⁵ for AI systems across all EU members; to promote the uptake of human-centric and trustworthy AI systems; and to ensure

a high level of protection of health, safety, fundamental rights as enshrined in the Charter of Fundamental Rights of the European Union (the Charter), including democracy, the rule of law and environmental protection, to protect against the harmful effects of AI systems in the Union, and to support innovation.⁴⁸⁶

In common with the Chinese regulatory system, the EU AI Act aspires to balance innovation with risks and emphasises the need for compliance with rights and the rule of law.

A risk-based approach

The EU AI Act regulates AI systems as follows based on the risks that they pose:

- 1) Systems that are considered to pose unacceptable risks are prohibited.⁴⁸⁷

These include AI systems that deploy subliminal techniques that distort behaviour

⁴⁸³ See <<https://www.deepseek.com>>.

⁴⁸⁴ Regulation (EU) 2024/1689 of the European Parliament and of the Council, 13 June 2024, 'EU AI Act.' See <https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=OJ:L_202401689>.

⁴⁸⁵ EU AI Act, Recital 1.

⁴⁸⁶ Ibid.

⁴⁸⁷ EU AI Act, Art 1.

and informed decision-making;⁴⁸⁸ those that exploit vulnerabilities related to age, disability or socio-economic circumstances to distort behaviour;⁴⁸⁹ biometric categorisation systems that infer sensitive personal attributes;⁴⁹⁰ social scoring mechanisms and manipulative AI unrelated to the context that data was originally collected or generated;⁴⁹¹ that assess the risk or criminality solely based on profiling or personal traits;⁴⁹² compiling facial recognition databases by untargeted scraping of facial images from the internet or CCTV footage;⁴⁹³ inferring emotions in workplaces or educational institutions other than for medical or safety reasons;⁴⁹⁴ and real time remote biometric identification in public places for law enforcement purposes.⁴⁹⁵

- 2) Other ‘high risk’ systems are subject to extensive regulation, as outlined below.
- 3) ‘Limited risk’ systems are subject to more minimal restrictions that focus on transparency, i.e., informing users that they are interacting with an AI system⁴⁹⁶ and that content with which they are provided is marked as having been generated artificially.⁴⁹⁷
- 4) ‘Minimal risk’ AI systems are unregulated unless they are general purpose AI. Examples include ‘AI-enabled video games and spam filters.’⁴⁹⁸

High-risk AI systems

A classification system has been designed to assist in determining whether an AI system is high risk.⁴⁹⁹ Essentially there are three ways in which an AI system can be considered ‘high risk’: when the AI system is a product;⁵⁰⁰ when it is a safety component of a product;⁵⁰¹ or where it meets the definition of a high-risk system set out in Annex III of the EU AI Act.

⁴⁸⁸ EU AI Act, Art 5(1)(a).

⁴⁸⁹ EU AI Act, Art 5(1)(b).

⁴⁹⁰ EU AI Act, Art 5(1)(c).

⁴⁹¹ See EU AI Act, Article 5.

⁴⁹² EU AI Act, Art 5(1)(d).

⁴⁹³ EU AI Act, Art 5(1)(e).

⁴⁹⁴ EU AI Act, Art 5(1)(f).

⁴⁹⁵ EU AI Act Art 5(1)(f). There are some exceptions such as for searching for missing persons, preventing a substantial and imminent threat to life or a foreseeable terrorist attack or identifying suspects in serious crimes.

⁴⁹⁶ EU AI Act, Art 50(1).

⁴⁹⁷ EU AI Act, Art 50 (2), (3) (4) and (5).

⁴⁹⁸ Future of Life Institute, *High-level Summary of the AI Act*, 27 February 2024. See < <https://artificialintelligenceact.eu/high-level-summary/> >.

⁴⁹⁹ EU AI Act, Ch III, s 1.

⁵⁰⁰ EU AI Act, Art 6(1)(a).

⁵⁰¹ Ibid.

The Annex III requirements broadly cover AI systems used in biometrics;⁵⁰² AI used in critical infrastructure;⁵⁰³ AI used to determine access or admission to education and vocational training,⁵⁰⁴ evaluating learning outcomes,⁵⁰⁵ assessing the level of education an individual will be able to access or receive,⁵⁰⁶ and monitoring and detecting prohibited student behaviour during tests;⁵⁰⁷ employment recruitment,⁵⁰⁸ and promotion and evaluation;⁵⁰⁹ AI systems intended to be used by public sector organisations to evaluate entitlements to benefits and services,⁵¹⁰ evaluation of creditworthiness or credit scoring (except credit fraud),⁵¹¹ and risk assessment and pricing for life and health insurance;⁵¹² a variety of law enforcement uses, including AI systems used to assess whether a person is likely to become a victim of crime,⁵¹³ evaluating evidence,⁵¹⁴ assessing the risk of offending or re-offending or assessing personality traits and characteristics or past criminal behaviour,⁵¹⁵ and profiling individuals in the course of detecting, investigating and prosecuting criminal offences;⁵¹⁶ immigration, asylum and border control management;⁵¹⁷ and the administration of justice and democratic processes,⁵¹⁸ including use by the judiciary to research and apply the law and facts,⁵¹⁹ and the use of AI systems to influence elections or voting behaviour.⁵²⁰

The EU regulatory obligations delineate between *providers*⁵²¹ and *deployers*⁵²² of high-risk AI systems and impose obligations on importers⁵²³ and distributors.⁵²⁴ The majority of obligations are imposed on providers, i.e., the developers of AI systems. Across each of these

⁵⁰² EU AI Act, Art 1, Annex III,

⁵⁰³ EU AI Act, Art 2, Annex III.

⁵⁰⁴ EU AI Act, Art 3(a), Annex III.

⁵⁰⁵ EU AI Act Art 3(b), Annex III.

⁵⁰⁶ EU AI Act, Art 3(c), Annex III.

⁵⁰⁷ EU AI Act, Art 3(d), Annex III.

⁵⁰⁸ EU AI Act, Art 4(a), Annex III.

⁵⁰⁹ EU AI Act, Art 4(d), Annex III.

⁵¹⁰ EU AI Act, Art 5(a), Annex III.

⁵¹¹ EU AI Act, Art 5(b), Annex III.

⁵¹² EU AI Act, Art 5(c), Annex III.

⁵¹³ EU AI Act, Art 6(a), Annex III.

⁵¹⁴ EU AI Act, Art 6(c), Annex III.

⁵¹⁵ EU AI Act, Art 6(d), Annex II.

⁵¹⁶ EU AI Act, Art 6(e), Annex III.

⁵¹⁷ EU AI Act Art 7, Annex III.

⁵¹⁸ EU AI Act, Art 8, Annex II.

⁵¹⁹ EU AI Act, Art 8(a), Annex III.

⁵²⁰ EU AI Act Art 8(b), Annex III.

⁵²¹ Under Article 3 of the EU AI Act a provider ‘means a natural or legal person, public authority, agency or other body that develops an AI system or a general-purpose AI model or that has an AI system or general purpose AI model developed and places it on the market or puts the AI system into service under its own name or trademark, whether for payment or free of charge.’

⁵²² Under Article 3 of the EU AI Act, a deployer is ‘a natural or legal person, public authority, agency or other body using an AI system under its authority except where the AI system is used in the course of a personal non-professional activity.’

⁵²³ Under Article 3 of the EU AI Act, an importer ‘means a natural or legal person located or established in the Union that places on the market an AI system that bears the name or trademark of a natural or legal person established in a third country.’

⁵²⁴ Under Article 3 of the EU AI Act, a distributor ‘means a natural or legal person in the supply chain, other than the provider or the importer, that makes an AI system available on the Union market.’

categories, except for prohibited AI, both *ex ante* and *ex post* regulatory measures are used. Despite this complex regulatory architecture, there are no liability provisions for end users who sustain loss or damage from an AI system. Remedies are limited to defined complaint mechanisms.

Providers of high-risk AI systems must comply with a number of regulatory requirements.⁵²⁵ These are primarily (although not exclusively) *ex ante* measures that are designed to ensure these systems are safe *before* they are provided and include establishing a proper risk-management system;⁵²⁶ where the model involves the use of training data, employing proper data governance and management practices;⁵²⁷ the development of technical documentation before the AI system is put on the market and that is kept up-to-date which documents compliance with the requirements set out in Chapter III, Section 2 of the EU AI Act; the establishment and maintenance of logs; transparency obligations such that the operation of a high-risk AI system 'is sufficiently transparent to enable deployers to interpret the system's output and use it appropriately';⁵²⁸ provision for human oversight;⁵²⁹ a requirement that high-risk AI systems be 'designed and developed in such a way that they achieve an appropriate level of accuracy, robustness, and cybersecurity, and that they perform consistently in those respects across their lifecycle';⁵³⁰ that quality management systems be put in place to ensure compliance;⁵³¹ that technical compliance documentation be retained for ten years;⁵³² a requirement that if non-compliance with regulatory requirements is apparent, corrective action be taken immediately or for the AI system to be withdrawn or disabled;⁵³³ and that providers have a duty to cooperate with competent regulatory authorities.⁵³⁴

In general terms, these obligations extend across the AI value chain, i.e., to distributors, deployers or any third party considered to be a provider.⁵³⁵ For most deployments of high-risk AI systems a fundamental rights impact assessment is required.⁵³⁶

⁵²⁵ These are set out in detail in EU AI Act, Art 16.

⁵²⁶ EU AI Act, Art 19.

⁵²⁷ EU AI Act, Art 10.

⁵²⁸ EU AI Act, Art 13 (1).

⁵²⁹ EU AI Act, Art 14.

⁵³⁰ EU AI Act, Art 15(1).

⁵³¹ EU AI Act, Art 17.

⁵³² EU AI Act, Art 18(1).

⁵³³ EU AI Act, Art 20.

⁵³⁴ EU AI Act, Art 21.

⁵³⁵ These arrangements are instantiated through a series of provisions including Art 22, Art 23, Art 24, Art 25 and Art 26 EU AI Act.

⁵³⁶ EU AI Act, Art 27.

General purpose AI (GPAI)

GPAI is defined as

an AI model ... that displays significant generality and is capable of competently performing a wide range of distinct tasks regardless of the way the model is placed on the market and that can be integrated into a variety of downstream systems or applications ...⁵³⁷

This definition has been supplemented by voluntary, soft law guidelines, the *General-Purpose AI Code of Practice*⁵³⁸ (GPAI Code) that sets out a regulatory interpretation. As explained in the Code, GPAI providers who sign-up to it 'can show they comply with the AI Act by adhering to the code. This will reduce their administrative burden and give them more legal certainty than if they proved compliance through other methods.'⁵³⁹

GPAI regulation is divided into two main categories. The first consists of regulatory measures that apply to GPAI in general. The second, a subset of the first, applies to GPAI that constitutes a systemic risk.

In overall terms, providers of GPAI must develop and maintain technical documentation about the model except where the model is released under a free and open-source licence; provide information and documentation to downstream providers to assist them to understand the model; establish a policy to comply with EU copyright law; and publish a summary of the content used for training the model.⁵⁴⁰

There are additional requirements where GPAI raises systemic risks. This is determined by reference to computing power thresholds:

The AI Act presumes (that) for models trained with a cumulative amount of computational resources exceeding 10^{25} floating point operations (the 'compute threshold') ...have such capabilities. This threshold suggests that the model could significantly impact the Union market due to its reach or potential negative effects on public health, safety, security, fundamental rights, or society.⁵⁴¹

The additional regulatory requirements involve 'model evaluation in accordance with standardised protocols and tools reflecting state of the art, including conducting and

⁵³⁷ EU AI Act, Art 3(63).

⁵³⁸ European Commission, *The General-Purpose AI Code of Practice*, 10 July 2025. See <<https://digital-strategy.ec.europa.eu/en/policies/contents-code-gpai>>.

⁵³⁹ Ibid.

⁵⁴⁰ European Commission, *Guidelines on Obligations for General-Purpose AI Providers*, 22 July 2025. See <<https://digital-strategy.ec.europa.eu/en/faqs/guidelines-obligations-general-purpose-ai-providers>>.

⁵⁴¹ Ibid.

documenting adversarial testing of the model;⁵⁴² ‘assess and mitigate possible systemic risks at Union level, including from their sources’;⁵⁴³ ‘keep track of, document, and report ... relevant information about serious incidents and possible corrective measures to address them’⁵⁴⁴ and ‘ensure an adequate level of cybersecurity protection’ for the AI model and its infrastructure.⁵⁴⁵

The overall thrust of these regulatory initiatives is to provide GPAI providers with a compliance pathway provided they comply with a range of transparency and notification measures.

Scope

A key feature of the EU AI Act is that it has extra-territorial effect. It applies to providers ‘placing on the market or putting into service AI systems or placing on the market general-purpose AI models in the Union, *irrespective of whether those providers are established or located within the Union or a third country*’⁵⁴⁶ as well as to ‘providers and deployers of AI systems that have their place of establishment or are located in a third country, where the output produced by the AI system is used in the Union.’⁵⁴⁷ ‘Output’ is not defined by the AI Act but it ‘could include any content, predictions, recommendations or decisions of an AI system (as per the OECD definition of outputs) ... [and] organisations [should] err on the side of caution and adopt a broad interpretation of the term ...’⁵⁴⁸ This broad approach to extra-territoriality means that organisations anywhere in the world that might wish to provide AI systems into the EU will be required to comply with its regulatory requirements.

