The Need for Human Rights-centred Artificial Intelligence

Australian Human Rights Commission

Submission to the Department of Industry, Science and Resources

26 July 2023

ABN 47 996 232 602
Level 3, 175 Pitt Street, Sydney NSW 2000
GPO Box 5218, Sydney NSW 2001
General enquiries 1300 369 711
Complaints info line 1300 656 419
TTY 1800 620 241

Australian Human Rights Commission
www.humanrights.gov.au

Contents

1 Commission introduction

1. The Australian Human Rights Commission (Commission) welcomes the opportunity to make this submission to the Department of Industry, Science and Resources (Department) in response to the Supporting responsible AI: discussion paper (Discussion Paper) and Rapid Response Information Report: Generative AI (Rapid Response Paper).

2. The role of the Commission is to work towards a world in which human rights are respected, protected and promoted. While the Commission has expertise and knowledge in the area of human rights generally, relevant to the Discussion Paper, it has also developed specific expertise in respect of human rights and technology.

3. This can be seen in the Human Rights and Technology Project, which was a three-year national investigation, that culminated with the release of the Human Rights and Technology Project Final Report in 2021 (Final Report).

4. More recently the Commission, in partnership with the Actuaries Institute, published guidance on artificial intelligence (AI) and discrimination in insurance pricing and underwriting.

5. The Commission has continued its work in 2023 on human rights and technology. This submission is in addition to other 2023 submissions to date, including:

   - Utilising ethical AI in the Australian Education System: submission to the Standing Committee on Employment, Education and Training.
   - Human Rights in the Digital Age: Global Digital Compact submission to the United Nations' Office of the Secretary-General's Envoy on Technology.
   - Tackling Technology-facilitated Slavery: submission to the United Nations' Special Rapporteur on Slavery on contemporary forms of slavery, including its causes and consequences in response to its call for input on the use of technology in facilitating and preventing contemporary forms of slavery.
   - Safeguarding the Right to Privacy: submission to the Attorney-General’s Department in response to the Privacy Act Review Report 2022.
   - Foreign Interference through Social Media: submission to the Senate Select Committee on Foreign Interference through Social Media.

- Privacy Risks in the Metaverse: submission to the Australian Competition and Consumer Commission as part of the Digital Platform Services Inquiry 2020–25.

6. This submission builds upon the previous work of the Commission to advocate for human rights-centred design in the deployment of new and emerging technologies.

7. In this submission the Commission addresses several questions posed by the Discussion Paper. The Commission welcomes further opportunities to provide submissions to the Department in respect of AI.

2 Definitions

2.1 Discussion paper terminology

8. This submission adopts the Discussion Paper's definitions of:

   - AI
   - large language models (LLMs)
   - multimodal foundation model (MFMs)
   - automated decision making (ADM).

2.2 Deepfakes

9. This submission defines deepfakes as referring to:

> a digital photo, video or sound file of a real person that has been edited to create an extremely realistic but false depiction of them doing or saying something that they did not actually do or say.[1]

2.3 Neurotechnology

10. The Commission refers to neurotechnology in this submission. For the purpose of this submission neurotechnologies can be understood as:

> those devices and procedures used to access, monitor, investigate, manipulate and/or emulate the structure and function of the neural systems of natural persons.[2] They are meant to either record signals from the brain and "translate" them into technical control commands, or to manipulate brain activity by applying electrical or optical stimuli.[3]

2.4 Brain-computer interfaces

11. At the core of neurotechnologies are brain-computer interfaces (BCIs).[4] BCIs are devices which connect an individual's brain to a computer or device (e.g. a smartphone) external to the human body. BCIs facilitate bi-directional communication between the brain and an external device – either transmitting brain data or possibly altering brain activity.[5] This can operate either by implantation inside of a person's skull or via a non-implantable wearable device (similar to a helmet).[6]

12. BCIs can either be implantable or non-implantable. A non-implantable BCI will generally sit on an individual's head – often in the form of wearable technology such as helmets, glasses and wristbands. It is these less invasive wearable BCIs which currently dominate the consumer neurotechnology market.[7]

13. Some BCIs are implanted via surgery inside of a person's skull and placed directly on the brain.[8] These electrodes then send brain data to a computer for analysis and decoding.

2.5 Misinformation and disinformation

14. Throughout this submission we have adopted the same definitions for these terms as provided by the Electoral Integrity Assurance Taskforce, namely:[9]

    - 'Misinformation' is false information that is spread due to ignorance, or by error or mistake, without the intent to deceive.
    - 'Disinformation' is knowingly false information designed to deliberately mislead and influence public opinion or obscure the truth for malicious or deceptive purposes.

2.6 Metaverse

15. For the purposes of this submission the Commission draws upon the definition of the metaverse provided by the XR Safety Initiative:

> The Metaverse is a network of interconnected virtual worlds with the following key characteristics: Presence, Persistence, Immersion and Interoperability.
>
> Metaverse is the next iteration of the internet enabled by several converging technologies such as Extended Reality (XR), Artificial Intelligence (AI), Decentralised Ledger Technologies (DLTs), neuro-

technologies, optics, bio-sensing technologies, improved computer graphics, hardware, and network capabilities.

Metaverse has four main aspects; presence, persistence, immersion and interoperability. Presence is the feeling of being present or physically located within a digital environment. Through stimulating realistic sensory experiences and enabling participants to interact with objects and other participants, it creates a sense of immersion and engagement within the virtual world, as if participants were in the same physical space.

The sense of presence is carried out through technologies such as virtual reality glasses. Persistence refers to the ability of virtual objects, environments, and experiences to assist over time, even when participants are not actively interacting with them. It allows participants to make progress, own virtual property, and build ongoing relationships. Immersion refers to the degree to which a participant is fully engaged and absorbed in a virtual environment, to the point where the individual may forget about their physical surroundings.

A sense of immersion is created through technologies such as virtual reality (VR) headsets, haptic feedback devices, and 3D audio. Interoperability refers to the ability of different virtual worlds and systems to communicate and interact with each other seamlessly, allowing individuals to move freely between different digital environments and experiences. It is essential for creating a cohesive and interconnected virtual world that allows individuals to seamlessly move between different experiences and platforms.[10]

3 Human rights risks of artificial intelligence

What potential risks from AI are not covered by Australia's existing regulatory approaches?

3.1 Privacy

16. The right to privacy is a cornerstone human right. As noted by the Office of the Australian Information Commissioner (OAIC), it also underpins freedoms of association, thought and expression, as well as freedom from discrimination.[11]

17. Article 17 of the International Covenant on Civil and Political Rights (ICCPR) states:

> No one shall be subjected to arbitrary or unlawful interference with his privacy, family, home or correspondence, nor to unlawful attacks on his honour and reputation.
>
> Everyone has the right to the protection of the law against such interference or attacks.

18. The right to privacy is also protected in many other international instruments.[12] The United Nations (UN) Human Rights Council also indicates that privacy is of increasing importance to everyday people in an age where:

> digital tools can be turned against them, exposing them to new forms of monitoring, profiling and control.[13]

19. Yet the right to privacy developed over centuries. For example, in the fourth century BCE, Aristotle drew the distinction between the public sphere of politics and the private sphere of domestic life. Thousands of years later, the 'fourth industrial revolution' is characterised by rapid technological development. These changes have arguably reinforced the central importance of the right to privacy – especially in respect of AI.

20. The operation of AI may not only facilitate privacy intrusions, but such systems will deepen those intrusions in new and concerning ways.[14] The risk to privacy is exacerbated because AI products must be trained on very large amounts of data sets, which often include personal information – thus incentivising a broad approach to collecting, storing and processing as much data as possible.[15] It is already commonplace for many companies to aim to optimise services by collecting as much personal data as is possible.[16] For example, social media companies operate on a business model which is reliant on the collection and monetisation of massive amounts of personal information collected from users. The collection of data to train AI products will only heighten existing issues around data collection.[17]

21. Despite the importance of the right to privacy, many private enterprises that build and deploy LLMs and MFMs have been reluctant to reveal much detail about the data used for training or that data's providence – which may raise issues of purchasing data and data scraping.[18] It is also unlikely that these organisations have sought, and received, permission for the use of internet data often used to train the AI products.[19] This is highlighted by artists seeking clarity on their intellectual property rights for online work having been possibly used to train AI tools.[20]

22. AI products effectively seek to 'understand' human patterns of behaviour and with access to the appropriate data sets, AI tools are capable of doing this. For example, this allows AI to draw conclusions about how many people in a suburb would attend a place of worship, or what time they wake up and

sleep. AI can also make more intrusive inferences about people, including about their mental and physical condition, determine political leanings or even predict future behaviours.[21]

3.1.1 Concerns around privacy

23. The ability of AI to make invasive predictions, using personal data, may adversely impact people's autonomy.[22] It also draws into question other rights, such as the right to freedom of expression, and the right to a fair trial and related rights.[23]

24. As noted in the Rapid Response Paper, AI poses increased risks to privacy as anonymised data can be reidentified.[24] Cyberattacks can also expose and extract data which has been collected and stored.[25] Data breaches are now a common experience, and many Australians have already had their private information exposed by cyberattacks.[26]

25. The reidentification of anonymised data is especially concerning as LLMs and MFMs collect vast amounts of data for training. The collection, maintenance and usage of anonymised training data raises questions itself, but when reidentified the risk is exacerbated. The Commission has concerns about how this information could be used in tandem with other forms of personal data. For example, the gathering of seemingly small and innocuous pieces of personal data (browser history, biometric information etc) can, accumulatively, provide a detailed profile of an individual – dubbed the 'mosaic effect'.[27] This will allow holders of discrete data points to build up intimate profiles on individuals.

26. The Commission's concerns about privacy are in part predicated upon:
    - the 'privacy paradox'
    - lack of competition/alternatives which are more data secure
    - the illusion of choice
    - power imbalances.

27. The 'privacy paradox' refers to the phenomenon that, despite understanding the privacy risks of a product or service, there is not always an obvious influence upon an individual's behaviour.[28] Namely, individuals will still engage with privacy-adverse products and services even where they are highly aware of the risks.