Penalties

EU AI Act penalties are significant. For example, breaches associated with prohibited AI systems under Article 5 are an administrative fine of up to €35,000,000, or where the offender is an undertaking, up to 7% of total worldwide annual turnover for the preceding financial year, whichever is higher.⁵⁴⁹ For breaches associated with high-risk AI systems can

⁵⁴² EU AI Act, Art 55(1)(a).

⁵⁴³ EU AI Act, Art 55(b).

⁵⁴⁴ EU AI Act, Art 55(c).

⁵⁴⁵ EU AI Act, Art 55(d).

⁵⁴⁶ EU AI Act, Art 2(1)(a). (My emphasis.)

⁵⁴⁷ EU AI Act, Article 2(1)(c).

⁵⁴⁸ Olivia Newbold, *Extra-territorial Application of the AI Act: How will it Impact Australian Organisations?*, 20 February 2024. See <<https://www.dlapiper.com/en-au/insights/publications/2024/02/extra-territorial-application-of-the-ai-act-how-will-it-impact-australian-organisations>>.

⁵⁴⁹ EU AI Act, Art 99(3).

involve penalties of up to €15,000,000 of, for an undertaking, up to 3% of total worldwide turnover.⁵⁵⁰

Summary

The EU AI Act is risk-based and technology neutral regulation that is intended to promote innovation but also to provide robust protection from AI harms through a precautionary approach and by explicitly requiring that AI systems comply with the rule of law and fundamental rights. It has, however, been criticised as stifling innovation through its complexity and over-prescriptive approach that has seen the technology industry lobby for change or to dilute implementation measures. The EU, under pressure from both the technology industry leading up to and at the Paris AI Action Summit in February 2025 and concerned about the slow growth of AI in the EU, has diluted its regulatory strategy. It has abandoned its AI Liability Directive (AILD).⁵⁵¹ It is also proposed that the *General Data Protection Regulation* (GDPR) be simplified.⁵⁵² The Code of Practice on GPAI has been under intensive pressure from aggressive Big Tech lobbying⁵⁵³ based on arguments that dilution is ‘essential to keep pace with US and Chinese giants’.⁵⁵⁴ This has led to criticism that this deregulatory turn ‘risks trading away democratic safeguards, without addressing systemic challenges to AI regulation.’⁵⁵⁵

While it seems unlikely that the EU will recast the AI Act, it is possible that it might retreat from its initial regulatory posture by limiting the AI Act’s regulatory oversight or by redirecting regulators from enforcement activities to a more collaborative, less punitive, approach. With full AI Act implementation not scheduled until August 2026 it is too early to speculate on these potential outcomes.

⁵⁵⁰ EU AI Act, Art 99(4).

⁵⁵¹ The AILD was omitted from the EU Commission’s 2025 work programme. See

<https://commission.europa.eu/strategy-and-policy/strategy-documents/commission-work-programme/commission-work-programme-2025_en>.

⁵⁵² See letter from the European Data Protection Board and the European Data Protection Supervisor to the Chair European Commission, 8 May 2025. See <https://www.edpb.europa.eu/system/files/2025-05/20250508_edpb-edps-letter-commission-draft-proposal-simplification-record-keeping-obligations-gdpr_en.pdf>.

⁵⁵³ Corporate Europe Observatory, ‘EU Under Heavy Big Tech Pressure to Weaken Rules on Advanced AI’, 29 April 2025. See <<https://corporateeurope.org/en/2025/04/eu-under-heavy-big-tech-pressure-weaken-rules-advanced-ai#:~:text=Brussels%2C%2030%20April%202025%20%2D%20As,social%20harms%20caused%20by%20AI>>.

⁵⁵⁴ Raluca Csernaton, *The EU’s AI Power Play: Between Deregulation and Innovation*, Carnegie Endowment for International Peace, 20 May 2025. See <<https://carnegieendowment.org/research/2025/05/the-eus-ai-power-play-between-deregulation-and-innovation?lang=en>>.

⁵⁵⁵ Ibid.

United States

Federal overview

Following the 2024 US Presidential election, much of the prior administration's light-touch, primarily self-regulation-based AI rules were jettisoned, replaced with a consultation process to develop an AI Action Plan.

An AI Bill of Rights developed by the previous administration had proposed five principles to 'help guide the design, use, and deployment of automated systems to protect the rights of the American public'.⁵⁵⁶ The non-enforceable, non-binding principles cover safe and effective systems; protection against algorithmic discrimination; safeguarding data privacy; providing notice and explanation; and provision of a human alternative to machine-based decision through an opt-out mechanism.

While the White House Blueprint for an AI Bill of Rights⁵⁵⁷ has been retained for now, its status remains unclear following the publication of the US AI Action Plan in July 2025.⁵⁵⁸ The latter expresses a policy goal 'to achieve global dominance in artificial intelligence', emphasising that it is 'imperative that the United States and its allies win this race'.⁵⁵⁹ The plan further states that in order to achieve this dominance the US 'will continue to reject radical climate change dogma and bureaucratic red tape ... Simply put, we need to "Build, Baby, Build."'⁵⁶⁰

The AI Action Plan goes beyond market-driven regulation given its expressly anti-regulatory intent that aims to 'dismantle unnecessary regulatory barriers that hinder the private sector'.⁵⁶¹ It further states that with the aim of preventing woke AI, 'our AI systems must be free from ideological bias and be designed to pursue objective truth rather than social engineering agendas ...'⁵⁶² The plan is intended to be implemented in the near-term.⁵⁶³

The AI Action Plan crystallises an explicitly pro-innovation approach to AI that minimises regulatory intervention and parallels the US default policy stance on internet regulation that

⁵⁵⁶ Ibid.

⁵⁵⁷ The White House, *Blueprint for an AI Bill of Rights*, October 2022. See

<<https://bidenwhitehouse.archives.gov/ostp/ai-bill-of-rights/what-is-the-blueprint-for-an-ai-bill-of-rights/>>.

⁵⁵⁸ Executive Office of the President, *Winning the Race: America's AI Action Plan*, July 2025. See <<https://www.whitehouse.gov/wp-content/uploads/2025/07/Americas-AI-Action-Plan.pdf>>. AI Action Plan is accompanied by three Executive Orders corresponding to the three 'pillars' identified within the AI Action Plan.

⁵⁵⁹ Ibid, p 1.

⁵⁶⁰ Ibid.

⁵⁶¹ Ibid.

⁵⁶² Ibid, p 2.

⁵⁶³ Ibid.

regulation should be left to the market. This does not, however, rule out some patchwork, sectoral regulation that responds to specific risks or policy concerns such as national security and public safety risks.

In the meantime some US states have started to fill this regulatory lacuna by developing their own AI regulation in a manner similar to their development of state-based data protection legislation.

State overview

Four of the US States, California, Colorado, Nevada and Texas have developed AI regulations that are now in force.⁵⁶⁴ Although a review of each is beyond the scope of a chapter that is focused on national approaches, a short account of some of California's AI regulation follows, given that it is the epicentre of AI technology development in the US.

California Assembly Bill (AB) 2885 provides a definition of AI for the state's purposes:

Artificial intelligence means an engineered or machine-based system that varies in its level of autonomy and that can, for explicit or implicit objectives, infer from the input it receives how to generate outputs that can influence physical or virtual environments.⁵⁶⁵

AB 2013, known as the *Generative Artificial Intelligence Training Data Transparency Act 2024*, (GAITDTA) requires 'developers of GenAI systems to publicly disclose detailed information about the datasets used in their development on their website. The law's transparency mandate applies to all GenAI systems and services made available to Californians.'⁵⁶⁶ Generative AI is defined broadly 'artificial intelligence that can generate derived synthetic content, such as text, images, video, and audio, that emulates the structure and characteristics of the artificial intelligence's training data.'⁵⁶⁷ The information that must be disclosed consists of a high-level summary of datasets used in the system's development; information on dataset sources, ownership, and intended purpose; descriptions of data types, including whether they include data protected by copyright, trademark, or patents, and whether the data includes personal information or aggregated consumer data; details on synthetic data usage, dataset cleaning or processing, and whether

⁵⁶⁴ As at 1 July 2025. See <<https://iapp.org/resources/article/us-state-ai-governance-legislation-tracker/#state-ai-governance-law-chart>>.

⁵⁶⁵ California Legislative Information, s 3 *Assembly Bill No. 2885, Chapter 843*. The Bill was enacted on 28 September 2024. See <https://leginfo.ca.gov/faces/billNavClient.xhtml?bill_id=202320240AB2885>.

⁵⁶⁶ California Legislative Information, *Assembly Bill No. 2014*. See <<https://legiscan.com/CA/text/AB2013/2023>>. See also, Baker Botts, *California's AB 2013: Challenges and Opportunities in Generative AI Compliance*, 25 November 2024. See <https://www.bakerbotts.com/thought-leadership/publications/2024/november/ca-ab-2013_gen-ai-compliance>.

⁵⁶⁷ GAITDTA s 3110(c).

datasets were purchased or licensed; the timeframes for data collection and the dates datasets were first used; and whether synthetic data generation was used in developing the GenAI system or service.⁵⁶⁸ These requirements come into effect on 1 January 2026.

California's approach also encompasses specific legislative initiatives to address AI-generated child pornography, non-consensual deep fake pornography, filtering of sexually explicit identity theft on social media platforms, safeguards against unauthorised digital replicas in contracts, clarification that personal information so that it covers AI-generated information about a person and unauthorised use of deceased personalities' replicas amongst others.⁵⁶⁹

Summary

The US federal government's retreat from AI regulation and anti-regulatory stance has meant that some US states have moved to fill this gap, with many others contemplating their own laws. This necessarily leads to a regulatory environment that will consist of a patchwork of different laws that will be complex and time consuming to navigate. These will necessarily undermine federal anti-regulation policy. Any US AI developer that wishes to make available AI systems to the citizens of California (or any other regulated US state jurisdiction) will need to undertake compliance activities. Equally that same developer seeking to provide its AI systems to the EU or China will also need to comply with those jurisdictions' regulatory settings.

⁵⁶⁸ GAITDTA § 3111.

⁵⁶⁹ [Jeewon K. Serrato](#), Christine Mastromonaco, Shruti Bhutani Arora, Andrew Caplan, Erin Choo, Mia Rendar, Leighton Watson, Anne M. Voigts, Shani Rivaux, Johanna Purcell, and Day Feyisayo Ajanaku, *California's AI Laws are here – Is Your Business Ready?* 2 July 2025. See <<https://www.pillsburylaw.com/en/news-and-insights/california-ai-laws.html>>.

United Kingdom

Overview

The UK is taking an explicitly pro-innovation approach to AI. Its regulatory posture is based on a March 2023 white paper, 'A Pro-innovation approach to AI',⁵⁷⁰ together with the government's 2024 response.⁵⁷¹

Instead of developing bespoke AI regulation the UK government's position is that 'a deliberately agile and iterative approach'⁵⁷² to regulation is appropriate. To support this, five principles have been developed 'to guide and inform the responsible development and use of AI in all sectors of the economy.'⁵⁷³ The five principles are safety, security and robustness; appropriate transparency and explainability; fairness; accountability and governance; and contestability and redress.⁵⁷⁴

The five principles will not be embodied in statute for the reason that:

[n]ew rigid and onerous legislative requirements on businesses could hold back AI innovation and reduce our ability to respond quickly and in a proportionate way to future technological advances. Instead the principles will be issued on a non-statutory basis and implemented by existing regulators. This approach makes use of regulators' domain-specific expertise to tailor the implementation of the principles to the specific context in which AI is used.⁵⁷⁵

Instead of introducing overarching AI laws, the UK has chosen to adopt a principles-based approach 'for existing sector-specific regulators to interpret and apply to the development and use of AI within their domains.'⁵⁷⁶ Existing regulators including the Competition and Markets Authority, the Information Commissioner's Office, the Equality and Human Rights Commission, the Financial Conduct Authority, the Intellectual Property Office, the Office of Communications and the Medicines and Healthcare Products Regulatory Agency have either developed AI-specific guidance or are in the process of producing it.

⁵⁷⁰ Department for Science, Innovation and Technology, *A Pro-innovation Approach to AI Regulation*, March 2023. See <<https://assets.publishing.service.gov.uk/media/64cb71a547915a00142a91c4/a-pro-innovation-approach-to-ai-regulation-amended-web-ready.pdf>>.

⁵⁷¹ Department for Science, Innovation and Technology, *A Pro-innovation approach to AI Regulation: Government Response*, 6 February 2024. See <<https://www.gov.uk/government/consultations/ai-regulation-a-pro-innovation-approach-policy-proposals/outcome/a-pro-innovation-approach-to-ai-regulation-government-response>>.