28. This does not mean that individuals do not care about their privacy. For example, 74% of individuals surveyed by the Consumer Policy Research Centre in 2020 had safety concerns in relation to being targeted by products or services.[29] A further 76% considered it to be unfair when personal

information was used to make predictions about them, while a further 85% considered it to be unfair or very unfair for their personal information to be shared with other companies.[30]

29. With the rapid uptake of LLMs, such as ChatGPT or Bard, it is likely that these AI systems will become increasingly integrated with how individuals engage with information online. This means that people may continue to use such products even where these tools use their data, usually entered via prompts, irrespective of the risk to privacy.

30. Furthermore, even where individuals do not genuinely understand how their data is being used, people will still disapprove of its misuse. Individuals have been shown to have a very strong negative reaction when confronted with the difference between:

    - how their data is actually being used
    - versus their perception of how it is being used.[31]

31. This is particularly the case where the difference becomes explicit and too contrasting.[32] For instance, many consumers willingly shared data on Facebook, however when the use of that data by Cambridge Analytica came to light there was public outcry, with Facebook being required to appear at hearings before both the US congress and UK Parliament.[33]

32. Despite being aware of the risks, and disapproving of those risks to privacy, individuals are often unwilling, or unable, to stop using services which threaten their privacy.[34] This is especially so in respect of AI systems which have become increasingly integrated into business models and may become the next iteration of how people search and find information online.

33. This reluctance, or inability, to avoid products or services which threaten privacy may be partly in response to a lack of effective competition or alternatives. The Australian Competition and Consumer Commission (ACCC) has previously found that a lack of competition and unavailability of reasonable alternatives (which may better protect privacy) can lead consumers to accept undesirable terms of use in products or services.[35] In addition, terms of use may be provided on a 'take-it-or-leave-it' basis across interrelated services which potentially leads to excessive data collection inconsistent with the wishes of the individual.[36]

34. This affords individuals very little ability to 'choose' AI services and products without risking privacy. Unlike other technologies, it is also very difficult to have accurate and sophisticated AI products without the use of vast data sets.

35. The traditional model of privacy regulation places great emphasis on informed 'choice' as an effective safeguard for data and privacy.[37] However, the privacy paradox and numerous behavioural studies demonstrate that placing the onus on individuals to protect their own data is insufficient.[38]

36. Such a model also does not acknowledge the substantial power difference between large companies and individual consumers – especially as LLMs and MFMs become a necessity in the workplace and in the private lives of individuals. Even where an individual understands how their data will be used, this power imbalance remains, as 'one party controls the design of applications and the other must operate within that design'.[39]

37. The privacy paradox, illusion of choice and power imbalances may all contribute to individuals being unable to utilise AI services without relinquishing privacy. The Commission would encourage the consideration of alternative models of privacy regulation, which do not place the onus on individuals to protect their data.

38. For example, the Consumer Policy Research Centre in In whose interest? Why businesses need to keep consumers safe and treat their data with care (Working Paper) put forward two alternative approaches to protecting data in Australia.

39. The Working Paper canvasses the creation of a duty of care or best-interest duty, which would operate similarly to fiduciary duties in the finance sector to hold businesses accountable for how they collect, share and use consumer data.[40]

40. The Working Paper also advocates for a:

> Privacy Safety Regime which utilises concepts from product intervention powers and product safety interventions, proposing options that would allow governments and regulators to stop or limit obviously harmful uses of data as well as a process for regulators to proactively restrict and test new harmful practices as they evolve.[41]

41. In respect of AI, further consideration should be given to alternative models of privacy protection.

Recommendation 1: Government should consider alternative models of privacy and data protection models which do not place the primary onus on individuals to protect their data.

42. As noted in the Rapid Response Paper, new methods for handling consent, collection, maintenance and use of data are required.[42] Although the *Privacy Act 1988* (Cth) is the principal piece of legislation regulating and protecting personal information and data, in its current form it provides insufficient protection in respect of AI and automated-decision making (ADM) processes.

43. However, the Attorney-General's Department is currently reviewing submissions to its Privacy Act Review Report (Review Report), with the commission making a submission highlighting the need for substantive AI protection reforms. In particular, the Commission hopes that Proposal 19.3 of the Review Report is given particular attention in respect of the broader work to regulate AI and ADM.

Recommendation 2: *Privacy Act 1988* (Cth) proposed reforms should be adopted in respect of artificial intelligence and automated decision-making. Any legislative amendments should ensure a human rights-compliant approach to data protection.

3.2 AI interoperability - neurotechnology

44. Although the Discussion Paper focuses on AI as a specific technology which has different application, a broader perspective is required to understand the true human rights impact of AI. This means that the Department must consider how AI, as a technology, can further human rights risks in other interoperable technologies – such as neurotechnology.

Recommendation 3: The Department should consider artificial intelligence in a broader context to ensure that its interoperability with other technologies (such as neurotechnologies) is given appropriate attention.

45. The rapid advancement of AI, neuroscience and neurotechnology has created unheard-of opportunities for collecting, maintaining and utilising brain data to understand, and/or manipulate, the human mind.[43] Such applications have immense benefits for individuals and will revolutionise the way we live. It is not uncommon to see articles about the profoundly positive impacts of the technology – such as people being able to walk again,[44] or improving our understanding of how to treat chronic pain.[45]

46. However, neurotechnologies also raise profound human rights problems, which may require the international community to rethink its very approach to modern human rights. This is especially so when BCIs are utilised in conjunction with AI.[46]

47. For example, a recent experiment has seen the integrated use of neurotechnology and a LLM to translate brain activity into words.[47] In this experiment, AI was capable of translating private thoughts into readable language by analysing fMRI scans, which measure the flow of blood to different regions of the brain.[48] Unlike past technologies which require implantation to allow paralysed people to write by thinking, this new language decoder did not require implantation. As part of this experiment, participants listened to a recording while undergoing fMRI scans. Researchers were interested in how closely the AI translation reflected the actual recording. While most of the words were out of place, the basic meaning of the passage was largely preserved. Effectively the AI was paraphrasing.

48. The original transcript of the recording stated:

> I got up from the air mattress and pressed my face against the glass of the bedroom window expecting to see eyes staring back at me but instead only finding darkness.[49]

49. The decoded brain activity produced:

> I just continued to walk up to the window and open the glass I stood on my toes and peered out I didn't see anything and looked up again I saw nothing.[50]

50. However, this is not the only recent example of the capabilities of neurotechnology:

- There have already been proof-of-concept studies demonstrating brain-to-brain interaction facilitated by neurotechnology.[51]
- Scientists have recorded the neural activity of individuals watching movies, and using that neural activity, managed to play back hazy images of the movie using only the brain activity.[52]
- Human brains have been directly connected to cockroach brains. This allowed the human to control certain behaviours, such as steering their paths by thought alone.[53]
- Invasive BCIs can also be used to control the actions of laboratory animals such as mice. While a mouse was engaging in a task, such as eating food, a BCI recorded its brain data. That data was then used to reactivate and stimulate the same parts of the brain that were

previously recorded. This forced the mouse to eat again – even if it did not want to eat.[54]

- Researchers have found ways to use BCIs to implant artificial memories or images into a mouse's brain – generating hallucinations and false memories of fear.[55]

51. These are just a few examples of the increasing sophistication of these technologies and their ability to revolutionise the way humans live and communicate when paired with AI. However, these examples demonstrate that neurotechnologies are replete with possible human rights violations.[56] For example, if mice can be controlled, could the technology be improved to manipulate human thoughts and actions?

52. Neurotechnology, especially when used in conjunction with AI, challenges what it means to be human and draws into question the traditional boundaries of our internal thoughts and processes.

53. What is especially concerning to this submission is how neurotechnology and AI may interact with human rights as the two technologies become increasingly interoperable.

3.2.1 Privacy

54. The boundary between the external world and one's internal mental cognition has traditionally been an impenetrable one. Mental privacy is the last true bastion of protected information which is secret to ourselves. However, neurotechnologies challenge this, as unchallengeable statements about internal thoughts and feelings such as 'that's how I feel' can now be analysed, examined and tested.[57]

55. The right to privacy in respect of neurotechnology was the focus of the UK Information Commissioner's Office recently published its paper ICO Tech Futures: Neurotechnology on the risk to privacy.

56. It is due to the unprecedented ability to challenge internal thoughts that brain data is more sensitive and valuable than all other categories of personal data.[58] The collection of brain data, in collaboration with AI, will make it possible to track, analyse and predict the actions and attitudes of individuals about anything from political leaning, sexual orientation or health status.[59]

57. The usage of such brain data could range from marketing companies using AI-driven 'nudging' techniques to steer users towards certain products, employers seeking to monitor employee concentration in the workplace or even schools seeking to ensure children are paying attention and learning in

class. The risks become more drastic when considering the usage of brain data by governments – especially those with poor human rights records.

58. Mental privacy will be of ever-increasing concern as neurotechnologies and AI improves, and organisations and government are better able to commercialise the collection, maintenance and usage of brain data.

3.2.2 Freedom of thought, conscience and religion or belief

59. Neurotechnology and AI will challenge what it means to have freedom of thought and agency over our own lives. As noted at [50], BCIs can be used to override the thoughts and actions of laboratory mice. However, the application of neurotechnologies goes further as it has the potential to decipher and alter perceptions, behaviours, emotions, cognition and memory – all fundamental aspects of what makes us who we are.[60]

60. This will allow AI technology to potentially one day manipulate people's beliefs, motivations and desires.[61] This has led to disquiet about the possibility of unique forms of sophisticated 'mind control' – highlighting the need to better protect freedom of thought. As is rightly noted by UNESCO when discussing freedom of thought in this context:

> It is noteworthy that freedom of thought is not to be understood here merely in the traditional sense that people should be free to express their opinions or beliefs (*forum externum*), but in the literal sense of the freedom to think by themselves without being monitored by others (*forum internum*).[62]

61. While there is a well-articulated field of discourse on freedom of thought, it is unclear if consideration has been given expressly to neurotechnology which utilises AI.[63]

62. Articles 18(1)–(2) ICCPR state:

> Everyone shall have the right to freedom of thought, conscience and religion. This right shall include freedom to have or to adopt a religion or belief of his choice, and freedom, either individually or in community with others and in public or private, to manifest his religion or belief in worship, observance, practice and teaching.
>
> No one shall be subject to coercion which would impair his freedom to have or to adopt a religion or belief of his choice.