⁵⁷² *Ibid.*, p6.

⁵⁷³ *Ibid.*

⁵⁷⁴ *Ibid.*

⁵⁷⁵ *Ibid.*

⁵⁷⁶ White & Case, *AI Watch: Global Regulatory Tracker – United Kingdom*, 25 March 2025. See <<https://www.whitecase.com/insight-our-thinking/ai-watch-global-regulatory-tracker-united-kingdom>>.

Centralised support and resourcing

These sector-specific initiatives are supplemented by the development of other public policy initiatives including the establishment of the AI Security Institute (AISI), a research organisation, whose primary role is ‘to equip governments with the scientific understanding of the risks posed by advanced AI.’⁵⁷⁷ Established as a Directorate within the Department for Science, Innovation and Technology, the AISI has been purpose-built ‘as a startup in the government, combining the authority of government with the expertise and agility of the private sector.’⁵⁷⁸ Its work focuses on a range of projects that range from technical evaluations through AI safety evaluation principles, AI cyber security and the development of sandboxes for AI evaluation. As such it functions as a central source of technical skill and expertise that can inform and be applied by government and the regulators tasked to undertake the UK government’s regulatory program. This is supplemented by additional funding for regulators to develop new skills and capabilities:

The funding will enable regulators to collaborate to create, adapt, and improve practical tools to address AI risks and opportunities within and across their remits. It will enable regulators to carry out research and development to produce novel, actionable insights that will set the foundation of their approaches for years to come.⁵⁷⁹

Continuing concerns

However, although the express UK policy does not favour bespoke AI regulation, there are several initiatives underway that are legislative in nature, but which have not received passage through the UK parliament.

These reflect an ongoing concern within the UK that using existing regulatory settings may not be sufficient to provide a sufficient foundation for managing and mitigating AI risks and harms. The 2024 King’s Speech foreshadowed that the government would ‘seek to establish the appropriate legislation to place requirements on those working to develop the most powerful artificial intelligence models.’⁵⁸⁰ No government legislation has yet been introduced⁵⁸¹ to implement this objective. However, a private member’s Bill introduced in

⁵⁷⁷ AI Security Institute, *Rigorous AI Research to Enable Advanced AI Governance*. See <<https://www.aisi.gov.uk>>.

⁵⁷⁸ AI Security Institute, *Organisation*. See <<https://www.aisi.gov.uk/about>>.

⁵⁷⁹ Department for Science, Innovation and Technology, *A Pro-innovation approach to AI Regulation: Government Response*, 6 February 2024, Section 5, para 19(ii).

⁵⁸⁰ The King’s Speech 2024, 17 July 2024. See <<https://www.gov.uk/government/speeches/the-kings-speech-2024>>.

⁵⁸¹ As at 31 August 2025.

the House of Lords, the *Artificial Intelligence (Regulation) Bill* has progressed to its second reading.⁵⁸² Key feature of this Bill include:

- 1) Establishing and conferring functions on an AI Authority;⁵⁸³
- 2) Establishing five high-level regulatory principles applicable to the AI Authority;⁵⁸⁴
- 3) Requiring any business that develops, deploys or uses AI to be transparent, to test it thoroughly, and to comply with applicable laws, including data protection, privacy and IP;⁵⁸⁵ and equalities legislation;⁵⁸⁶ and
- 4) Introducing a principle of proportionality whereby any 'burden or restriction which is imposed ... should be proportionate to the benefits.'⁵⁸⁷

Although there have been no official UK government statements about how the King's Speech objectives are to be achieved, media commentary suggests that '[p]roposals to regulate artificial intelligence have been delayed by at least a year as UK ministers plan a bumper bill to regulate the technology and its use of copyright material.'⁵⁸⁸

Summary

The UK's embrace of an agile and iterative approach to regulation means that its response to AI is a work in progress, the precise contours of which will not be known until a later date. Its emphasis on using existing regulatory mechanisms has the benefit of providing industry and consumers with the continuity of existing and familiar regulatory schemes and their institutional oversight.

What is less clear is how the inevitable gaps between existing regulatory frameworks informed by non-legally binding principles will adequately safeguard against the new risks posed by AI. An obvious limitation is that five non-binding principles cannot override existing statutory provisions. There is no UK government policy material that addresses this issue.

⁵⁸² UK Parliament, Parliamentary Bill, *Artificial Intelligence (Regulation) Bill*, 4 March 2025. See <<https://bills.parliament.uk/publications/59353/documents/6094>>.

⁵⁸³ *Ibid*, clause 1.

⁵⁸⁴ *Ibid*, clause 2(1)(a).

⁵⁸⁵ *Ibid*, clause 2(1)(b).

⁵⁸⁶ *Ibid*, clause 2(1)(c).

⁵⁸⁷ *Ibid*, clause 2(1)(d).

⁵⁸⁸ Eleni Courea and Kiran Stacey, *UK ministers delay AI regulation amid plans for more 'comprehensive' bill*, The Guardian, 8 June 2025. See <<https://www.theguardian.com/technology/2025/jun/07/uk-ministers-delay-ai-regulation-amid-plans-for-more-comprehensive-bill>>.

What can Australia learn from these developments?

A high-level examination of international AI regulation yields a number of insights.

The US has been the international technology powerhouse and its approach to regulation has often served as a blueprint for that success, as other jurisdictions seek to emulate it. But its current federal policy settings stand on their own. The US has arguably moved from the market-based approach described by Anu Bradford to an anti-regulation stance albeit subject to a stated qualification to prevent “woke” AI. At a federal level, with this one exception, the US is committed to dismantling the minimal AI regulation it had developed in the period leading up to the election of the current administration.

That said, its retreat from regulation has left a vacuum that some US states are beginning to fill. A likely consequence is a patchwork of state-based US regulation that will inevitably lead to complexity and additional compliance costs for organisations that seek to commercialise their AI offerings in the US. This situation is not dissimilar to US data protection regulation where many US states have chosen to enact their own laws in the absence of comprehensive federal legislation.⁵⁸⁹ Just as the Brussels Effect has led to EU data protection law serving as a *de facto* global standard as international enterprises find it more practical to adopt higher standards than to modify their technology for compliance with the laws of individual jurisdictions, US state-based AI regulation could operate in a similar way for US-based organisations.

Both China and the EU have developed comprehensive, national AI regulatory schemes. While they differ in significantly different they share a number of policy preoccupations. For example, they share a commitment to the transparency and explainability and are explicitly designed to promote AI innovation and to mitigate its risks. They acknowledge the need for the development of AI that is safe and predictable and include a range of *ex ante* transparency and disclosure requirements. They seek to safeguard data protection and copyright and recognise cyber security risks. Disclosure of training data and requirements to ensure that it is fit for purpose are also featured. In broad terms, California’s regulatory framework also imposes a range of transparency requirements.

⁵⁸⁹ The International Association of Privacy Professionals (IAPP) reported that as at 7 July 2025, nineteen US states have passed data protection laws with active legislative proposals under consideration in another five states. See <https://iapp.org/media/pdf/resource_center/us_state_privacy_laws_report_2024_session_overview.pdf><https://iapp.org/pdf/resource_centre/State_Comp_Privacy_Law_Chart.pdf>.

Although the UK's regulatory posture appears to be in a state of transition, the five principles it has developed to guide responsible AI development and to inform the actions of existing regulators are not dissimilar to those embodied in, or which underpin, other international regulatory initiatives.

China's approach to AI regulation has not featured in Australian AI policy papers or discussions. Even in international discussions it is often overlooked. While the principles it seeks to enforce are not inconsistent with those adopted elsewhere, their implementation, interpretation and enforcement are the product of a legal system that is unlike those more familiar, at least to western sensibilities, common and civil law traditions, thus making them less immediately accessible and comprehensible.

However, it is important to acknowledge that China's regulatory system for AI has not impeded the development of a vibrant AI sector. Its success not only tends to undermine assertions that regulation and innovation are trade-offs that cannot coexist but could be considered to support the notion that well designed regulation provides a foundation for innovation by removing regulatory uncertainties.

Chapter Six – A suggested way forward for Australia

Approach

Existing Australian regulation applies to AI. That regulation represents prevailing and accepted policy settings. Although they may be imperfect, inadequate or contested in a pre-AI environment, they provide a foundation from which to launch an inquiry into their continued appropriateness in an AI environment.

Moving from that foundation involves ascertaining whether existing regulation adequately covers newly instantiated AI risks and, where it does not, identifying new regulatory interventions that address these new risks. However, before doing so there are some important intermediate steps that require consideration. These relate to the role and function of the law in an era of AI.

The first of these intermediate steps concerns the rule of law itself. The rule of law mandates that everyone is subject, and accountable, to the law. It ‘restricts the powers of government, corporations and individuals, and protects against the exercise of power without a lawful basis.’⁵⁹⁰ AI displays a range of characteristics, such as autonomy and its lack of transparency, contestability and accountability that tend to evade or sidestep the rule of law. This thesis argues that any Australian regulatory architecture for AI must ensure that the rule of law applies to AI.

The second intermediate step is a correlative of the rule of law and a consequence of it. I refer to this as the principle of equivalence. Equivalence concepts and principles are common and have featured prominently in internet-related regulation. The United Nations Commission on International Trade Law’s (UNCITRAL) Model Law on Electronic Commerce⁵⁹¹ serves as an example. In providing a model legal framework for electronic communications in a commercial context:

[t]he Model Law thus relies on a new approach, sometimes referred to as the “functional equivalent approach”, which is based on an analysis of the purposes and functions of the traditional paper-based requirement with a view to determining how those purposes or functions could be fulfilled through electronic-commerce techniques.⁵⁹²

⁵⁹⁰ Australian Constitution Centre, *Principle 2: The Rule of Law*, See <<https://www.australianconstitutioncentre.org.au/the-rule-of-law/>>.

⁵⁹¹ United Nations Commission on International Trade Law, Model Law on Electronic Commerce, 1996. See <https://uncitral.un.org/sites/uncitral.un.org/files/media-documents/uncitral/en/19-04970_ebook.pdf>.

⁵⁹² *Ibid*, p 20.

Although ‘functional equivalent’ does not completely account for the AI principle of equivalence I suggest, given that the latter extends to substantive equivalence, the concept is closely similar. Equivalence approaches also have a counterpart in ‘national treatment’ requirements in a range of international copyright treaties. For example, Article 5(1) of the *Berne Convention for the Protection of Literary and Artistic Works*⁵⁹³ (Berne Convention) requires Convention members to grant copyright protection to authors from other member countries on the same terms and having the same rights as they provide to their own citizens, thus also embodying an equivalence principle that has applied since the nineteenth century.

For AI, a principle of equivalence entails that where individuals, organisations and governments are required to observe laws and regulations, AI should do so such that those substantive obligations apply equally to it. To do otherwise privileges the artificial over the real and creates perverse incentives. For example, if a human is legally liable for providing negligent advice, AI should bear an equivalent liability. Equally, if a human is liable for traffic offences, those who are responsible for the AI that controls autonomous vehicles should bear equivalent responsibility. Without a principle of equivalence, there is an incentive to evade the law by increasingly using AI to undertake activities that exist, in effect, outside of the law. To ensure that the principle of equivalence is not avoided by powerful and well-resourced technology enterprises, contractual provisions that attempt to limit or negate it should be legally void.

Adopting this approach provides a regulatory compass that enables a more sustainable and consistent scaffolding from which to develop AI regulatory architecture and assists in adopting a coherent approach to AI regulation that is not reliant on an overarching regulatory scheme such as the EU AI Act.

The current position in Australia

Two different parts of the Australian government have produced, albeit incomplete, opposing approaches to AI regulation. The 2024 Proposals Paper recommends bespoke AI legislation that introduces guardrails for high-risk AI. The Productivity Commission rejects this approach, arguing that ‘AI-specific regulations should only be considered as a last

⁵⁹³ World Intellectual Property Organisation, *Berne Convention for the Protection of Literary and Artistic Works*, September 1886. See <<https://www.wipo.int/wipolex/en/text/283693>>.

resort⁵⁹⁴ for use cases ‘where existing regulatory frameworks cannot be sufficiently adapted to handle the issue’⁵⁹⁵ and ‘where technology-neutral regulations are not feasible.’⁵⁹⁶ In the meantime, a number of other government agencies have signalled that their approach to AI regulation within their portfolio responsibilities is informed by and remains based on the 2024 Proposals Paper.

Overall, the Commonwealth’s position appears to be shifting in favour of the Productivity Commission’s interim recommendations. The Minister now responsible for the Department of Industry, Science and Resources (DISR) is ‘working on a lighter touch model that will mostly adopt existing regulations including privacy and copyright, avoiding new red tape’⁵⁹⁷ and is:

resisting mounting calls for comprehensive new laws to address the risks of potentially “game-changing” AI technology’ and has “pushed back on the case for unified new AI architecture. “Australian law applies now. It’ll be treated no differently to the internet ... all of those technological platforms engage very similar issues, actually, and we’re going to deal with it in the same way.”⁵⁹⁸

As noted in Chapter four, a number of the regulatory problems associated with AI stem from unresolved issues about how to regulate the internet. Duplicating that approach is fraught with difficulties. At the same time, the Treasurer is reported as seeking a ““responsible middle path” between restraining AI and “letting it rip.””⁵⁹⁹

There is a clear shift in government attitudes to AI regulation in Australia. While the exact nature and extent of this remains unclear,⁶⁰⁰ a move away from developing a bespoke regulatory architecture for AI in favour of adapting existing laws and regulatory frameworks is clearly under way. More specifically, the Australian position appears to signal a move away from EU-type AI Act regulation towards the position adopted in the UK, although the political and policy discourse has not yet articulated this labelling.

In the meantime, however

⁵⁹⁴ Productivity Commission, *Harnessing Data and Digital Technology: Interim Report*, August 2025, p 20.

⁵⁹⁵ Ibid.

⁵⁹⁶ Ibid.

⁵⁹⁷ Paul Sakkal, *Split in Albanese’s Caucus as Government Moves to Kill AI Laws*, The Age, 28 August 2025. See

<<https://www.theage.com.au/politics/federal/split-in-albanese-s-caucus-as-government-moves-to-kill-ai-laws-20250808-p5mlic.html>>.

⁵⁹⁸ The Saturday Paper, ‘Labor Shrinks from AI Rules’, August 16-22, 2025. See

<<https://www.thesaturdaypaper.com.au/news/politics/2025/08/16/labor-shrinks-ai-rules>>.

⁵⁹⁹ Ibid.

⁶⁰⁰ As at 31 August 2025.

The pace of reform in Australia in responding to the challenges of generative AI is slow. Uncertainty about the legislative response to concerns about generative AI might be said to stifle innovation and impede the adoption of useful technology. These concerns should not be exaggerated.⁶⁰¹

Australia's delay in developing its own approach to AI regulation in the face of its increasing adoption across many sectors of the economy is not conducive to engendering regulatory certainty and trust.

AI regulation: first principles

Any proposal for an Australian approach to regulating AI should be anchored in a clear understanding of what regulation is designed to achieve. Following standard approaches to public policy development, the next step is to closely scrutinise its characteristics and then assess what legal instruments are appropriate to address those regulatory objectives.

The Productivity Commission argues that:

applied AI can *exacerbate* existing risk of harm but does not create *wholly new* risks where none existed before. Indeed, issues like the environmental pressures posed by high energy demands when training and running AI models, or even the potential issue of copyright infringement, are not unique to AI and continue trends seen with previous phases of the information technology revolution ... The regulatory response should focus on mitigating the *additional risk* of these harms occurring, rather than treating them as wholly new problems.⁶⁰²

It also argues that these additional risks should be identified and addressed by reference to:

a comprehensive set of 'gap analyses' to understand the risks stemming from AI and whether these risks can be dealt with under existing regulatory frameworks ... These gap analyses into the additional risk posed by AI should be completed with urgency, and until they are, consideration of new, economy wide regulations before the gaps in the existing regulatory framework are identified creates the risk of duplication and unnecessarily high compliance burden with little benefit to end consumers.⁶⁰³

This approach is broadly consistent with the approach this thesis suggests. However, it does so for reasons that are significantly different to those proposed by the Productivity Commission. Many of the risks associated with AI are already well known. Equally, legal and regulatory inadequacies are readily apparent. Some existing risks are poorly covered by current regulation making it impractical to merely focus, as the Productivity Commission suggest, on additional risks.

⁶⁰¹ Jeannie Marie Paterson, *Regulating Generative AI in Australia: Challenges of Regulatory Design and Regulator Capacity* in Philipp Hacker et al (eds), *The Oxford Handbook of the Foundations and Regulation of Generative AI* (Oxford University Press, 1st ed, 2025) p 12.

⁶⁰² Productivity Commission, (above, n 594), p 15.

⁶⁰³ *ibid*, p 18.

Much of the AI regulatory debate focuses on individual harms, such as deep fakes, various types of hallucinations and children's online safety and how laws might be developed to address these. Although these are important issues, as noted in Chapter Four, a patchwork risk-by-risk, issue-by-issue regulatory approach necessarily means that regulation will always lag risk by significant margins. It also fails to learn from the lessons of failed attempts to regulate the internet discussed in the same chapter and will lead to a prolonged policy debate as each contentious regulatory initiative is debated, inquired into and eventually resolved.

This constant cycle of piece-by-piece regulatory change is not conducive to producing the regulatory certainty that is important to encouraging investment in AI and its broader use within the community. It is an issue that does not appear to have received the attention it deserves from the Productivity Commission. An important element of providing these regulatory incentives is to create an environment of certainty and trust: 'regulation is about creating the conditions of certainty and trust, and the management of risk ... While too much certainty can inhibit innovation and restrict competition, too much uncertainty can inhibit investment and creativity.'⁶⁰⁴ Innovation and regulation are not inherently mutually exclusive binary choices. However, badly designed and implemented regulation is likely to produce adverse outcomes.

Taking account of these observations about the role of function of regulation in conjunction with rule of law and equivalence principles, the next step is to understand how existing regulation of the foundational risks associated with AI identified in Chapter Three can be adapted to an AI environment.

Data protection

Protecting personal data in an environment where machine learning AI relies on it for training purposes is one of the most intractable problems for AI regulation. There is a clear interest in ensuring that this type of second-wave AI is as accurate and effective as possible. But this requires it to process vast quantities of personal data the use of which is overwhelmingly undertaken without the consent of the individuals concerned, or otherwise on legal foundations that are questionable. In Australia, in particular, the small business exemption in the Privacy Act means that it fails to protect significant amounts of personal

⁶⁰⁴ Arie Freiberg, (above, n 9), p 642.

information at all. When combined with adhesion terms and conditions for digital platforms and other internet-based services and ‘take it or leave it’ privacy policies, individuals are left with rights that are not practical or manageable at an individual level.

The effectiveness of data protection law rests on assumptions that are no longer valid. Two of the main assumptions are linked and involve first, the conferral of rights on individuals that are intended to give a measure of control over their personal data that can be enforced against those who use it and secondly that individuals have sufficient agency to exercise these rights.

A main impetus for rights involves a desire to address the problem that individuals lack much power in their relationships with the gigantic organizations that have massive digital dossiers of their personal data ... Privacy laws were developed with the aim of putting individuals back in control of their personal data – and providing for individual rights was an essential way to do so.⁶⁰⁵

Together these are known as ‘privacy self-management.’ A third assumption is that personal data is located at the level of the individual. This assumption is now also false as AI can generate personal data through aggregating personal data sources and creating inferences and analysis that result in it creating new personal data. The individual may not know that this personal data exists, where it exists or the nature of its content.

Solove advances ‘three reasons why rights are not practical for individuals to exercise.’⁶⁰⁶ He argues that ‘[r]ights put too much of the onus on individuals to fight a war they can’t win ... (using) privacy rights as a primary way to protect privacy is akin to arming an individual to with a dagger to fight an entire army.’⁶⁰⁷ He identifies privacy self-management as being ineffective as ‘[i]ndividuals lack the expertise to understand and assess the risks. Even experts lack the knowledge about how the data will be used in the future and how algorithms will reach decisions regarding the data.’⁶⁰⁸ Privacy self-management requires individuals to undertake an ‘endless burden of chores.’⁶⁰⁹ Finally, he contends that privacy:

can’t be protected at the level of the atomistic individual. Individuals make privacy choices that have effects not just for themselves but for many others ... sharing one’s genetic data also shares the genetic data of one’s family members. In today’s world of machine learning,

⁶⁰⁵ Daniel Solove, *The Limitations of Privacy Rights*, (2023) 98(3) Notre Dame L. Rev., 975, 977.

⁶⁰⁶ *Ibid*, p 978.

⁶⁰⁷ *Ibid*.

⁶⁰⁸ *Ibid*.

⁶⁰⁹ *Ibid*, p 985.

the personal data of everyone in a data set has an impact on the decisions that the system makes.⁶¹⁰

While Solove offers a powerful critique of the limitations of data protection law, he is unable to provide a solution other than to identify that while '[r]ights should certainly be part of privacy laws ... they can only play a small supportive role. Meaningful protection must be large-scale and structural in nature.'⁶¹¹

Clearly data protection law is failing in an AI environment. However, there has to date been no meaningful reconceptualisation of it that could lead to the large-scale and structural regulatory innovation that is necessary to address the new issues posed by AI. This leaves the options limited to more short-term regulatory initiatives that attempt to alleviate the difficulties that individuals face in exercising data protection rights in a meaningful way. These reforms could usefully address the problems associated with data protection law's reliance upon privacy self-management and providing alternate options so that individual rights can be asserted collectively.

One potential approach is to enable individuals to directly enforce the rights conferred on them under the Privacy Act and the APPs. This type of direct right of action is a solution that has previously been considered by the ACCC in its *Digital Platforms Inquiry* where it recommended that individuals be given 'a direct right of action to bring actions and class actions against APP entities in court to seek compensation for an interference with their privacy under the Privacy Act.'⁶¹² The ACCC's recommendations were subsequently canvassed in the Attorney-General's Department *Privacy Act Review Discussion Paper*,⁶¹³ which proposed a direct action model where any individual or group could initiate proceedings for an interference with privacy in the Federal Court or the Federal Circuit and Family Court after first having undertaken and exhausted the OAIC's conciliation process. This model was subsequently considered in the Attorney-General Department's *Privacy Act Review Report*⁶¹⁴ which made a recommendation to '[a]mend the Act to allow for a direct right of action in order to permit individuals to apply to the courts for relief in relation to an

⁶¹⁰ Ibid, p 978-979.

⁶¹¹ Ibid, p 1035.

⁶¹² Australian Competition and Consumer Commission, *Digital Platforms Inquiry*, June 2019. Recommendation 16(e), p 35.

⁶¹³ Attorney-General's Department, *Privacy Act Review Discussion Paper*, October 2021. See <https://consultations.ag.gov.au/rights-and-protections/privacy-act-review-discussion-paper/user_uploads/privacy-act-review-discussion-paper.pdf>.

⁶¹⁴ Attorney-General's Department, *Privacy Act Review Report 2022*. See <https://www.ag.gov.au/sites/default/files/2023-02/privacy-act-review-report_0.pdf>.

interference with privacy.⁶¹⁵ The Australian government agreed in-principle to this recommendation,⁶¹⁶ stating that ‘individuals should have more direct access to the courts to seek remedies for breaches of the Act through a direct right of action ... Such a right would be an important measure to enhance individuals’ control over their personal information.’⁶¹⁷

Although a first tranche of Privacy Act amendments based on these reviews was enacted in 2024 and introduced, for the first time in Australia, a cause of action in tort for serious invasions of privacy,⁶¹⁸ they did not include a direct right of action for interference with privacy under the APPs. Second tranche reforms are expected, but it is unclear when they will be introduced or if they will include a direct right of action.

Enhanced information privacy protection through a direct right of action exercised through class action litigation mechanisms would allow individual infringements of rights to be approached collectively, thus removing one of the key impediments to the enforcement of data protection rights. Collective, class action-based data protection rights enforcement would also reduce the burdens and resource limitations faced by the OAIC and, through litigation funding for class actions, would tend to address the power and resource imbalances faced by individual rights holders against well-resourced international technology organisations such as the major digital platforms and AI companies. Enforcement of compensation rights provided those affected by the Robodebt programme with redress through class actions against the Commonwealth and is an effective way in which to pursue collective rights through the courts.⁶¹⁹ Although it is conceded that this approach does not address the fundamental shortcomings of Australian data protection law, such as the small business exemption which government has also agreed in-principle to remove⁶²⁰ and concerns regarding enforcement of Australian privacy law where interferences with privacy occur offshore, it is arguably the only practical way forward until new approaches to data protection are developed.

⁶¹⁵ Ibid, Recommendation 26.1, p 279.

⁶¹⁶ Australian Government, *Government Response: Privacy Act Review Report*, 28 September 2023. See <<https://www.ag.gov.au/sites/default/files/2023-09/government-response-privacy-act-review-report.PDF>>.

⁶¹⁷ Ibid, p 18.

⁶¹⁸ The *Privacy and Other Legislation Amendment Act 2024 (Cth)*

⁶¹⁹ See <<https://gordonlegal.com.au/services/class-actions/robodebt-class-action-appeal/>>.

⁶²⁰ Above (n 616) p 6.

Copyright

Issues associated with the application of copyright law in respect of, in particular, the use of works and other subject matter for training data can also be addressed using collective, rather than individually based rights management approaches. Unlike data protection law, copyright law has long established, existing mechanisms that enable end users to access and use large bodies of copyright materials in circumstances where identifying and negotiating with individual copyright owners is difficult, impractical or impossible. These mechanisms involve either voluntary or compulsory licensing schemes and, in Australia, are established under Part V Division 5 of the Copyright Act.

Copyright collective licensing schemes have operated successfully internationally and in Australia for decades.⁶²¹ They serve the dual purpose of ensuring that copyright owners are compensated for the use of their works and that end users are provided with a convenient way in which to licence large quantities of works without being required to identify individual copyright owners and negotiate with each of them, thus reducing complex administration and higher transaction costs, particularly for copyright works and other subject matter owned by individuals or organisations located outside Australia.