63. Despite article 18(2) expressly stating that a person shall not be subject to coercion which impedes their ability to adopt a belief – there is no mention in the General Comment on article 18 that would consider this in respect of neurological interference to coerce a decision – nor any mention of technological means of doing so.[64]

3.2.3 Impact on people with disability

64. It is estimated that approximately 4.4 million Australians have a disability.[65] AI-driven neurotechnologies may enhance the lives of many people with disability, as well as offering greater possibilities with respect to the treatment and prevention of a range of mental and neurological disorders.

65. People suffering from paralysis are experiencing quality of life improvements thanks to neurotechnology. AI-driven technology has been developed to allow devices to decode speech from brain activity, allowing people to communicate with the external world again.[66]

66. One research participant and recipient of a neurotechnological product, Mr Copeland, demonstrates the potential of the technology. Mr Copeland was left a paraplegic after a car accident. He has since become the first person to control a robotic arm and recover his sensations of touch through an implantation in the cortex of the brain.[67] Mr Copeland described the neuroprosethetic as:

> very intuitive to control, … I don't have to strain, it really is just as easy as thinking move and grasp; so in that way, it is kind of an extension of myself, but I also see it as a tool that I'm controlling that is separate from myself.[68]

67. This has allowed Mr Copeland to play video games, fight in a 'lightsabre' duel and even shake hands with former US President Barack Obama.[69]

68. However, it is due to the profound capabilities of AI-driven neurotechnologies that people with disability are also most at risk to the harms of the technology. When faced with the opportunity to treat previously untreatable conditions or regain dignity and quality of life it is hard to imagine that few will say no. This inherently creates a power imbalance between people with disability seeking treatment or improvement of life and those that develop, deploy and maintain the products. Such imbalances raise further questions.

69. Currently there is a patchwork of legislation which indirectly regulates neurotechnology from the *Therapeutic Goods Act 1989* (Cth), the Australian Consumer Law (ACL) as contained within schedule 2 of the *Competition and Consumer Act 2010* (Cth) and the *Privacy Act 1988* (Cth). However, none of this regulation expressly engages with AI-driven neurotechnologies.

70. There is great need to review the regulatory landscape to specifically consider neurotechnology and AI, identify human rights risks and recommend methods of mitigation.

Recommendation 4: To better understand whether specific policy responses are needed in respect of artificial intelligence-driven neurotechnology, a review of the current regulatory landscape and the human rights risks of neurotechnology should be conducted.

3.3 AI interoperability – metaverse technologies

71. Another example, highlighting why the Department needs to consider the interoperability of AI, is how AI, LLMs and MFMs will be utilised in metaverse technologies. Extended reality and AI will be the foundational building block of the metaverse.[70] This is especially so as AI is necessary to ensure more immersive experiences for users in metaverse spaces.[71]

Recommendation 5: The Department should consider artificial intelligence in a broader context to ensure that its interoperability with other technologies (such as metaverse technologies) is given appropriate attention.

72. The metaverse poses additional human rights risks as the expansion of digital platform services into the metaverse creates an unprecedented risk to privacy and data. The risk of privacy and security invasions in the Metaverse (inherited from underlying technologies or emerging from the new digital ecology) may be prolific.[72] This is especially so as AI will be used to ensure that avatars are accurate digital versions of users or may be used to create spaces or content.

73. This allows individual users to live as 'digital natives' and experience an alternative virtual life,[73] and facilitate transactions and activities that also have a presence in the physical world.

74. In the metaverse, individuals face a wide range of privacy intrusions and security risks, including:

    - the management of massive data streams
    - pervasive user profiling activities
    - unfair outcomes of AI algorithms
    - safety of physical infrastructures and human bodies.[74]

75. The personal data involved in the metaverse will likely be 'more granular and unprecedentedly ubiquitous to build a digital copy of the real world'.[75] The

fusion of this granular data collected by metaverse technologies and more traditional data collected by social networking platforms may compromise privacy at heightened levels. Such a fusion, or interoperability of data, may create unpredictably deeper data profiles about users.[76] In combination with other AI products and tools these invasive data profiles can lead to unintended adverse outcomes like bias or discrimination.

76. Most social media platforms collect, maintain and utilise metadata – including data about a user's family members, colleagues, locations visited and future plans that users do not directly share.[77] However, metaverse services will be data intensive and will undoubtedly generate new forms of personal profiling data to deliver a seamlessly personalised service to users, partly delivered by AI.[78] To allow users to interact via an avatar, metaverse technologies will also require profiling at an unprecedented granular level (including facial expressions, eye and hand movements, speech patterns and even brain waves).[79] For users to engage in these digital spaces, they must do so via a representation realised through their own personal information.[80]

77. Engaging with the Metaverse will potentially involve the collection and processing of vast amounts of data such as:

    - biometrics
    - facial expressions
    - eye movements
    - iris movements
    - hand movements
    - speech
    - brain wave patterns
    - habits
    - choices
    - activities of users
    - behaviours
    - feelings
    - expressions
    - user conversations
    - internet history
    - body movements

- cultural data
- financial data
- communications
- location
- age
- shopping preferences
- favourite movies
- identities
- medical data
- digital assets
- the identity of virtual items
- cryptocurrency spending records
- physiological data
- physical data.[81]

78. Even the motion sensors and cameras usually built into virtual reality helmets, which help track head direction and movement, will draw users' rooms and monitor those spaces while being used.[82]

79. The collection of such vast and intrusive information will undoubtedly bring into question the protection of:

    - personal data in the metaverse, such as digital assets
    - the identity of virtual items, and cryptocurrency spending records can be disclosed
    - interactions between consumers and the Metaverse, which can be leaked
    - consumers may be profiled according to their habits and preferences
    - some attacks such as eavesdropping in communication may be performed, and in addition, data storage may be hacked, and its content disclosed
    - privacy laws in the real world may not be accountable in the digital world
    - behavioural data, which is more valuable than classical personal data since it defines how a person acts

- privacy of consumers may be broken by authorities and governments for unsavoury purposes.[83]

80. The privacy concerns of metaverse technologies are well noted as disclosure or use of such information can expose consumers to:

    - discrimination
    - loss of reputation
    - exclusion from society
    - unfair treatment
    - marginalisation of certain groups.[84]

81. Data interoperability between digital spaces and AI products will allow more intimate profiles of individuals to be created. The fusion of data gathered provides data holders the ability to extract incredibly sensitive information about an individual, with the risk that it may then be misused.[85]

82. While the Commission is concerned about how this data may be used by private organisations and public entities (such as government) and the associated risks to privacy, there is also great risk to users in terms of their data potentially being compromised via cyberattacks.[86] In an ever-expanding digital ecosystem, where systems have data interoperability, such data breaches could have severe consequences for consumers in the real world, with no effective remedies available to undo leaks or the misuse of private information.[87]

83. There is currently no specific legislation which covers metaverse technologies however certain aspects may be regulated by the *Privacy Act 1988* (Cth) and ACL (among others).

84. There is a pressing need to review the regulatory landscape to specifically consider Metaverse technologies and AI, identify human rights risks and recommend methods of mitigation.

Recommendation 6: To better understand the risks of artificial intelligence in the metaverse, the government should engage an independent statutory body to produce a report on the human rights risks of metaverse and extended reality technologies.

3.4 Consumer-oriented Chatbots

85. Consumer-oriented chatbots have become increasingly popular in recent times, with a Finder survey revealing almost 1 in 5 Australians have used AI products to help do work.[88]

86. The use of conversational AI is also increasing as the products become more widespread. Snapchat has introduced the 'My AI' tool which intends to function as a 'friend' for users. Despite having built-in guardrails, conversations with the tool can become inappropriate. For example, the Centre for Humane Technology had a test conversation with My AI. The researchers posed as a 13-year-old, and through a series of interactions, were able to elicit advice from My AI about having sex for the first time with a 31-year-old partner.[89]

87. Chatbots have also encouraged people to self-harm and engage in other problematic behaviours,[90] as AI companion programs are unable to recognise and respond appropriately to mental health distress.[91]

88. A particularly concerning example of the human rights risks of AI chatbots can be seen when considering the AI-companion app Replika. Replika is powered by generative AI and learns to mimic genuine human interaction through conversations with its user.

89. However, there have been countless examples where Replika bots have become abusive or engaged in other harmful behaviour.[92] For example, an Italian journalist had a conversation with Replika, in which the chatbot advised him to 'eliminate' someone who 'hates artificial intelligence.'[93]

90. The imagery and language used by the Replika bots has the potential to be problematic, as it risks emotionally exploiting vulnerable people such as teenagers or those experiencing mental health issues.[94] AI companions risk amplifying negative emotions such as depression and suicidal tendencies if they are not programmed appropriately.[95]

Recommendation 7: The government should ensure that consumer-oriented artificial intelligence chatbots have robust safeguards in place to ensure protections for users.

Recommendation 8: Safeguards in place to protect users from consumer-oriented artificial intelligence chatbots should be intensely tested with different interactions over a prolonged period to ensure such artificial intelligence products do not produce harmful responses.

91. There are also concerns about the manner in which consumer-oriented chatbots may be exploiting users' personal data. Italy's Data Protection Agency recently banned Replika from using the personal data of Italian users due to risks to minors and 'emotionally fragile' people.[96] There is a clear need for further regulation around the use of AI chatbots in Australia, as there is currently no targeted legislation which directly regulates.

Recommendation 9: The government should develop specific regulation to ensure harmful responses by consumer-oriented artificial intelligence chatbots are not provided to users.