The most established of these is the worldwide network of performing right societies that collectively license performance, communication (and sometimes mechanical⁶²²) rights in musical works and sound recordings to end users. These voluntary licensing arrangements cover a range of venues, including restaurants, retail premises and broadcasters. In Australia, APRA AMCOS, the copyright society that administers these licensing arrangements, describes itself as a 'music rights management organisation'⁶²³ that pays 'royalties to music creators when their music is played or copied, both locally and overseas.'⁶²⁴ Through reciprocal arrangements it represents the musical copyright catalogues and repertoires of copyright owners who are members of like collecting societies around the world, such as the

⁶²¹ The Australasian Performing Rights Association (APRA) was established in 1926.

⁶²² The mechanical right is the right of the owner of copyright in a musical work (and any accompanying literary work) to authorise its reproduction in the form of a sound recording.

⁶²³ See <<https://www.apraamcos.com.au/about/what-we-do>>.

⁶²⁴ Ibid.

UK's Performing Right Society, PRS,⁶²⁵ equivalent societies in the US such as ASCAP⁶²⁶ and BMI⁶²⁷ and many other similar societies across the world.⁶²⁸

Another example of collective licensing arrangements is the Copyright Agency Ltd (CAL), 'a not-for-profit organisation that provides simple licensing solutions to allow (the) use of copyright-protected words and images.'⁶²⁹ CAL administers the Statutory Licence Scheme for Education,⁶³⁰ the Statutory Licence Scheme for the Commonwealth, State and Territory governments,⁶³¹ the Artists Resale Royalty Scheme⁶³² as well as visual arts licensing⁶³³ as well as a variety of commercial licensing arrangements for business.⁶³⁴ CAL has undertaken collective licensing activities since 1974.⁶³⁵

Voluntary copyright licensing is subject to statutory checks and balances, primarily through the Copyright Tribunal of Australia (Copyright Tribunal) established under s 138 of the Copyright Act. Copyright collecting societies must also adhere to the voluntary *Code of Conduct for Copyright Collecting Societies*.⁶³⁶ Importantly, the Copyright Tribunal has oversight over collective licensing arrangements under references from either a licensor or licensee, including licence terms, licence charges and distribution arrangements⁶³⁷ as well as 'own motion' powers to refer questions of law to the Federal Court of Australia.⁶³⁸

Collective voluntary licensing of copyright works and other subject matter used for training AI provides an effective balance between the competing interests of AI developers needing efficient access to large volumes of copyright material and the interests of creators to be compensated for the use of their works and other subject matter. It also supports equity by ensuring that AI developers do not 'free ride' at the expense of copyright owners by obtaining the essential raw material for their technology at no cost to themselves but to the significant disadvantage of creators. As a voluntary scheme, certain copyright owners may

⁶²⁵ See <<https://www.prsformusic.com/what-we-do>>.

⁶²⁶ See <<https://www.ascap.com>>.

⁶²⁷ See <<https://www.bmi.com>>.

⁶²⁸ A list of international music copyright collecting societies can be found at <<https://www.apraamcos.com.au/about/cost-to-revenue-ratio>>.

⁶²⁹ Copyright Agency Limited. See <<https://www.copyright.com.au>>.

⁶³⁰ See <<https://www.copyright.com.au/licences-permission/educational-licences/>>.

⁶³¹ See <<https://www.copyright.com.au/licences-permission/government-licences/>>.

⁶³² See <<https://www.resaleroyalty.org.au>>

⁶³³ See <<https://www.copyright.com.au/licences-permission/visual-art-licences/>>

⁶³⁴ See <<https://www.copyright.com.au/licences-permission/corporate-licences/>>

⁶³⁵ See <<https://www.copyright.com.au/about-us/our-heritage/>>

⁶³⁶ *Code of Conduct for Copyright Collecting Societies*. See <<https://www.copyrightcodeofconduct.org.au>>

⁶³⁷ See Part VI, Division 3, Copyright Act.

⁶³⁸ s 161 Copyright Act.

choose to opt out of any voluntary licensing arrangement if they chose to do so, such as where large copyright owners decide to undertake their own direct licensing arrangements, for example the arrangements that News Corporation has made with Open AI to permit use of its copyright catalogue to train ChatGPT.⁶³⁹

In short, copyright issues associated with AI can be addressed within existing, established voluntary licensing frameworks. It follows that text and data mining exemptions from copyright are unnecessary and, equally importantly, inequitable and unfair.

AI liability

As discussed earlier existing forms of legal liability have adequately addressed liability issues arising out of first-wave AI. For fault-based liability for second-wave AI, there are significant problems in applying negligence law doctrines. These are attributable to the inherent characteristics of that form of AI such as its lack of transparency and explainability as well as with difficulties in tracing causation issues across the AI value chain.

To address these issues this thesis suggests that where second-wave AI is concerned several regulatory changes are essential. These changes would only apply to AI that falls within the OECD's definition of an 'AI system.'

I argue that the principles of strict liability should apply to all AI systems as a default regulatory position. However, to avoid strict liability those responsible for AI systems could do so by satisfying transparency and explainability requirements that would be legally defined, drawing on equivalent measures in the EU AI Act and China's AI regulatory system. Implementing this regime would require establishing an institutional capacity under which a purpose-built, statute-based organisation is given functions and powers to receive and hold the required transparency and explainability material and to make it available to litigants who claim that an AI system has caused loss or damage. Compliance with these requirements would constitute a 'safe harbour' where strict liability is replaced by fault-based liability. Standard negligence law doctrines would then apply within the safe harbour.

⁶³⁹ T.J. Thomson and James Meese, *What OpenAI's Deal with News Corp Means for Journalism (and for you)*, The Conversation, 24 May 2024. See < <https://theconversation.com/what-openais-deal-with-news-corp-means-for-journalism-and-for-you-230773>>.

Strict liability as a default

The common law doctrine of strict liability pre-dates fault-based approaches to liability. In a high technology AI environment it is easy to overlook the fact that the common law developed solutions to questions of liability for the actions of things that functioned autonomously, inexplicably and non-transparently as early as the fourteenth century: the law of livestock.

Strict liability is a mode of responsibility that does not require proof of fault. It is primarily concerned with issues of causation and was the general approach of the ancient common law to address harms caused by one individual to another:

the early common law, preoccupied with breaches of the peace displayed little concern about the moral responsibility of 'fault' of the wrongdoer ... primitive law imposed liability for causation rather than fault.⁶⁴⁰

Early forms of strict liability included causes of action such as trespass to the person⁶⁴¹ and the law of livestock which, as early as the 1300s addressed regulating things whose activities and actions are unpredictable and unknowable:

By the fourteenth century it was well established that if stock strayed into neighbouring land *their owner was held strictly liable*. At first liability existed only for damage caused to the surface of the land and to crops. Later it was extended to injury to the plaintiff's stock, whether directly by attack or indirectly by infection or misbreeding. Later again it was extended to injury to the plaintiff himself.⁶⁴²

The early development of the common law 'imposed liability for causation rather than fault'⁶⁴³ across a range of tortious conduct based on two forms of action – trespass and case. The main difference between the two forms of action were:

Trespass was actionable per se, that is, without proof of actual damage whereas generally damage was the gist of a case ... trespassory harm was prima facie wrongful and it was for the defendant to raise any issue of justification or excuse, such as self-defence or unavoidable necessity. In contrast, the action on the case required proof by the plaintiff of either wrongful intent or negligence on the part of the defendant.⁶⁴⁴

Although strict liability torts waned in their scope and popularity in favour of fault-based causes of action, such as negligence and other torts that involved a mental element,

⁶⁴⁰ Fleming, John G, *Fleming's the Law of Torts*, ed C Sappideen and Prue Vines (Thomson Reuters (Professional) Australia Limited: Lawbook Co., 11th edition, 2024) p 379.

⁶⁴¹ For example, see *Venning v Chin* (1974) 10 SASR 299.

⁶⁴² John Toohey, *Liability for Straying Stock*, (1966) 7(4) University of Western Australia Law Review, 490 at p 491. My emphasis.

⁶⁴³ Fleming, John G, (above n 640) p 379.

⁶⁴⁴ *Ibid*, p24.

particularly after the enactment of the *Judicature Act 1873*⁶⁴⁵ and its replication across most of the common law world, ‘the impetus of this movement towards fault-based liability’⁶⁴⁶ diminished during the latter part of the twentieth century. In policy terms this movement reflects a desire to shift the cost of harmful activity from victims and society as a whole to ‘those benefitting from the accident-producing activity.’⁶⁴⁷ Consistent with this rationale there are now a range of laws that impose strict or absolute liability.

The *Criminal Code Act 1995 (Cth)* (CCA) defines ‘strict liability’ and offences that are created under Commonwealth legislation. By s 6.1:

- (1) If a law that creates an offence provides that the offence is an offence of strict liability:
 - (a) there are no fault elements for any of the physical elements of the offence; and
 - (b) the defence of mistake of fact under section 9.2 is available
- (2) If a law that creates an offence provides that strict liability applies to a particular physical element of the offence:
 - (a) there are no fault elements for that physical element; and
 - (b) the defence of mistake of fact under section 9.2 is available in relation to that physical element.

Section 6.1 of the CCA ‘implicitly rejects attempts to rationalise strict liability as a form of liability for negligence..The defining features of strict liability are the absence of any requirement of fault, whether for all or some of the physical elements of an offence, coupled with the provision of the defence of reasonable mistake of fact.’⁶⁴⁸ Although this formulation has been developed in the context of the criminal law it also accurately describes the position for strict liability torts.

Strict liability applies across a range of subject matters. It is ‘a common feature of regulatory frameworks underpinning corporate and prudential regulation,’⁶⁴⁹ as well as environmental

⁶⁴⁵ *Supreme Court of Judicature Act 1873*, 36 & 37 Vict., Ch 66.

⁶⁴⁶ Fleming, John G, (above n 640) p 380.

⁶⁴⁷ *Ibid.*

⁶⁴⁸ Australian Government, Attorney-General’s Department, *Commonwealth Criminal Code: Guide for Practitioners*, March 2002. See <<https://www.ag.gov.au/crime/publications/commonwealth-criminal-code-guide-practitioners>>

⁶⁴⁹ Australian Government, Australian Law Reform Commission, *Traditional Rights and Freedoms-Encroachments By Commonwealth laws*, (2016) ALRC report 129, para 10.36. See, for example, s 1308B(1)(b), *Corporations Act 2001 (Cth)* which establishes a strict liability offence where a corporation advertises or publishes a statement of the amount of its capital that is misleading and s 1308(3), *Corporations Act 2001 (Cth)* which establishes a strict liability offence where a person makes or authorises the making of a statement in a document required under the Act and the document is materially false or misleading and the person did not take all reasonable steps to ensure it was not materially false or misleading. Prudential regulation strict liability offences include s29JA(3) *Superannuation Industry (Supervision) Act 1993 (Cth)* where it is an offence where a Registered Superannuation Entity (RSE) fails to notify the Australian Prudential Regulation

protection regulation,⁶⁵⁰ workplace health and safety laws,⁶⁵¹ customs and border protection,⁶⁵² copyright⁶⁵³ and aviation.⁶⁵⁴ Australian defamation law is strict liability-based and the Australian Consumer Law also embodies strict liability for some categories of consumer products and services protections.⁶⁵⁵ Strict liability laws often provide for flexible remedies such as civil penalties and can also involve criminal penalties.⁶⁵⁶

There is thus nothing unusual in imposing civil strict liability principles to activities that entail significant harm or where it is equitable to ensure that responsibility for harm is borne by those that benefit from the activity rather than requiring others, or society in general, to bear the cost. Those responsible for AI systems are uniquely placed to ensure that they are safe and should accordingly bear the responsibility for the loss and damage they cause on a strict liability basis.

What is a safe harbour?

In legal terms, a safe harbour is a legislative provision or scheme that provides relief or protection from liability if certain requirements are met. In Australia, examples of safe harbour schemes can be found in the *Corporations Act 2001 (Cth)* (Corporations Act) and the Copyright Act.

The Corporations Act safe harbour is designed to protect company directors from personal civil and criminal liability for insolvent trading where certain conditions are met. Under s 588G of the Corporations Act a duty is imposed on company directors to prevent insolvent trading. However, if:

at a particular time after the person starts to suspect the company may become or be insolvent, the person starts developing one or more courses of action that are reasonably likely to lead to a better outcome for the company⁶⁵⁷

Authority (APRA) of a significant breach of an RSE licence.

⁶⁵⁰ For example, s 303EK, *Environment Protection and Biodiversity Act 1999 (Cth)* which creates an offence of importing into Australia a regulated live specimen without a permit; engaging in commercial fishing in the Australian Fishing Zone without a permit under s 95(1)(a) *Fisheries Management Act 1991(Cth)*

⁶⁵¹ For example, breaches of general employer duties under s 21, *Occupational Health and Safety Act 2004 (Vic)*

⁶⁵² For example, an extensive range of offences under Division 4, *Customs Act 1901 (Cth)*.

⁶⁵³ For example, s 132AD(5) *Copyright Act 1968 (Cth)* which creates a strict liability offence where a person makes an article in preparation for or in the course of selling it, letting it for hire, or obtaining a commercial advantage or profit and the article is an infringing copy of a work or other subject-matter and copyright subsists in it.

⁶⁵⁴ See s 11 *Damage by Aircraft Act 1999 (Cth)*. Part 2 Divisions 2 and 3 of the *Crimes (Aviation) Act 1991 (Cth)* also create a range of absolute liability offences in relation to certain conduct on board or affecting the safe operation of aircraft.

⁶⁵⁵ These laws also provide for a range of defences but their onus of proof must be discharged by the defendant. Chapter 4 Division 1 of the *Competition and Consumer Act 2010 (Cth)* creates a number of strict liability offences relating to unfair practices. For example, s 151(1) prohibits a range of false or misleading practices relating to goods and services. By virtue of s151(4), these are strict liability offences.

⁶⁵⁶ For example, Part V, Division 5 of the Copyright Act includes criminal offences for certain categories of copyright infringement.

⁶⁵⁷ S 588GA(1)(a), *Corporations Act 2001 (Cth)*.

and certain other requirements are met, such as continuing to pay employee entitlements and tax reporting compliance, insolvent trading obligations are attenuated. The burden of proving that the safe harbour requirements have been met lies on the person who seeks to rely on it.⁶⁵⁸

The Copyright Act safe harbour provisions are contained in Part V, Division 2AA. Essentially these protect a range of ‘service providers’ such as carriage service providers,⁶⁵⁹ organisations that assist disabled persons,⁶⁶⁰ and archives, cultural and educational institutions from liability for copyright infringement if they establish systems to respond to take-down notices and to terminate accounts of repeated infringers. The copyright safe harbour constitutes a defence to an infringement action. Under s 116AI of the Copyright Act, there is a lower standard of proof than is required under the Corporations Act safe harbour:

If a service provider, in an action relating to this Division, *points to evidence*, as prescribed, that suggests that the service provider has complied with a condition, the court must presume, in the absence of evidence to the contrary, that the service provider has complied with the condition.⁶⁶¹

An AI safe harbour

This thesis proposes that an AI safe harbour should be conceptually informed by those adopted in the Corporations Act and Copyright Act but adapted to align with the particular requirements needed to provide for appropriate AI transparency and accountability and, as far as possible, to align Australian transparency approaches with international benchmarks such as the Chinese and EU regulatory approaches to transparency for second-wave AI.

An Australian approach would consist of the enactment of legislation to first, establish, and confer functions and powers on, an Office of AI and secondly to establish the parameters of the safe harbour scheme.

In broad terms the safe harbour would enable AI developers, providers, deployers or suppliers (AI Provider) to lodge with the AI Office material they considered explained and made transparent the operations of an AI system and to evidence how it produced an outcome in any particular instance. There would be no obligation to lodge this material: participation in the safe harbour scheme would be entirely voluntary.

⁶⁵⁸ S 588GA(3), *Corporations Act 2001* (Cth).

⁶⁵⁹ S 116ABA(1)(a) *Copyright Act 1968*.

⁶⁶⁰ S 116ABA(1)(b) *Copyright Act 1968*.

⁶⁶¹ My emphasis.

The Office of AI would function as a repository of all the material provided by an AI Provider necessary to enable AI systems to be scrutinised, explained and made transparent. It would be empowered to receive that material from either public or private sector organisations. Drawing on EU and Chinese regulatory approaches this would include details of algorithms used and a copy of the AI system's training data. Other material, drawing on EU approaches, that could be lodged could include risk assessments, risk management plans, impact assessments and technical details pertaining to the development and operation of the AI system: in other words, any material that the AI provider or deployer considered that would satisfy transparency and explainability requirements and that would support its contention that it was entitled to the benefit of the safe harbour. The Office of AI's role would not be to define the nature of the transparency materials that could be provided to it. Apart from holding the material lodged with it, its role would also be, similar to EU and Chinese regulation, to establish and maintain a public-facing register of transparency material in summary form.

Where a claim for liability for an AI system is made the AI Office would be empowered to provide it to courts or tribunals as a normal part of litigation discovery or subpoena procedures. Once given to a court or tribunal the defendant AI developer or provider would bear the onus of proving that the material it had lodged with the Office of AI provided a sufficient basis on which its AI system could be sufficiently explained. On discharging this onus, it would be entitled to conduct its defence in the same way as any standard negligence action.

The AI Office would have no role in certifying or approving the adequacy or otherwise of any material lodged with it. That function would be vested in the relevant court or tribunal to determine, on a case-by-case basis, whether the sufficiency of the material met safe harbour transparency requirements. As the AI provider's material would only be disclosed to a litigant in the context of court proceedings, the *Harman principle*⁶⁶² would apply, restricting a plaintiff from using any such material produced in proceedings for any other purpose, and thus addressing any AI provider confidentiality or trade secret concerns.

An AI Office could also conveniently function as an institutional source of research and capacity building in a manner similar to the capacity established in the UK by functioning as

⁶⁶² *Harman v Secretary of State for the Home Department* [1983] 1 AC 280. See also *Hearne v Street* (2008) 235 CLR 125.

source of expertise across the public and private sectors. It is unreasonable to expect that all actors who wish to use AI have access to appropriate and up-to-date specialised information. In particular, '[r]egulators need expertise and resources to make their statements of expectations creditable'⁶⁶³ and it is economically inefficient for each potential regulator to establish and maintain a specialised AI research capability.

Aside from establishing the AI Office, the main other regulatory reform would involve legislation that established the safe harbour scheme, including adopting the OECD definition of an AI system and specifying that the deployment or supply of the AI system was governed by strict liability principles unless (a) the provider or supplier lodged with the AI Office material that it considered provided an adequate explanation of its functions; and (b) that the provider or supplier proved to a court or tribunal on the balance of probabilities that transparency requirements had been met by the material lodged with the AI Office.

A useful basis for the transparency requirements, drawing on existing Australian policy work, would be the Department of Industry, Science and Resources' *Voluntary AI Safety Standard*⁶⁶⁴ which consists of ten 'guardrails:'

- 1) Establish, implement and publish an accountability process including governance, internal capability and a strategy for regulatory compliance.
- 2) Establish and implement a risk management process to identify and mitigate risks.
- 3) Protect AI systems, and implement data governance measures to manage data quality and provenance.
- 4) Test AI models and systems to evaluate model performance and monitor the system once deployed.
- 5) Enable human control or intervention in an AI system to achieve meaningful human oversight.
- 6) Inform end-users regarding AI-enabled decisions, interactions with AI and AI-generated content.
- 7) Establish processes for people impacted by AI systems to challenge use or outcomes.

⁶⁶³ Jeannie Marie Paterson, (above, n 601) p 5.

⁶⁶⁴ Department of Industry, Science and Resources, *Voluntary AI Safety Standard*, August 2024. See <<https://www.industry.gov.au/sites/default/files/2024-09/voluntary-ai-safety-standard.pdf>>.

- 8) Be transparent with other organisations across the AI supply chain about data, models and systems to help them effectively address risks.
- 9) Keep and maintain records to allow third parties to assess compliance with guardrails.
- 10) Engage your stakeholders and evaluate their needs and circumstances, with a focus on safety, diversity, inclusion and fairness.

These 'guardrails' would constitute a non-exhaustive list of relevant transparency requirements that a court would take into account in determining whether safe harbour transparency requirements are met in any particular case.

Establishing a safe harbour that enables fault-based liability serves a variety of connected policy purposes. It provides a powerful incentive for AI providers to address transparency and explainability issues. It provides end users with a level of assurance that AI systems they are using have been designed safely by being able to access a public facing register. It could also provide individuals with a means to establish whether their personal information appears in training data utilising existing access mechanisms established under the Privacy Act, APP 12. It could constitute the evidence base for the calculation and distribution of copyright royalties under collective licensing schemes. It would subject unexplainable black box AI to strict liability.

Indirect Effects

Although the regulatory architecture described in this Chapter does not directly address AI risks such as deep fakes, discrimination, identity theft or online hate speech, the regulatory impact of a default strict liability is likely to encourage AI developers and providers to give greater consideration to the impacts of their technology over all of its potential applications. For example, if a claimant can establish causation between an AI system and rejection for rental accommodation on grounds prohibited by discrimination legislation and no safe harbour can be established, this enhanced level of accountability and liability will necessarily lead to either better and safer AI that complies with the law or the abandonment of the AI system in question.

International Alignment

Although the safe harbour approach does not explicitly adopt either EU AI Act or Chinese regulatory transparency approaches, it enables any Australian AI provider to develop the transparency material that is required if it wished to supply or commercialise its AI system in either of those markets.

Conclusion

As outlined in Chapter One, the aim of this thesis is to elucidate an optimal approach to the design of a regulatory regime for AI in Australia¹. In so doing it adopts an approach based on the principle of rule of law and the idea of equivalence between human and artificial intelligence as a foundation for a legal and regulatory investigation into AI. The thesis has been structured to address 4 specific sub-questions.

- 1) What are the specific features of current AI and the foundational risks and harms flowing from the use of AI that need to be appreciated in designing an appropriate regulatory framework for Australia?
- 2) What considerations should be taken into account in developing a regulatory framework for Australia, having regard to the broader context of prior information, communications and technology regulation and key international developments in AI regulation and governance?
- 3) What insights are available from key international developments in relation to the regulation and governance of AI?
- 4) How might it be possible to address the risks and harms posed by AI without enacting a sui generis risk based regulatory regime as originally proposed for Australia?

Chapters Two and Three deals with question (1). Chapter Two identifies the key features of AI and argues for the adoption of a broad definition that is inclusive of both first-wave and later iterations. What is relevant for the purposes of this thesis is that AI is used to computationally generate outputs that typically require human intelligence to produce and it, to varying degrees, contains levels of complexity and opacity that make it difficult to comprehend how these outputs are generated. Chapter Three then identifies the key foundational risks and harms presented by AI and the challenges they create. These arise in part from the gathering of the vast troves of data used for processing and training and in

part from deficiencies in AI outputs and the manner in which they are implemented. It focusses on the 5 key areas of law where these issues arise: copyright law, data protection law, administrative law, discrimination law and negligence.

Chapter Four answers question (2) and contextualised these issues with reference to regulation of the internet more generally. It points out that many of the regulatory dilemmas raised by AI are a residue of unresolved issues in respect the internet more broadly, and the arguably problematic adoption of a laissez faire approach based on the rationale that regulation impedes innovation

Chapter Five then answers question (3) and provides an overview of overseas' developments relating to the regulation of AI, focussing on those in the US, EU and China and highlighting the current state of play and salient features of relevance to Australian regulators.

This final chapter builds on these earlier chapters by answering question (4). As it now seems likely that Australia will abandon the sui generis guardrails approach originally proposed for regulating AI, it suggests an alternative way forward for addressing the regulatory gaps in the 5 areas of law discussed in Chapter Three.