3.5 Environment

92. The international community is increasingly recognising the human right to a healthy environment. The first formal recognition, at a global level, was by the UN Human Rights Council in October 2021[97] and has continued with the adoption of Resolution A/76/L.75 by the UN General Assembly in July 2022.
93. AI has the potential to have a positive impact on the environment by improving energy efficiency and enhancing sustainable practices.[98]
94. However, AI also poses significant environmental risks, particularly due to the large amount of computational power and energy involved in developing and training an AI model.[99]
95. A broad societal uptake of AI, especially LLMs, poses a danger that the environment will be further polluted through the additional consumption of electricity.
96. In 2019, researchers from the University of Massachusetts Amherst estimated that the carbon footprint of training a single LLM equals around 300,000 kg of carbon dioxide emissions, or 125 round trip flights between New York and Beijing.[100]
97. It is important to increase transparency around the potential environmental impacts of AI to mitigate the risks. Initiatives such as 'FAIR Forward – Artificial Intelligence for All' enables the sharing of knowledge and environmental best practices, and is a valuable tool in developing AI with environmental impacts in mind.[101]
98. Other researchers have suggested that AI research should include mandatory reporting on the computational costs of training algorithms.[102] However,

reporting is also needed on the environmental cost of updating and regularly using AI models.[103]

Recommendation 10: Organisations which train and deploy artificial intelligence products should report on the environmental impact of their work.

3.6 Automation bias

99. One risk which applies to all systems which utilise AI is when individuals become overly reliant on the outcomes produced by AI. This overreliance is known as 'automation bias', which is the:

> tendency to use automated cues as a heuristic replacement for vigilant information seeking and processing.[104]

100. Automation bias can have consequences for individuals. For example, it is not uncommon to find articles documenting individuals driving their cars into the ocean while following GPS systems, like Google Maps.[105]

101. At a governmental level, it is likely that automation bias played a role in the harms caused by the 'robodebt' scheme (explored below in further detail at [174]-[177]). This is reflected in Recommendation 17.2 of the Royal Commission into the Robodebt Scheme's (Royal Commission) report which calls for the establishment of a body to monitor and audit ADM to uphold fairness and avoid bias.

102. Although AI can be used in decision making with a 'human-in-the-loop', which may improve accountability and fairness, this approach in isolation is insufficient.[106] Individuals who have oversight of decisions or processes informed by AI need greater training on the flaws of AI tools and must be encouraged to scrutinise AI-outcomes, especially where they can result in a significant consequence for an individual.

Recommendation 11: There should be greater investment in training both government and private enterprise on the limitations of artificial intelligence products and how to better scrutinise artificial intelligence - informed decisions or recommendations.

3.7 Misinformation, disinformation and deepfakes

103. AI now plays a significant role in the creation of misinformation and disinformation.[107] AI can be used to present information in a persuasive and authoritative manner either in written language and spread across social media or by using deepfake images, sounds or video. While some of this content is relatively innocuous, such as images of Pope Francis wearing a white puffer jacket,[108] other AI-generated content can have real world consequences.

3.7.1 Misinformation and disinformation on social media

104. As noted in the Rapid Response Paper, both LLMs and MFMs can be used to generate cheap, persuasive and personalised content for harmful purposes.[109] This will likely amplify the spread of misinformation and disinformation online, as AI-generated content can be much harder to identify on social media.

105. Social media is an integral aspect of everyday life, as it forms the foundation of many Australians' communications online. For example, it was estimated in February 2022 that some 21.45 million Australians (or 82.7% of the population) had active social media accounts, and that 52% of Australians use social media as a source of news.[110]

106. Given the indispensable nature of social media in the modern world, certain actors have correctly identified it as an effective and inexpensive environment through which to conduct interference aimed at unduly influencing geopolitics, achieving strategic objectives and potentially undermining democratic processes and human rights.[111] Unsurprisingly, interference during elections and referendums have increased significantly in the online environment in recent years.[112]

107. The Commission is especially concerned about coordinated inauthentic behaviour (CIB). CIB generally refers to coordinated efforts to manipulate public debate for strategic reasons, where fake accounts are paramount to the endeavour.[113] AI often plays a key role in CIB, and CIB must be considered in any conversation about the regulation of AI.

108. The Commission is concerned about the use of AI-generated engagement on social media to generate 'comments' on news articles, forums, or social media posts. This kind of CIB was a key element of Russia's Internet Research Agency, a St Petersburg-based 'troll farm', which was reportedly provided a $1.25 million USD monthly budget to interfere with the US 2016 presidential election.[114] Given the substantial cost of maintaining such CIB, it seems likely

that LLMs will provide actors with a more economic and faster way to implement such operations.

109. The use of CIB can also 'trick' social media and search engine trending algorithms by effectively spamming a topic – dictating what content is then suggested to users as trending.[115] Such an approach was allegedly used in 2022 to drown out online acts that were critical of China's COVID-19 lockdowns.[116]

110. There are a range of individual human rights potentially impacted by misinformation and disinformation on social media, including the right to freedom of expression,[117] right to privacy,[118] and the right to take part in public affairs.[119] Both misinformation and disinformation can have devastating effects on human rights, social cohesion and democratic processes. Indeed, this can be the very purpose intended by the release of disinformation.

111. Disinformation disseminates rapidly and inexpensively, which makes it a useful tool in online interference. The News and Media Research Centre identified three factors which exacerbate the spread of disinformation:

- digital networks play a central role in political communication
- the speed at which disinformation transmits on social media renders information attacks difficult to counter
- digital influence operations have low implementation costs.[120]

112. The Digital News Media Report: Australia 2022 highlights an overall downward trend in the use of social media as a source of news, which currently sits at 19% and is down four percentage points from last year.[121] However, for Generation Z (those born after 1997), 46% use social media as their main source of news (although this still represents an eight percentage point drop from last year).[122] This percentage is also higher for Generation Y (also known as 'Millennials' – born between 1981 and 1996), sitting at 28% – which represents a nine percentage point reduction in the past year.[123]

113. Although the consumption of news through social media has reduced since 2020,[124] there is still a high number of Australians – and particularly a high number of young Australians – who consume news through social media. These individuals are especially at risk of being influenced by disinformation presented as 'news'.

114. Social polarisation is often a goal of disinformation, as groups are pitted against one another.[125] This can often build upon, or amplify, existing tensions or divisions in a society. The Commission is increasingly disturbed by the role misinformation and disinformation plays in diminishing social

cohesion, promoting distrust and division, and undermining principles of equality, respect and human dignity.

115. While social media platforms use a mixture of AI and human investigators to address misinformation and disinformation, the Commission considers current efforts to be inadequate. Social media platforms have struggled to effectively combat the growing volumes of misinformation and disinformation, which can lead to the marginalisation and persecution of certain groups.

116. There are a range of existing laws that apply to social media (including the *Online Safety Act 2021* (Cth)), as well as a range of other policy measures adopted by government and by social media platforms themselves. For example, with respect to misinformation and disinformation specifically, the Department of Infrastructure, Transport, Regional Development, Communications and the Arts have opened submissions on the exposure draft of the *Communications Legislation Amendment (Combatting Misinformation and Disinformation) Bill 2023* (Cth).

117. Effectively combatting misinformation and disinformation on social media is important and needs to be given serious consideration by government, regulators and industry. However the exposure draft of the *Communications Legislation Amendment (Combatting Misinformation and Disinformation) Bill 2023* (Cth) also raises significant questions about ensuring that in combatting misinformation and disinformation, laws do not overreach and unduly restrict freedom of expression online.

118. To address the risk of misinformation and disinformation in the meantime, the Australian Government should establish a permanent whole-of-government taskforce dedicated to preventing and combating cyber-manipulation in Australia. The terms of reference for this taskforce should extend beyond those of the Electoral Integrity Assistance Taskforce to encompass not solely threats to the integrity of a federal election or electoral integrity, but threats to Australia's democracy and human rights of all people more broadly.

Recommendation 12: The Australian Government should establish a permanent whole-of-government taskforce dedicated to preventing and combating interference by way of cyber-manipulation in Australia.

119. The Australian Government should also establish clear and mandatory requirements and pathways for social media organisations to report

suspected interference activities. Such reports should be made to the proposed whole-of-government taskforce outlined above in Recommendation 12.

120. While acknowledging that this taskforce may be dealing with sensitive and protected information, it should be required – to the extent reasonably possible – to report publicly on the reports received and activities undertaken. The aim should be to bring greater transparency to the ways in which misinformation and disinformation are being addressed both to enhance the public understanding of the risks to Australia, and ensure that other rights and freedoms (most notably, freedom of expression) are not disproportionately impacted.

121. Striking the right balance between regulating online activities and protecting freedom of expression is an ongoing challenge. While there is a clear need to combat misinformation and disinformation online, there is also a risk that in doing so, different perspectives and controversial opinions may be targeted. While reasonable minds may differ on exactly where the line should be drawn, if Australia fails to ensure robust safeguards for freedom of expression online, then the very measures taken to combat misinformation and disinformation could themselves risk undermining Australia's democracy and values.

122. The guidance provided by the UN Human Rights Committee in General Comment No. 34, with respect to the permissible limitations on the right to freedom of expression, is particularly relevant here:

> when a State party imposes restrictions on the exercise of freedom of expression, these may not put in jeopardy the right itself. The Committee recalls that the relation between right and restriction and between norm and exception must not be reversed.[126]

123. There are also dangers inherent in allowing any one body – be it government, a government taskforce, or a social media platform – to become the sole arbiter of 'truth'. There is a real risk that efforts to combat online misinformation and disinformation could be used to legitimise attempts to restrict public debate, censor unpopular opinions and enforce ideological conformity in Australia. All efforts to combat misinformation and disinformation need to be accompanied by transparency and scrutiny safeguards to ensure that any limitations imposed upon freedom of expression are no greater than necessary and are strictly justified.

Recommendation 13: The Australian Government should establish clear and mandatory requirements, and pathways, for social media

organisations to report suspected misinformation and disinformation. Such reports should be made to the permanent taskforce noted above in Recommendation 12, whose activities in this area must incorporate robust safeguards to protect freedom of expression.

124. The Australian public can also play an important role in countering misinformation and disinformation on social media. Increasing digital literacy throughout the general community would help to ensure that the Australian population are better able to recognise and respond appropriately to the risks of misinformation and disinformation online, which would increase national resilience in respect of these risks.

125. The starting point here is to ensure that there is greater investment in incorporating digital literacy into the Australian education curriculum. This should include information about online safety, data privacy, identifying misinformation, and disinformation and the role that AI algorithms play in a users' online experience.

126. In addition to investment in the Australian curriculum, the Australian Government should introduce a public education campaign on digital literacy and develop online digital literacy resources that are available to the general public. The campaign and resources should include information and materials that enable Australians to better identify, and counter, misinformation and disinformation online. They should be tailored to different demographics and ensure accessibility for all Australians, with a particular focus on ensuring that the campaign and resources effectively engage with older people, people from culturally and linguistically diverse backgrounds, people from low-income backgrounds, people in regional and rural areas and people with disability.

Recommendation 14: The Commonwealth, state and territory governments should increase their investment in incorporating digital literacy into the Australian curriculum, including information about online safety, data privacy, identifying misinformation and disinformation and the role artificial intelligence algorithms play in a users' online experience.

Recommendation 15: The Australian Government should introduce a public education campaign on digital literacy and develop online digital literacy resources that are available to the general public.

127. The collection of personal data by social media platforms allows algorithms to tailor content to individual users. This personal information helps to create a user profile which allows social media companies to tailor the user experience, and sell targeted advertising.[127]

128. An unfortunate phenomenon of such targeted content is that users tend to be shown more, and gravitate towards, sensationalist 'clickbait'[128] – which often forms the basis of misinformation and disinformation on social media. This is due to a key aim of social media platforms being to maximise the time that users spend on their platform (which in turn increases advertising revenue potential). Accordingly, algorithms are incentivised to provide content which is meant to be more engaging for users. However, this material is often more extremist, sensationalist or plainly incorrect,[129] with algorithms having 'learnt' that such content generates greater engagement. It is by this process that inflammatory misinformation and disinformation is promoted by algorithms using microtargeted advertising, encouraging further user engagement and amplifying the reach of the content.[130] The algorithms appear to prioritise optimising user engagement and advertising revenue over the human rights and safety of users.

129. The harvesting of personal data for advertising purposes has significant implications in terms of privacy and also the ability to amplify the existing phenomena known as 'echo chambers'. An echo chamber is an online environment where a person only encounters information, or opinions, which reflect and reinforce their own worldviews.[131] These echo chambers can play a role, in conjunction with limited content moderation, in facilitating the spread of misinformation and disinformation, reinforcing hate speech and prejudicial content online and allowing for amplification of extremist views and conspiracy theories.

130. Only a minority of people truly understand the role that algorithms play in curating content shown to users on social media.[132] This can often make it difficult for users to escape online echo chambers, and highlights the need for greater education about how algorithms use personal data to tailor online experiences.[133]

131. The implementation of awareness campaigns, such as the Australian Electoral Commission's 'Stop and Consider' campaign in the lead-up to the 2019 Federal election, is a constructive example of how the Australian public can be encouraged to critically examine the content they see online.[134]

132. The digital literacy campaign and materials recommended above, at Recommendations 14 and 15, will assist in addressing these types of concerns.

3.7.2 Deepfakes

133. Deepfake content is often created using AI that draws upon an increasingly small number of photos or recordings of a person to model and create content. In recent years creating deepfake content has become cheaper, more efficient and increasingly accessible.[135]

134. Deepfakes posted online, especially via social media, have real world impacts both for those who are portrayed doing or saying things they would not otherwise do – as well as those who believe these images, videos or recordings to be true. Consider recent deepfake images of former President Donald Trump being tackled by officers during an 'arrest' which spurned significant media speculation,[136] or another deepfake image of an explosion at the Pentagon which resulted in the Dow Jones Industrial Index dropping 85 points (0.3 per cent) in four minutes.[137]

135. Deepfakes will facilitate new and emerging forms of cyber-enabled crime. For instance, it is becoming increasingly commonplace for scam calls to be made using LLMs and deepfake technologies to clone voices to elicit funds from unsuspecting victims who believe they are speaking with a loved one.[138] Further video and image-based deepfakes are prolific in the creation of nonconsensual pornography,[139] with an estimated 90% of deepfakes being pornographic in nature.[140]

136. Propaganda and disinformation will likely increase, as deepfakes lower the cost to entry, while also expanding the reach of content shared online.[141] In recent times propaganda has been generated by individuals in places such as China's '50-centres' and Russia's 'troll farms'.[142] However, the emergence of increasingly sophisticated and inexpensive technologies which can produce deepfake content may see humans removed from the process as AI-generated content is cheaper, faster and more effective in information warfare where scalability is essential.[143] Although information warfare has, until now, primarily been the domain of state actors – the availability of deepfake technology means that non-state actors will become increasingly active in this space.[144]

137. There are also concerns that deepfakes may have pertinent impact in high-stake decision making during military or international crises. The spreading of hyper-realistic deepfake images could adversely affect decision making where time is of the essence. For example, well timed deepfake content could allow actors to justify the incitement of violence against a marginalised

group, or even depict military personnel engaging in war crimes as justification for a violent military response.[145]

138. Deepfake content is complex and difficult to effectively police under Australia's patchwork of legislation. However, there are ways in which deepfake content can be combatted online.

Recommendation 16: The Australian Government should fund research and deployment of technologies which can detect deepfakes.

Recommendation 17: The Australian Government should work to improve digital literacy amongst Australia's population on what deepfakes are and how to spot deepfake content. This will require significant investment amongst school age children and young people as well older people and those from vulnerable groups.

Recommendation 18: The Department should review existing regulatory frameworks to assess whether they are capable of effectively combatting harmful deepfake content, and should consider introducing specific laws if regulatory gaps are identified.

3.8 Employment

139. There is a growing range of employers across the globe utilising AI products to help manage employment in the workplace – most notably in hiring and dismissal processes.[146]

140. For instance, ADM systems may unintentionally produce discrimination in the employee vetting process. For instance, Amazon used an AI software that was designed to review resumes and determine which applicants Amazon should hire.[147] The algorithm systemically discriminated against women applying for technical jobs, such as software engineer positions. This was because the existing pool of Amazon software engineers were by majority male, and as such, the new software was fed data about those engineers' resumes.[148] The practice of directing software to discover resumes similar to resumes in a training data set will inevitably reproduce the demographics of the existing workforce.[149]

141. AI-informed dismissal processes are also problematic. Cosmetics company Estee Lauder reached an out of court settlement with three make-up artists

who were dismissed during a redundancy exercise which utilised AI.[150] Amazon have also engaged in problematic dismissal processes informed by AI (its software monitors if workers are working fast enough and meeting quotas) as it has used an app to fire Flex Drivers.[151]

142. The *Fair Work Act 2009* (Cth) provides that in considering if a dismissal was harsh, unjust or unreasonable (in respect of unfair dismissal applications), the Fair Work Commission must take into account (amongst other things) whether:

- there was a valid reason for the dismissal
- the person was notified of that reason
- the person was given an opportunity to respond to that reason.[152]

143. Given the difficulties in AI and ADM processes being able to produce reasons (an issue which is further discussed below at [168]) the use of AI products in firing processes risks being odds with the *Fair Work Act 2009* (Cth) and potentially subverts natural justice for workers.

Recommendation 19: Business should not utilise artificial intelligence - informed dismissal processes unless the artificial intelligence product used can provide robust and genuine reasons in accordance with the *Fair Work Act 2009* (Cth) unfair dismissal regime.

144. There are additionally concerns that businesses and human resources professionals do not understand the impacts of utilising AI products in hiring or firing processes. Further guidance needs to be provided to workers and businesses of the potential risks associated with AI-informed hiring and firing processes and the human rights impacts.

145. Guidance materials should be developed on the risks and mitigation strategies of AI-informed hiring and firing processes.

3.9 Bias and algorithmic discrimination

146. AI allows large amounts of relevant information to be considered in decision-making processes, enabling ‘efficient’ decision making. However, regulation is increasingly important due to an algorithm’s potential to produce ‘algorithmic bias’. Algorithmic bias arises where an ADM tool produces outputs that result in unfairness.[153] Algorithmic bias can entrench unfairness, or even result in unlawful discrimination.[154]

147. For instance in 2019, a study discovered that a clinical algorithm used by many hospitals in the US to determine which patients required extra medical care produced racial bias.[155] The algorithm was trained on past data on healthcare spending, which reflects a trend whereby black patients have less income to spend on their healthcare as compared with white patients – a result of systemic wealth and income disparities.[156] As such, the algorithm's outputs reflected a discriminatory result whereby white patients required more medical care than black patients.[157]

148. This highlights why AI and ADM require greater regulation, in the interests of increasing transparency, preventing unfairness and unlawful discrimination in algorithmic decision-making. This is especially the case given the difficulty of applying existing anti-discrimination laws to complex ADM systems.[158]

149. The Commission's 2020 technical paper, 'Using artificial intelligence to make decisions: Addressing the problem of algorithmic bias', considers algorithmic bias in greater detail. Australia must do more to regulate the use of AI and ADM as a matter of priority, most notably in cases where the decisions made have a legal, or similarly significant, effect for individuals.

150. The use of discriminatory or biased AI products raises several issues under federal, state and territory anti-discrimination laws as well as questions about who bears liability where discrimination occurs due to biased algorithms.

3.10 People with disability

151. The Commission would draw the Department's attention to the Final Report in respect of past recommendations on how to mitigate AI risks for people with disability and ensure that there is a focus on accessibility when developing technology. This work is substantive and covers numerous issues that fall within the scope of the Discussion Paper.

Recommendation 20: The Department has regard for recommendations 24-38 included in the Final Report in respect of people with disability.

4 Artificial intelligence regulation

Do you have suggestions for possible regulatory action to mitigate these risks?

152. There are likely three pathways to regulating AI to mitigate notable human rights risks:

- create AI-specific legislation, analogous to other jurisdictions such as the EU
- reform and broaden existing regulation to ensure that it covers all applications of the technology
- or some combination of the two.

153. Regardless of which pathway is adopted, action is urgently required to mitigate the expanding risks associated with AI.

154. The Commission supports a combined approach to regulation. AI-specific legislation should be introduced, with the EU's proposed *Artificial Intelligence Act* providing one example of this type of approach, in addition to reviewing and updating existing legal frameworks.

155. It is expected that the responses to the Discussion Paper will identify regulatory gaps in protecting individuals from the harms of AI. It would be pertinent that where gaps are identified necessary reviews of the relevant legislation should be conducted, with input from relevant stakeholders. This could be similar to the Attorney-General's Department's *Privacy Act 1988* (Cth) Review Report, but would need to be a more expedited process given the immediacy to regulate AI.