Bibliography

Articles/Books/Reports

Alford, J, Head B.W., 'Wicked and Less Wicked Problems: a Typology and a Contingency Framework,' (2017) 36(3) Policy and Society 397

Althus, C, Bridgman, P, Davis, G, *The Australian Policy Handbook*, (Allen and Unwin, 2007)

Australian Competition and Consumer Commission, *Digital Platforms Inquiry: Final Report*, (2019)

Australian Competition and Consumer Commission, *Privacy Act Review Report: ACCC Submission*, (2023)

Australian Government, Australian Law Reform Commission, *Laws that Impose Strict or Absolute Liability*, (2016) ALRC Report 129

Australian Government, Australia's Chief Scientist, *Rapid Response Information Report, Generative AI: Language Models and Multimodal Foundation Models*, (2023)

Australian Government, Department of Health, Disability and Ageing, *Safe and Responsible Artificial Intelligence in Health Care – Legislation and Regulation Review Final Report*, (2025)

Australian Government, Department of Industry, Science and Resources, *Australia's Artificial Intelligence Ethics Principles*, (2019, 2024)

Australian Government, Department of Industry, Science and Resources, *Safe and Responsible AI in Australia*, Discussion Paper, (2023)

Australian Government, Department of Industry, Science and Resources, *Voluntary AI Safety Standard: Guiding Safe and Responsible Use of Artificial Intelligence in Australia*, (2024)

Australian Government, Department of Industry, Science and Resources, *The Australian Government's Interim Response to Safe and Responsible AI Consultation*, (2024)

Australian Government, Department of Industry, Science and Resources, *Introducing Mandatory Guardrails for AI in High-Risk Settings*, Proposals Paper, (2024)

Australian Government, The Treasury, *Review of AI and the Australian Consumer Law*, Discussion Paper (2024)

Australian Senate, Senate Select Committee on Adopting Artificial Intelligence, *Report*, (2024)

Austrian, Geoffrey, Herman Hollerith: *The Forgotten Giant of Information Processing*, (Columbia University Press, 1982)

Baldwin, Robert and Cave, Martin, *Understanding Regulation: Theory, Strategy, and Practice*, (Oxford, 1999)

Bagchi, S, 'What is a black box? A computer scientist explains what it means when the inner working of AI are hidden,' *The Conversation*, (online, 22 May 2023) <<https://theconversation.com/what-is-a-black-box-a-computer-scientist-explains-what-it-means-when-the-inner-workings-of-ais-are-hidden-203888>>

Bateman, J, and others, 'Beyond Open v Closed: Emerging Consensus and Key questions for Foundation AI Model Governance,' *Carnegie Endowment for International Peace* (2024)

Bell, G, Burgess, J, Thomas, J, Sadiq, S, *Rapid Response Information Report: Generative AI - Language Models and Multimodal Foundation Models*, (2023), Australian Council of Learned Academies

Bikkasani, D.C., 'Navigating artificial general intelligence (AGI): societal implications, ethical considerations, and governance strategies,' (2025) 5 *AI Ethics* 2021

Bingham, T, *The Rule of Law*, London, 2010,

Bostrom, Nick, *Superintelligence: Paths, Dangers, Strategies*, (Oxford University Press, 2014)

Bradford, Anu, *Digital Empires: The Global Battle to Regulate Technology*, (Oxford University Press, 2023)

Bradford, Anu, *The False Choice Between Digital Regulation and Innovation*, (2024) 119 *Northwestern University Law Review*, 377

Baldwin, R, and Black, J, "Driving Priorities in Risk-based Regulation: What's the Problem?," (2016) 43(4) *Journal of Law and Technology* 565

Brown, Micah, 'Global AI Legal Frameworks: Challenges and Implications' [2025] *Cingularity Press*

Brownsword, R, 'Generative AI and the Rule of Law' in Philipp Hacker et al (eds), *The Oxford Handbook of the Foundations and Regulation of Generative AI* (Oxford University Press, 1st ed, 2025)

Brożek, Bartosz et al, 'The Black Box Problem Revisited. Real and Imaginary Challenges for Automated Legal Decision Making' (2024) 32(2) *Artificial Intelligence and Law* 427

Calo, R, 'Artificial Intelligence Policy: A Primer and Roadmap,' (2017) 51 *U.C.D.L. Rev.* 399

Carney, T, 'Robo-Debt Illegality: A Failure of Rule of Law Protections?,' *AUSPUBLAW*, (online, 30 April 2018) <<https://www.auspublaw.org/blog/2018/04/robo-debt-illegality>>.

Chabert, Jean-Luc (ed), *A History of Algorithms: from the Pebble to the Microchip*, (Springer, 1999)

Coglianese, C , and Lai, A, 'Algorithm v Algorithm,' (2022) 71 *Duke L.J.*, 1281

Cole, Mark D, 'International AI Regulation and Governance on a Global Scale: An Overview of International, Regional and National Instruments' (2024) 1(1) *Journal of AI Law and Regulation* 126

Commonwealth Ombudsman, *Automation in the Targeted Compliance Framework: when the law is changed but the system isn't: An investigation into the cancellation of income support payments under the Targeted Compliance Framework*, Report (2025)
Commonwealth, *Royal Commission into the Robodebt Scheme*, Report (2023)

Commonwealth, States and Territories, *National Framework for the Assurance of Artificial Intelligence in Government: A Joint Approach to Safe and Responsible AI* by the Australian, State and Territory Governments, (2024)

Correa, N.K., and ors, 'Worldwide AI Ethics: A Review of 200 Guidelines and Recommendations for AI Governance,' (2023) *Patterns* 4

Crawford, Kate, *Atlas of AI*, (Yale University Press, 2021)

Crawford, Kate, and Paglen, Trevor, *Excavating AI: The Politics of Images in Machine Learning Training Sets*, (2021) 36(4) *AI and Society*, 1105

Crawford, Kate and Jason Schultz, 'The Work of Copyright Law in the Age of Generative AI' [2024] (94) *Grey Room* 56

Creemers, Rogier, 'The Regulation of Generative AI in China' in Philipp Hacker et al (eds), *The Oxford Handbook of the Foundations and Regulation of Generative AI* (Oxford University Press, 1st ed, 2025)

Csernaton, R, 'The EU's AI Power Play: Between Deregulation and Innovation,' *Carnegie Endowment for International Peace*, (2025)

Déchaux, R, 'The Issues of Legal Regulation of Artificial Intelligence,' in Charalampos, S, (ed) *Universal Solutions to Global Problems: Holistic Analysis of Law Connecting Theory and Practice*, 98 (Cambridge Scholars Publishing 2023)

Deng, L, *Artificial Intelligence in the Rising Wave of Deep Learning: The Historical Path and Future Outlook*, (2018) 35 *IEEE Signal Processing Magazine*, Vol.1, p172
Digital Platform Regulators Forum, *Examination of Technology: Large Language Models*, Working Paper, (2023)

Easterbrook, Frank H, *Cyberspace and the Law of the Horse*, University of Chicago Legal Forum, 1996)

European Commission, On Artificial Intelligence – A European Approach to Excellence and Trust, White Paper, (2020)

European Commission, Digital Single Market Policy, Big data. (2020).

Finn, Ed, What Algorithms Want: Imagination in the Age of Computing, (Massachusetts Institute of Technology, 2017)

Fleming, John G, *Fleming's the Law of Torts*, ed C Sappideen and Prue Vines (Thomson Reuters (Professional) Australia Limited: Lawbook Co., 11th edition., 2024)

Floridi, L, *The Philosophy of Information*, (Oxford, 2011)

Floridi, L, *The Fourth Revolution: How the Infosphere is Reshaping Human Reality*, Oxford, 2014

Forum of Small States and Digital Forum of Small States, AI Playbook for Small States, (2024)

Franklin, U, *The Real World of Technology*, revised edition, (Anansi 1999)

Fraser, Henry L and Nicolas P Suzor, 'Locating Fault for AI Harms: A Systems Theory of Foreseeability, Reasonable Care and Causal Responsibility in the AI Value Chain' (2025) 17(1) *Law, Innovation and Technology* 103

Frieberg, Arie, *Regulation in Australia*, edn 2 (The Federation Press, 2025)

Gordon-Tapiero, A, 'A Liability Framework for AI Companions' (March 10, 2025). 1 Geo. Wash. J. L. & Tech (Forthcoming)

Gray, Pamela N., *Artificial Legal Intelligence*, (Dartmouth, 1997)

Grobelnik, M, Perset, M, Russell, S, Organisation for Economic Cooperation and Development, 'What is AI? Can you make a clear distinction between AI and non-AI systems?,' (online, 6 March 2024) <https://oecd.ai/en/work/definition>

Gryz, J, Rojszczak, M, 'Black Box Algorithms and the Rights of Individuals: No Easy Solution to the "Explainability" Problem' (2021) 10(2) *Internet Policy Review* 1

Handler, M, Rolph, D, 'A Real Pea Souper': the Panel Case and the Development of the Fair Dealing Defences to Copyright Infringement in Australia,' (2003) 27(2) *Melbourne University Law Review* 381

Hanks, P, 'Administrative Law and Welfare Rights: a 40-year Story from Green v Daniels to Robo debt recovery,' (2017) 89 *AIAL Forum* 1.

Huq, Aziz Z, 'Artificial Intelligence and the Rule of Law,' in *Routledge Handbook of the Rule of Law* (Routledge, 1st ed, 2024)

Hutchinson, T, 'Doctrinal Research: Researching the Jury' in Dawn Watkins and Mandy Burton (eds), *Research Methods in Law*, (Taylor & Francis Group, 2013)

International Technology Industry Council, 'Understanding Foundation Models & The AI Value Chain: ITI's Comprehensive Policy Guide,' (2023)

Jackson, Bernard S., *Semiotics and Legal Theory*, (Routledge, 1985)

Johnson, David R., and Post, D, 'Law and Borders—The Rise of Law in Cyberspace' (1996) 48(5) *Stanford Law Review* 1367

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

King, J, Meinhardt, C, 'Rethinking Privacy in the AI Era: Policy Provocations for a Data-Centric World,' *Stanford University Human-Centred Artificial Intelligence*, White Paper, (2024)

King, S, Productivity Commission, 'The Best way to Regulate AI might be not to Specifically Regulate AI. Here's why,' (2024)

Lim, Hannah Yee-Fen, *Regulatory Compliance*, in Charles Kerrigan (ed), *Artificial Intelligence: Law and Regulation*, (Elgar Online, 2022)

Lindsay, D, 'Some Problems of 'Consent' Under Australian Data Privacy Law,' (2022) 3(3) *Global Privacy Law Review*, p 116

Kretschmer, M, et al, 'The Risks of Risk-Based AI Regulation: Taking Liability Seriously' [2023] *SSRN Electronic Journal* <https://www.ssrn.com/abstract=4622405>

KPMG and the University of Melbourne, *Trust, Attitudes and Use of Artificial Intelligence: A Global Study 2025, Australian Insights*.

Kowalkiewicz, Marek, *The Economy of Algorithms: AI and the Rise of the Digital Minions*, (La Trobe University Press, 2024)

Li, Tiffany C, 'Privacy and Disinformation,' (online 2025) <<https://ssrn.com/abstract=5318708>>

Jack Lindsay and ors, 'On the Biology of a Large Language Model,' *Transformer Circuits Thread*, (online, 27 March 2025) < <https://transformer-circuits.pub/2025/attribution-graphs/biology.html>>

Manwaring, K, 'Click Here to (Dis)agree: Australian Law and Practice in Relation to Informed Consent,' (2022) 3(3) *Global Privacy Law Review* 127

McCarthy, J, Minsky, M, Rochester, N, Shannon, C, *A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence*, 31 August 1955, (online)
<<https://raysolomonoff.com/dartmouth/boxa/dart564props.pdf>>

McCorduck, Pamela, *Machines Who Think: A Personal Inquiry into the History and Prospects of Artificial Intelligence*, (Boca Raton, 2019)

Mendelson, Danuta, *The New Law of Torts, edn 3*, (Oxford University Press, 2014)

Xiao-Le Meng, 'Data Science and Engineering with Human in the Loop, Behind the Loop and Above the Loop,' (2023) 5(2) *Harvard Data Science Review* 1

Moses, Lyria Bennett, 'How to Think about Law, Regulation and Technology: Problems with 'Technology' as a Regulatory Target,' (2013) 5(1) *Law, Innovation and Technology* 1.

Muggleton, S, 'Alan Turing and the Development of Artificial Intelligence' (2014) 27 *AI Communications* 3.

National Institute for Science and Technology, *NIST Big Data Interoperability Framework: Volume 1, Definitions* (version 3, 2019)

New South Wales, The Treasury, *NSW Government Guide to Better Regulation*, January 2019
Nicholls, R, 'Informed Consent to Online Standard Form Agreements' (2022) 3(3) *Global Privacy Law Review* 163

National Transport Commission, *Automated Vehicle Safety Reforms: Public Consultation*, Consultation Report (2024)

National Transport Commission, *National Rail Action Plan Newsletter: April 2025*, May 2025

Nissenbaum, Helen, *Privacy in Context: Technology, Policy, and the Integrity of Social Life*, (Stanford Law Books, 2010)

O'Neil, C, *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy*, (Allen Lane, 2016)

OpenAI and ors, *GPT-4 Technical Report*, (online 4 March 2024)
<<https://arxiv.org/abs/2303.08774>>

Organisation for Economic Cooperation and Development, *AI Openness: A Primer for Policy Makers*, OECD Artificial Intelligence Papers No.44 (2025)

Pagallo, U, 'Algo-Rhythms and the Beat of the Legal Drum,' (2018) 31 *Philosophy and Technology*, 507

Pasquale, Frank, *The Black Box Society: The Secret Algorithms That Control Money and Information* (Harvard University Press, First Harvard University Press paperback edition, 2016)

Ready Research and the University of Queensland, *AI Governance in Australia, 2024 Australian Survey Assessing Risks from AI*, Survey Report (2024)

Organisation for Economic Cooperation and Development, 'Intellectual Property Issues in Artificial Intelligence Trained on Scraped Data.' OECD Artificial Intelligence Papers No.33 (2025)

Paterson, Jeannie Marie, 'Regulating Generative AI in Australia: Challenges of Regulatory Design and Regulator Capacity' in Philipp Hacker et al (eds), *The Oxford Handbook of the Foundations and Regulation of Generative AI* (Oxford University Press, 1st ed, 2025)

Productivity Commission, *Making the Most of the AI Opportunity: AI Uptake, Productivity, and the Role of Government*, Research Paper 1 (2024)

Productivity Commission, *Making the Most of the AI Opportunity: The Challenges of Regulating AI*, Research Paper 2 (2024)

Productivity Commission, *Making the Most of the AI Opportunity: AI Raises the Stakes for Data Policy*, Research Paper 3 (2024)

Productivity Commission, *Harnessing Data and Digital Technology*, Interim Report (2025)