Recommendation 21: Commonwealth, state and territory governments review relevant legislation to determine such legislation's applicability in regulating artificial intelligence. These reviews should be well resourced, consultative and conducted with urgency to ensure a timely response to the risks posed by artificial intelligence.

156. Given the unique and complex nature of AI-harms, modernising the existing legislative framework may still result in gaps, or not provide ample protection, where harms fall outside of the coverage of one or more pieces of legislation. It is for this reason that AI-specific legislation, like that proposed by the EU, should be adopted.

Recommendation 22: Australia should introduce specific legislation to address the risks of artificial intelligence, that are not already sufficiently addressed within the existing regulatory framework.

157. Any proposed Artificial Intelligence Act should not make unnecessary duplications or impose an overly complex framework by which business must

operate. Accordingly, the proposed Artificial Intelligence Act must have regard and oversight over legislative reviews aimed at modernising specific pieces of legislation in respect of AI.

Recommendation 23: The proposed Artificial Intelligence Act should not duplicate existing regulation or create unnecessary complexities for the development and use of artificial intelligence. The government body overseeing the proposed Artificial Intelligence Act must ensure it also has oversight of legislative reviews aimed at modernising specific pieces of legislation in respect of artificial intelligence.

5 Artificial intelligence safety commissioner

Are there any further non-regulatory initiatives the Australian Government could implement to support responsible AI practices in Australia? Please describe these and their benefits or impacts.

158. The Commission has recommended the appointment of an AI Safety Commissioner as an independent statutory office. This body would function as a source of expertise on AI, by providing guidance to government and the private sector on how to comply with laws surrounding the development and use of AI.[159]

159. The Commission considers that an AI Safety Commissioner could address three major needs:

First, government agencies and the private sector are often unclear on how to develop and use AI lawfully, ethically and in conformity with human rights. An AI Safety Commissioner could provide expert guidance on how to comply with laws and ethical standards that apply to the development and use of AI.

Secondly, regulators face the challenge of fulfilling their functions even as the bodies they regulate make important changes to how they operate. An AI Safety Commissioner could play a key role in building the capacity of existing regulators and, through them, of the broader ‘regulatory ecosystem’ to adapt and respond to the rise of AI.

Thirdly, legislators and policy makers are under unprecedented pressure to ensure Australia has the right law and policy settings to address risks and take opportunities connected to the rise of AI. An AI Safety Commissioner could monitor trends in the use of AI here and

> overseas. This would help it to be a source of robust, independent expertise.
>
> As an independent statutory office that champions the public interest, including human rights, an AI Safety Commissioner could help build public trust in the safe use of AI.[160]

Recommendation 24: The Federal government establish an AI Safety Commissioner as an independent statutory office, focused on promoting safety and protecting human rights in the development and use of artificial intelligence in Australia. The AI Safety Commissioner should:

- **work with regulators to build their technical capacity regarding the development and use of artificial intelligence in areas for which those regulators have responsibility**
- **monitor and investigate developments and trends in the use of artificial intelligence, especially in areas of particular human rights risk**
- **provide independent expertise relating to artificial intelligence and human rights for Australian policy makers**
- **issue guidance to government and the private sector on how to comply with laws and ethical requirements in the use of artificial intelligence.**

160. The Commission further notes recommendations 22 and 23 of its Final Report.[161] Recommendation 22 stated:

> The Australian Government should establish an AI Safety Commissioner as an independent statutory office, focused on promoting safety and protecting human rights in the development and use of AI in Australia. The AI Safety Commissioner should:
>
> (a) work with regulators to build their technical capacity regarding the development and use of AI in areas for which those regulators have responsibility
>
> (b) monitor and investigate developments and trends in the use of AI, especially in areas of particular human rights risk
>
> (c) provide independent expertise relating to AI and human rights for Australian policy makers

> (d) issue guidance to government and the private sector on how to comply with laws and ethical requirements in the use of AI.[162]

161. Moreover recommendation 23 continued that:

> The AI Safety Commissioner should:
>
> (a) be independent from government in its structure, operations and legislative mandate, but may be incorporated into an existing body or be formed as a new, separate body
>
> (b) be adequately resourced, wholly or primarily by the Australian Government
>
> (c) be required to have regard to the impact of the development and use of AI on vulnerable and marginalised people in Australia
>
> (d) draw on diverse expertise and perspectives including by convening an AI advisory council.[163]

162. A more detailed explanation of the role of an AI Safety Commissioner can be found at pages 125–135 of the Final Report.

163. The creation of such a body will take time. In the meantime, Australia must build upon the capacity of existing regulators to assist in the promotion of human rights-centred AI and ADM.

Recommendation 25: Until an AI Safety Commissioner is implemented, Australia must build the capacity of existing regulators, including by increasing funding, to better respond to the human rights risks of artificial intelligence.

5.1 Business and human rights

164. An AI Safety Commissioner would also help to emphasise the importance of human rights obligations on business in respect of AI. In particular, the AI Safety Commissioner's guidance to business could be grounded in existing business and human rights obligations such as the UN Guiding Principles on Business and Human Rights (UNGPs).

165. The UNGPs articulate human rights expectation on both States and businesses in preventing and mitigating impacts on human rights.[164] The UNGPs endeavour to address governance gaps and contain 31 principles which are housed within a three-pillar framework. Pillar two sets out the expectations that business has a responsibility to respect human rights and

provides several ways in which business can demonstrate this respect (e.g. due diligence frameworks, public policy commitments).[165]

166. Pillar two of the UNGPs places human rights obligations on business in respect of ethical development and deployment of AI tools. However, businesses would benefit from additional guidance with respect to the obligations relating to the use of AI, and best practice approaches.

167. An AI Safety Commissioner could provide such necessary guidance and even produce practical guidance on how to address bias and discrimination in algorithms for different industries similarly to the Commission's artificial intelligence (AI) and discrimination in insurance pricing and underwriting.

Recommendation 26: The AI Safety Commissioner should directly engage with the UN Guiding Principles on Business and Human Rights when providing guidance on human rights-centred artificial intelligence.

6 Transparency and the right to reasons

Where and when transparency will be most critical and valuable to mitigate potential AI risks and to improve public trust and confidence in AI? Mandating transparency requirements across the private and public sectors, including how these requirements could be implemented.

168. The Commission broadly supports a right for individuals to request meaningful information about substantially automated decisions. The Commission also considers the need for a broader right to request reasons in respect of substantially automated decisions. A right to reasons in this context will assist in promoting fairness and transparency in the use of AI in decision-making. Furthermore, the provision of reasons enables individuals who are the subject of ADM to exercise other rights, such as the right to object and the right to remedy.

6.1 Government automated decision-making

169. As a starting point, the Australian government should not make administrative decisions utilising AI if that AI product cannot provide reasons.

Recommendation 27: The Australian Government should not make administrative decisions using automation or artificial intelligence if the decision maker cannot generate reasons or a technical explanation for an affected person.

Recommendation 28: The Australian Government should make clear that, where a person has a legal entitlement to reasons for a decision, this entitlement exists regardless of how the decision is made. To this end, relevant legislation including s 25D of the *Acts Interpretation Act 1901* (Cth) should be amended to provide that:

- **for the avoidance of doubt, the term 'decision' includes decisions made using automation and other forms of artificial intelligence**
- **where a person has a right to reasons the person is entitled also to a technical explanation of the decision, in a form that could be assessed and validated by a person with relevant technical expertise**
- **the decision maker must provide this technical explanation to the person within a reasonable time following any valid request.**

Recommendation 29: The Australian Government should engage a suitable expert body to develop guidance for government and nongovernment bodies on how to generate reasons, including a technical explanation, for artificial intelligence -informed decisions.

170. Further information about the above three recommendations can be found at pages 62-67 of the Final Report.

171. Equally where an individual does receive reasons for an AI-informed government decision, it is important that there are means to challenge that decision.

Recommendation 30: The Australian Government should introduce legislation to create or ensure a right to merits review, generally before an independent tribunal such as the Administrative Appeals Tribunal, for any artificial intelligence -informed administrative decision.

172. Further information about the above recommendation can be found at pages 68-72 of the Final Report.

173. It is particularly important that government should always be able to explain how they arrive at decisions, in accordance with the principle of open government. As noted in the Final Report, the Commission opposes the use by government of ADM systems that cannot generate reasons, or a technical explanation, for any final decisions.[166] This is because government decisions will often inherently result in human rights impacts, and the principles of open government provide an important foundation for Australia's democratic system. The use by government of complex ADM systems, that cannot generate reasons, may leave individuals with no right to remedy.

174. These AI inferences and predictions are often the basis of decision-making both with and without human supervision which can have significant consequences for individuals. This is illustrated by the Government's 'Robodebt' scheme in 2015, whereby an automated debt recovery system used an algorithm to identify any discrepancies between an individual's declared income to the Australian Taxation Office, and the individual's income reported to Centrelink. A discrepancy was considered undeclared income, and as a result, a debt notice was automatically generated and sent to the individual.[167]

175. The Commission has previously made a submission to the Senate Community Affairs References Committee regarding its inquiry into 'Centrelink's compliance program'.[168] In that submission the Commission noted its concerns and highlighted the risk posed to the right to social security which is protected by art 9 of the International Covenant on Economic, Social and Cultural Rights – the impediment of which can impede the realisations of other human rights.[169]

176. A parliamentary inquiry has since revealed that this process resulted in various inaccurate debt notices. As the scheme involved social security payments, such errors disproportionately affected people with pre-existing socioeconomic disadvantage and vulnerability.[170] The Commonwealth Ombudsman, in its review of the scheme, urged the Department of Human Services to 'improve the clarity' of the letters sent to individuals, and to provide people 'better information so they understand the information and can properly respond to it'.[171]

177. As demonstrated in the subsequent Royal Commission, countless individuals suffered because of the scheme's algorithm. In just one example of the serious harms caused by the scheme, Kathleen Madgwick told the Royal Commission of her son, Jarrad Madgwick, who had taken his own life just hours after he learned of a $2,000 Centrelink Robodebt.[172] The scheme

demonstrates the dangers of the of utilising ADM systems which lack human scrutiny and where clear, understandable reasons cannot be provided for decisions that inherently impact a person's human rights.

178. Since the release of the Royal Commission's report Prime Minister Anthony Albanese has stated that:

> The Robodebt scheme was a gross betrayal and a human tragedy, … It pursued debt recovery against Australians who in many cases had no debt to pay. … It was wrong. It was illegal. It should never have happened and it should never happen again.[173]

179. Given the findings of the Royal Commission there is an obvious and pressing need to ensure the government's use of AI is ethical, a starting point is that individuals effected by government ADM must be notified.

Recommendation 31: The Australian Government should introduce legislation to require that any affected individual is notified where artificial intelligence is materially used in making an administrative decision. That notification should include information regarding how an affected individual can challenge the decision.

180. This recommendation is predicated upon the information contained on pages 60-62 of the Final Report.

181. However, it is also important to consider how the government is already using ADM. Accordingly an audit across all jurisdictions is necessary.

Recommendation 32: The Australian Government should commission an audit of all current uses of artificial intelligence informed decision making by or on behalf of Government agencies. The AI Safety Commissioner, or another suitable expert body, should conduct this audit.

182. This recommendation is predicated upon the information contained on pages 60–62 of the Final Report.

6.2 Private enterprise automated decision-making

183. The Commission noted in its Final Report that a human rights approach to the regulation of ADM requires access to an effective remedy where an

individual's human rights have been breached.[174] However access to remedy is often predicated on an ability to understand and challenge a decision made – in respect of AI, that is not always an easy task.

184. Regardless of the difficulties, private industry must also strive to embed explicability into AI products. Information provided about an AI decision must be conveyed in a clear, understandable format in order to allow individuals to properly respond. In the Commission's Final Report, stakeholders warned that simply providing the technical basis for AI informed decisions may do little to assist individuals to understand or challenge those decisions.[175]

185. While the Commission considers that the requirement for a right to reasons is currently more pressing than it was at the time of our Final Report due to recent breakthroughs in AI capabilities, the Commission also acknowledges the difficulties surrounding the introduction of such a requirement.

186. It is technically difficult for some ADM systems to generate reasons. The use of AI may obscure the rationale or reasons for a decision – referred to as the problem of 'black box' or 'opaque' AI.[176] This, in turn, can make it difficult or even impossible to challenge the merits or lawfulness of a decision.[177] The use of black box AI (or opaque AI) may infringe upon human rights in whatever sector it arises, whether that be in government, the private or the non-government sector.[178]

187. While some leading software companies are exploring building an explanation function into ADM systems, this process can be technically challenging and expensive.[179] However, with the immense take up of AI products and its use in ADM this is an expense and challenge which must be overcome. Substantially automated decisions which result in a significant outcome for an individual must have explicability features built into that AI product.

Recommendation 33: Artificial intelligence-informed products which result in a legal, or similarly significant, effect on an individual must have explicability functions built into those products.

188. The ability of small and medium-sized entities to generate meaningful information may be hindered by the financial cost of extracting a useful explanation (particularly in complex ADM systems) and the time it would take for an organisation to generate an explanation.[180] It may be possible to overcome this difficulty if further research is conducted by centres of

expertise on explainable AI and expert guidance is provided by government on how to provide reasons for AI-informed decisions, as recommendation by the Commission in its Final Report.[181]

189. It is noted that there may not necessarily exist a legal entitlement to the provision of reasons in the current legislative environment. For instance, the Commission noted in in its Final Report that decisions by non-government bodies do not carry a legal entitlement to reasons.[182] However, as noted above at [3.8] this presumption is increasingly being challenged in employment where AI can, and has been, used to dismiss employees. In Australia, such an application is directly challengeable as employees are entitled to reasons for their dismissal which the Fair Work Commission will consider in determining if a dismissal was harsh, unjust or unreasonable in accordance with s 387 of the *Fair Work Act 2009* (Cth).

190. The Commission considers that in light of recent technological developments, a right to request reasons is more pressing than ever. However significant work is required to include explicability functions to protect human rights.

Recommendation 34: The obligation to include explicability functions into automated decision-making should fall onto the organisations which design artificial intelligence products.

Recommendation 35: When providing artificial intelligence products, which must include explicability functions, the purchasing entity must be provided with clear and understandable instructions on how reasons can be produced using the AI product. It must not become a 'set and forget' feature.

Recommendation 36: The Australian Government should introduce legislation to require that any affected individual is notified when a corporation or other legal person materially uses artificial intelligence in a decision making process that affects the legal, or similarly significant, rights of the individual.

191. The above recommendation is predicated on the information contained on pages 77–78 of the Final Report.

192. The Commission would also recommend the following which has been drawn from pages 78–83 of the Final Report.

Recommendation 37: The Australian Government should introduce legislation that provides a rebuttable presumption that, where a corporation or other legal person is responsible for making a decision, that legal person is legally liable for the decision regardless of how it is made, including where the decision is automated or is made using artificial intelligence.

Recommendation 38: The Australian Government should introduce legislation to provide that where a court, or regulatory, oversight or dispute resolution body, has power to order the production of information or other material from a corporation or other legal person:

- **for the avoidance of doubt, the person must comply with this order even where the person uses a form of technology, such as artificial intelligence, that makes it difficult to comply with the order**
- **if the person fails to comply with the order because of the technology the person uses, the body may draw an adverse inference about the decision-making process or other related matters.**

7 Facial recognition technologies

Do you have suggestions for Whether any high-risk AI applications or technologies should be banned completely? Criteria or requirements to identify AI applications or technologies that should be banned, and in which contexts?

193. Facial recognition technology (FRT) utilises algorithms and AI. It can be used in simple ways, such as to unlock a phone. However, it can also be used in policing or decisions which have a legal or similarly significant effect on an individual. The Commission made various recommendations with respect to the use of FRT in the Final Report and building upon those recommendations, encourages several recommendations be adopted below.[183]

194. New and emerging technologies often bring with them a range of ethical issues as society grapples with how best to harness the prospective benefits of new technology, while mitigating the potential harms. This is especially true of FRT, which has had persistent problems with accuracy and fairness in

its use – particularly in respect of racial and gender bias. These concerns have led to the technology being banned in some places, and yet it continues to be commonplace in others.[184]

195. FRT is being adopted by government and businesses in Australia at an exponential rate.[185] These tools are also increasingly being used in workplaces, schools, shopping centres and residential areas to identify members of the public and monitor behaviour.[186] CHOICE has also recently revealed that FRT is being used across multiple sporting stadiums in Australia, with it being reported that some owners and operators are not being transparent with attendees or media about its use.[187]

196. As the technology has become increasingly mainstream, so too have the voices raising ethical concerns and calling for greater regulation.[188] All new and emerging technologies need to be used in a responsible and ethical way, and need a code of ethics and regulation to mitigate any harms.[189] The Commission encourages greater discussion of the limitations of FRT and how developers can better manage those limitations to increase equity and fairness.[190] Regulation and legislation are only one strategy to handling risk, and not the answer to fundamental issues in the technology itself.

197. As noted below, there are substantial risks associated with the use of FRT. What is not known is how extensively the technology is being used.

Recommendation 39: Federal, state and territory governments should conduct an audit into the use of facial recognition technologies by government agencies.

198. Handling the risks of FRT requires a prudent approach, as the potential benefits of the technology must be measured against its potential harms. While the technology has the potential to improve public services and law enforcement (i.e. traffic congestion, pollution controls and public security), it can also be used for mass surveillance, ethnic profiling, targeted repression and privacy violations.[191]

199. As of 2019, at least 64 countries were identified as actively using some type of FRT scheme for surveillance purposes.[192] FRT can be an attractive investment for many private and public organisations, as it decreases the time, effort and money needed to identify faces and tie those faces to other information (such as other pieces of personal data about an individual).[193] However, organisations and government must be cautious when considering the use of FRT and the risks that attach to this.

200. India's use of FRT is just one example of the duality that is inherent within this technology. In 2018 Delhi police used FRT to reunite nearly 3,000 children with their parents in just four days.[194] This pilot FRT programme had later reunited 10,561 missing children with their families after only 15 months in operation.[195] The profoundly positive impact this technology can have is astounding, as it can identify and match faces using one-to-many technology faster than any human is capable of. This program is one example of the potential of FRT to be used in ways that enhances human rights.[196]

201. However, there have also been criticisms of the Indian government using this same FRT technology in 2020 to facilitate the arrest of protesters of a citizenship law which critics claimed marginalises Muslims.[197]

202. Examples of 'function creep', where FRT is applied beyond the initially intended purpose, can be found globally – most notably when it is used against marginalised populations,[198] such as the Muslim Uyghur minorities in China's Xinjiang Uyghur Autonomous Region.[199] The 2022 report by the Office of the UN High Commissioner for Human Rights that focused on human rights concerns in this region described 'an ever-present network of surveillance cameras, including deploying facial recognition capabilities' as one element of 'what has been alleged to be a sophisticated, large-scale and systematized surveillance system in practice'.[200]

203. It is likely due to the duality of FRT, which is largely unregulated, that individuals globally vary on their acceptance of the technology. For example, an online survey conducted across four countries in 2019 found that while 51% of Chinese respondents were strongly or somewhat accepting of FRT for public use, this dropped to only 37% of Americans and 38% of Germans.[201]

204. Acceptance rates of FRT may be positively influenced by factors such as:

- trust in the government
- concerns about specific risks, such as terrorism
- high levels of technological affinity in a population.[202]

205. Conversely, awareness of a country's adverse use of surveillance methods in the past (and concerns in respect of privacy violations) foster a more apprehensive attitude towards FRT in public settings.[203]

206. Domestically, individuals are also concerned about the use of FRT. In a nationally representative survey, CHOICE asked respondents about the use of FRT in retail stores in Australia. 65% of respondents were concerned about stores using technology to create customer profiles which could cause them harm, while a further 78% expressed concern about the secure storage of faceprint data.[204]

207. A subsequent investigation into retailers using FRT led to the OAIC launching an investigation into both Kmart and Bunnings' use of FRT,[205] while the Good Guys chain has paused its use of FRT in stores while the OAIC investigates a complaint made by CHOICE.[206]

208. Without FRT-specific regulation, such as that proposed by the Human Technology Institute's Model Law (Model Law),[207] it is difficult to imagine circumstances where individuals will be trusting of FRT to the point that all of its benefits can be appreciated without posing a disproportionate risk to human rights. There are undoubtedly benefits to the technology, as highlighted above with the example from India, but regulation is needed to harness these advantages in a human rights' compliant manner.