Rachum-Twaig, O, 'Whose Robot is it Anyway?: Liability for Artificial-Intelligence-Based Robots, [2020] University of Illinois Law Review 1141

Reed, C, 'How Should We Regulate Artificial Intelligence?' (2018) 376 (2128) *Philosophical Transactions of the Royal Society*

Russell, S, and Norvig, P, *Artificial Intelligence: A Modern Approach*, (Global edition, Pearson, 2021)

Scherer, M, 'Regulating Artificial Intelligence Systems: Risks, Challenges, Competencies, and Strategies,' (2016) 29 Harvard Journal of Law and Technology, No 2, 353

Schwartz, O, 'Untold History of AI: When Charles Babbage Played Chess with the Mechanical Turk,' (IEEE Spectrum 2019) (online <<https://spectrum.ieee.org/untold-history-of-ai-charles-babbage-and-the-turk>>)

Selbst, A.D, 'Negligence and AI's Human Users,' (2020) 100 Boston University Law Review 1315

Sheard, N, 'No Notice and No Explanation: The Incontestability of Hiring Discrimination by Algorithm,' [2022] Australian Journal of Labour Law, 119

Soh, J, Lim, L, Yeong, Z.K, 'Artificial Intelligence in the Regulatory Wonderland' [2025] *SSRN Electronic Journal* <https://www.ssrn.com/abstract=5226111>

- Solove, D, 'Conceptualizing Privacy,' (2002) 90 California Law. Review 1087
- Solove, D, 'The Limitations of Privacy Rights,' (2023) 98(3) Notre Dame Law Review 975
- Solove, D.J, Hartzog, W, 'The Great Scrape: The Clash Between Scraping and Privacy,' [2024] Boston University School of Law, Scholarly Commons at Boston University Law School, 1
- Solove, D. J, 'Artificial Intelligence and Privacy' (2025) 77(1) Florida Law Review 1
- Soyer, Baris and Andrew Tettenborn, 'Artificial Intelligence and Civil Liability—Do We Need a New Regime?' (2022) 30(4) International Journal of Law and Information Technology 385
- Sparrow, Malcolm K, *The Regulatory Craft: Controlling Risks, Solving Problems, and Managing Compliance*, (Brookings Institution Press, 2000)
- Supreme Court of Victoria, *Guidelines for Litigants: Responsible Use of Artificial Intelligence in Litigation*, May 2024.
- Tricker, Bob, *Corporate Governance: Principles, Policies, and Practices*, (Oxford University Press, 2009)
- Turing, A, 'Computing Machinery and Intelligence,' *Mind*, Vol. LIX. No. 236., Oct 1950
- Turner, Jacob, *Robot Rules: Regulating Artificial Intelligence*, (Palgrave McMillan, 2019)
- Ulbricht, Lena and Karen Yeung, 'Algorithmic Regulation: A Maturing Concept for Investigating Regulation *of* and *through* Algorithms' (2022) 16(1) *Regulation & Governance* 3
- Vardi, M, 'Artificial Intelligence: Past and Future,' (2012) 55(1) *Communications of the ACM*, 5
- University of Technology Sydney, Human Technology Institute, *Essential Components of AI Governance: AI Governance Snapshot #1*, Shaping our Future Symposium 2024, (31 January 2024)
- Walsh, T, Levy, N, Bell, G, Elliott, A, Maclaurin, J, Mareels, I, Wood, F, *The effective and Ethical Development of Artificial Intelligence: An Opportunity to Improve our Wellbeing*, Report for the Australian Council of Learned Academics (2019)
- Wachter, S, 'Limitations and Loopholes in the EU AI Act and the AI Liability Directives: What this means for the European Union, the United States, and Beyond,' (2024) 26(3) *Yale Journal of Law and Technology* 671
- Zuboff, Shoshana, *The Age of Surveillance Capitalism: The Fight for a Human Future at the Frontier of Power*, (Public Affairs/Hachette Book Group, 2019)

B Cases

AGL Energy Ltd v Greenpeace Australia Pacific Ltd (2021) 395 ALR 275

Burnie Port Authority v General Jones Pty Ltd (1994) 179 CLR 520.

Church of Scientology v Woodward (1982) 154 CLR 25

Donoghue v Stevenson [1932] AC 562

Facebook Inc v Australian Information Commissioner [2022] FCAFC 9

Gala v Preston (1991) 172 CLR 243

Hubbard v Vosper [1972] 2QB 84

Minister for Immigration and Multicultural and Indigenous Affairs: Ex parte Palme (2003) 216 CLR 212

MZAPC v Minister for Immigration and Border Protection (2021) 273 CLR 506

Norkin v University of New England [2023] NSWCA 194,

Prygodicz v Commonwealth (No 2) [2021] FCA 634

Ravenscroft v Herbert and New English Library [1980] RPC 193

Roads and Traffic Authority of NSW v Dederer [2007] HCA 42 [43].

Rylands v Fletcher (1886) LR 1 Ex 265

TNT Management Pty Limited v Brooks (1979) ALJR 267

Venning v Chin (1974) 10 SASR 299

C Legislation

Australian Constitution

Corporations Act 2001 (Cth)

Competition and Consumer Act 2010 (Cth)

Copyright Act 1968 (Cth)

Crimes (Aviation) Act 1991 (Cth)

Criminal Code Act 1995 (Cth)

Customs Act 1901 (Cth)

Damage by Aircraft Act 1999 (Cth)

Environment Protection and Biodiversity Act 1999 (Cth)

Occupational Health and Safety Act 2004 (Vic)

Peoples Republic of China, *Provisions on the Management of Algorithmic Recommendations in Internet Information Services*, 31 December 2021, <

https://www.cac.gov.cn/2022-01/04/c_1642894606364259.htm>

Peoples Republic of China, *Interim Measures for the Management of Generative Artificial Intelligence Services*, 23 May 2023.

<https://www.cac.gov.cn/2023-07/13/c_1690898327029107.htm>

Peoples Republic of China, Ministry of Foreign Affairs, *Global AI Governance Action Plan*, 26 July 2025. < https://www.fmprc.gov.cn/eng./xw/zyxw/202507/t20250729_11679232.html>

Privacy Act 1988 (Cth)

Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation)

Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence and amending Regulations (EC) No 300/2008, (EU) No 167/2013, (EU) No 168/2013, (EU) 2018/858, (EU) 2018/1139 and (EU) 2019/2144 and Directives 2014/90/EU, (EU) 2016/797 and (EU) 2020/1828 (Artificial Intelligence Act)

Social Security Act 1991 (Cth)

Superannuation Industry (Supervision) Act 1993 (Cth)

Supreme Court of Judicature Act 1873, 36 & 37 Vict., Ch 66

Therapeutic Goods Act 1989 (Cth)

Therapeutic Goods (Medical Devices) Regulations 2002 (Cth)

White House, Executive Office of the President of the United States, 'Winning the Race: America's AI Action Plan,' (July 2025)

White House, Executive Order, *Accelerating Federal Permitting of Data Centre Infrastructure*, (July 2025)

White House, Executive Order, *Preventing Woke AI in the Federal Government*, (July 2025)

White House, Executive Order, *Promoting the Export of the American AI Technology Stack*, (July 2025)

Wrongs Act 1958 (Vic)

D Treaties and International Instruments

Berne Convention for the Protection of Literary and Artistic Works (1886)

Bletchley Declaration by Countries Attending the AI Safety Summit, (1-2 November 2023)

Convention for the Protection of Individuals with regard to Automatic Processing of Personal Data (1981, 2018)

Council of Europe Framework Convention on Artificial Intelligence and Human Rights, Democracy and the Rule of Law (2024)

International Covenant on Civil and Political Rights (1966)

Paris Statement on Inclusive and Sustainable Artificial Intelligence for People and the Planet (2025)

Seoul Declaration for Safe, Innovative and Inclusive AI: AI Seoul Summit (2024)

Seoul Statement of Intent Towards International Cooperation on AI Safety Science, AI Seoul Summit (2024)

Seoul Ministerial Statement for Advancing AI Safety, Innovation and Inclusivity (2024)

E Other

Anthropic, 'Anthropic's Interpretability Research,' (online, August 2025)
<<https://transformer-circuits.pub>>

John Perry Barlow, 'A Declaration of the Independence of Cyberspace', Davos, (online, 8 February 1996) <<https://www.eff.org/cyberspace-independence>>

Jill Barshay, 'Kids who use ChatGPT as a Study Assistant do Worse on Tests,' The Hechinger Report, (online, 2 September 2024) <<https://hechingerreport.org/kids-chatgpt-worse-on-tests/>>

Daniel Castro and Michael McLaughlin, 'Ten Ways the Precautionary Principle Undermines Progress in Artificial Intelligence,' Information Technology and Innovation Foundation, (online, 4 February 2019) <<https://itif.org/publications/2019/02/04/ten-ways-precautionary-principle-undermines-progress-artificial-intelligence/>>

Defence Advanced Research Projects Agency, 'AI Next Campaign.' (online, 2023) <https://www.darpa.mil/work-with-us/ai-next-campaign>

Gilbert and Tobin, 'Is AI Training Fair Use of Copyright Works?,' (online, 26 May 2025) <<https://www.gtlaw.com.au/insights/is-ai-training-fair-use-of-copyright-works>>

Vice President Al Gore, 'Information Superhighways Speech,' International Telecommunications Conference, (online, 21 March 1994) <<http://vlib.iue.it/history/internet/algorespeech.html>>

Melissa Heikkila, 'OpenAI's hunger for data is coming back to bite it,' MIT Technology Review, (online 19 April 2023) <<https://www.technologyreview.com/2023/04/19/1071789/openais-hunger-for-data-is-coming-back-to-bite-it/>>

HistoryofInformation.com, *Blaise Pascal Invents a Calculator*. (online) <<https://historyofinformation.com/detail.php?entryid=440>>

HistoryofInformation.com, *Leibniz Invents the Stepped Drum Gear Calculator*, (online) <<https://www.historyofinformation.com/detail.php?id=394>>

Jim Holdsworth and Mark Scapicchio, IBM, 'What is Deep Learning?,' (online, 17 June 2024) <<https://www.ibm.com/topics/deep-learning>>

Paul Keating, 'Statement by the Prime Minister, Hon. P.J. Keating, 'A National Strategy for Information and Communications Services and Technologies', Media Release, (online, 6 April 1994) < <https://pmtranscripts.pmc.gov.au/release/transcript-9541>>

Fangfang Lee, IBM, 'What is a Neural Network?,' (online, 10 September 2025) <<https://www.ibm.com/think/topics/neural-networks>>

Steve Lohr, 'A.I. Is Prompting an Evolution, Not Extinction, for Coders,' *The New York Times*, (online, 20 February 2025) <<https://www.nytimes.com/2025/02/20/business/ai-coding-software-engineers.html>>

Christopher Manning, Stanford University, Human-Centered Artificial Intelligence, *Artificial Intelligence Definitions*, (online 2020.) <<https://hai-production.s3.amazonaws.com/files/2020-09/AI-Definitions-HAI.pdf>>

Bernard Marr, 'A Short History of Machine Learning – Every Manager Should Read,' Forbes, (online, 19 February 2016) <https://www.forbes.com/sites/bernardmarr/2016/02/19/a-short-history-of-machine-learning-every-manager-should-read/>

News Corporation, 'News Corp and OpenAI Sign Landmark Multi-Year Global Partnership,' Press Release (online, 22 May 2024) < <https://investors.newscorp.com/node/15946/pdf>>

Bryan Robinson, 'AI Goes Rogue: Do 5 Things If Your Chatbot Lies, Schemes or Threatens,' Forbes (online, 3 July 2025) < <https://www.forbes.com/sites/bryanrobinson/2025/07/03/ai-goes-rogue-do-5-things-if-your-chatbot-lies-schemes-or-threatens/>>

Ian Stewart, 'The Mathematical Equation that Caused the Banks to Crash,' The Guardian, (online, 11 February 2012) See <<https://www.theguardian.com/science/2012/feb/12/black-scholes-equation-credit-crunch>>

Cole Stryker, IBM, 'What is Advanced Analytics?,' (online 10 July 2024) <https://www.ibm.com/think/topics/advanced-analytics>

University of California Berkeley, School of Information, *What is Machine Learning (ML)?*, (online, 26 June 2020) <https://ischoolonline.berkeley.edu/blog/what-is-machine-learning/>

Chris Westfall, 'The Dark Side of AI: Tracking the Decline of Human Cognitive Skills,' Forbes Magazine, (online, 18 December 2024)

<https://www.forbes.com/sites/chriswestfall/2024/12/18/the-dark-side-of-ai-tracking-the-decline-of-human-cognitive-skills/>

Cam Wilson, 'Clearview AI is Still Collecting Photos of Australians for its Facial Recognition Database,' Crikey (online, 8 February 2025) <https://www.crikey.com.au/2024/02/08/clearview-ai-australia-facial-recognition-data/>

Adam Zwe, 'Explained: Generative AI's Environmental Impact,' MIT News, (online, 17 January 2025) < <https://news.mit.edu/2025/explained-generative-ai-environmental-impact-0117>>