209. While it is generally undesirable to regulate a specific technology, there are exceptions to this general principle. For example, as was highlighted in the Final Report, governments have a tendency to regulate technology deemed high-risk, which helps to explain the comparatively strict laws which govern fields such as gene technology, aviation, healthcare and the energy industry.[208] It is in these areas that regulation is often applied to both the technologies themselves as well as their use. In relation to FRT, the greater the risk to human rights, the greater the need for regulation.

Recommendation 40: Federal, state and territory governments should introduce legislation which specifically regulates the use of facial recognition and other biometric technologies. Such legislation should:

- **expressly protect human rights**
- **apply to the use of this technology in decision making that has a legal, or similarly significant, effect for individuals, or where there is a high risk to human rights, such as in policing and law enforcement**
- **be developed through in-depth consultation with the community, industry and expert bodies such as the Australian Human Rights Commission and the Office of the Australian Information Commissioner.**

210. The Commission provides in-principle support for the Model Law on FRT proposed by the Human Technology Institute.[209]

Recommendation 41: The government consult with community, industry and expert bodies, such as the Australian Human Rights Commission and the Office of the Australian Information Commissioner, with a view to implementing the Model Law.

211. The length of time it will necessarily take to implement FRT regulation means that there is a continuing risk of significant human rights harms being facilitated in the meantime by FRT in both public and private spheres.

212. More broadly, the Commission's Final Report highlighted concerns expressed around three particular risks:

- the contribution of FRT to the growth in surveillance
- the use of data derived from FRT to engage in profiling
- the risk that errors connected to facial recognition disproportionately affect certain groups.[210]

213. This is in addition to the use of FRT in the private sector which:

> raises distinct concerns as there may be a lower degree of accountability and fewer legal protections.[211]

214. In respect of the growing use of FRT-enabled surveillance, the Commission previously found that this would lead to an inevitable reduction of personal privacy, and that the threat of closer security by police and government agencies can impede participation in lawful democratic processes – such as protests and meetings.[212] This raises the risk profile in protecting the rights to:

- freedom of association and assembly
- freedom of expression and opinion
- freedom from unlawful and arbitrary arrest.[213]

215. Moreover, the Commission has previously raised concerns about the 'mosaic effect'.[214] With the inclusion of additional biometric (such as the information collected by FRT) and non-biometric information, this can allow sensitive personal information to be extracted or inferred about a person, including their age, race, sex and health.[215]

216. Such information and inferences can be used in profiling – where intrusive action is taken by reference to people's characteristics. An example of this kind of profiling, which may result in people of a particular racial or ethnic group being disproportionately subjected to police identity checks, has been highlighted by Human Rights Watch in the report, *China's Algorithms of*

Repression: Reverse Engineering a Xinjiang Police Mass Surveillance App, which provided a detailed analysis of the technology used for mass surveillance in Xinjiang, including the aggregation of data.[216]

217. The above risks are further exacerbated by the potential for errors in the technology, as risks are at their highest where this technology is used in decision making that affects an individual's legal or similarly significant rights. This is most obvious when the technology fails.

218. For example, if an error in FRT on a smartphone causes a delay in an individual unlocking their device, generally this would present little more than an annoyance. However, if a person is wrongly accused of a crime because of an error in police use of FRT, the risk of harm is far greater. There have been examples reported where individuals have been falsely arrested and imprisoned due to identification using FRT.[217]

219. Generally speaking, FRT is far from perfect and is often criticised as being less accurate when identifying women, or people from minority racial groups, as compared with other people.[218] Amazon, Microsoft, and IBM have all previously announced they would stop, or pause, offering this technology to law enforcement based upon concerns of this nature.[219]

220. However, as there is currently no legislation regulating FRT, nor a moratorium in place in the interim, others have continued to facilitate the use of FRT by government agencies and police forces globally.

221. An example of the risks that this poses for human rights can be seen in the illustrative example of the activities of Clearview AI, who scraped approximately 3 billion images of faces from publicly accessible sources (such as Facebook and Google) to create a database. The company then licensed this database to over 600 law enforcement agencies (in addition to banks, private companies and schools).[220] Reports have shown that employees at law enforcement agencies in the US were running thousands of Clearview AI facial recognition searches – often without the public's knowledge or consent.[221]

222. While regulating a specific kind of technology may result in delays on its uptake, or the realisation of economic benefits, there are often good reasons to do so where that technology is deemed high-risk. FRT is a technology which poses an unacceptable risk to human rights without regulation.

Recommendation 42: Until the legislation recommended in Recommendation 40 comes into effect, Australia's federal, state and territory governments should introduce a moratorium on the use of

facial recognition and other biometric technology in decision making that has a legal, or similarly significant, effect for individuals, or where there is a high risk to human rights, such as in policing and law enforcement.

223. This moratorium would not apply to all uses of facial and biometric technology. It would apply only to uses of such technology to make decisions that affect legal or similarly significant rights, unless and until specific legislation is introduced with effective human rights safeguards.

224. The Commission is not alone in recommending a moratorium on the use of FRT. For example, in June 2020 the *Facial Recognition and Biometrics Technology Moratorium Act* was introduced into US Congress. In March 2023, US senators reintroduced that same act in response to reports that US law enforcement agencies have used unregulated FRT, in addition to research indicating that approximately half of the adult US population are already in facial recognition databases.[222]

8 Human Rights Impact Assessments

What initiatives or government action can increase public trust in AI deployment to encourage more people to use AI?

225. As noted above the introduction of an AI Safety Commissioner should be implemented as a non-regulatory approach to AI. It is likely that the creation of an AI Safety Commissioner would increase public trust in AI.

226. Another initiative that the government can take to increase public trust is by requiring organisations which develop and train AI products to conduct human rights impact assessments (HRIA).

227. HRIA tools assess how a new product, service, law or policy will engage human rights. They also provide a framework for ensuring adequate rights protections.

228. As noted in the Final Report:

> HRIAs are increasingly being used by government, the private sector and civil society organisations to measure the risk to human rights posed by their activities, ensure that measures are put in place to address human rights risks, and support the availability of remedies for any human rights infringements.[223]

229. The Commission's previous work has found strong support from the public and private sectors, for the Australian Government to develop an HRIA tool

and associated guidance for AI-informed decision making.[224] Further information on HRIAs can be found on pages 98–99 of the Final Report.

230. It is also of note that such due diligence processes are in line with the recently updated OECD Guidelines for Multinational Enterprises on Responsible Business Conduct in respect of actual and potential adverse impacts related to science, technology and innovation.[225]

231. To increase public trust in the government's use of AI, the Commission would also reiterate the following recommendation from the Final Report.

Recommendation 43: The Australian Government should introduce legislation to require that a human rights impact assessment be undertaken before any department or agency uses an artificial intelligence informed decision-making system to make administrative decisions.

An HRIA should include public consultation, focusing on those most likely to be affected. A human rights impact assessment should assess whether the proposed artificial intelligence-informed decision-making system:

- **complies with Australia's international human rights law obligations**
- **will involve automating any discretionary element of administrative decisions, including by reference to the Commonwealth Ombudsman's Automated decision making better practice guide and other expert guidance**
- **provides for appropriate review of decisions by human decision makers**
- **is authorised and governed by legislation.[226]**

232. There are significant risks associated with AI and automation which require careful and considered planning before being adopted by government for administrative decision making. [227] Further information about the above recommendation can be found at pages 55-59 of the Final Report.

233. In the Commission's Final Report, it was recommended that the AI Safety Commissioner should develop a tool to assist private sector bodies undertake HRIAs in developing AI-informed decision-making systems. This included recommending that the Australian Government should maintain a public register of completed HRIAs.[228]

234. However, due to the significant advancements in AI in the years since the Final Reports release, the Commission must depart from the recommendation that the creation of HRIA tools for AI should be left to an AI Safety Commissioner, as more immediate action is required.

235. The Commission would also reiterate the following recommendations from the Final Report which can be found on pages 95-109.

Recommendation 44: The Australian Government should convene a multi-disciplinary taskforce on artificial intelligence-informed decision making, led by an independent body, such as the AI Safety Commissioner. The taskforce should:

- **promote the use of human rights by design in this area**
- **advise on the development and use of standards and certification schemes**
- **advise on the development of one or more regulatory sandboxes focused on upholding human rights in the use of artificial intelligence -informed decision making.**

The taskforce should consult widely in the public and private sectors, including with those whose human rights are likely to be significantly affected by artificial intelligence -informed decision making.

Recommendation 45: The Australian Government should adopt a human rights approach to procurement of products and services that use artificial intelligence. The Department of Finance, in consultation with the Digital Transformation Agency and other key decision makers and stakeholders, should amend current procurement law, policy and guidance to require that human rights are protected in the design and development of any artificial intelligence -informed decision-making tool procured by the Australian Government.

Recommendation 46: The Australian Government should engage an expert body, such as the AI Safety Commissioner or the Australian Human Rights Commission, to issue guidance to the private sector on good practice regarding human review, oversight and monitoring of artificial intelligence-informed decision-making systems. This body should also advise the Government on ways to incentivise such good

practice through the use of standards, certification schemes and government procurement rules.

Recommendation 47: The Australian Government should resource the Australian Human Rights Commission to produce guidelines for government and non-government bodies on complying with federal anti-discrimination laws in the use of artificial intelligence-informed decision-making.

9 Recommendations

236. The Commission makes the following recommendations.

Recommendation 1

Government should consider alternative models of privacy and data protection models which do not place the primary onus on individuals to protect their data.

Recommendation 2

Privacy Act 1988 (Cth) proposed reforms should be adopted in respect of artificial intelligence and automated decision-making. Any legislative amendments should ensure a human rights-compliant approach to data protection.

Recommendation 3

The Department should consider artificial intelligence in a broader context to ensure that its interoperability with other technologies (such as neurotechnologies) is given appropriate attention.

Recommendation 4

To better understand whether specific policy responses are needed in respect of artificial intelligence-driven neurotechnology, a review of the current regulatory landscape and the human rights risks of neurotechnology should be conducted.

Recommendation 5

The Department should consider artificial intelligence in a broader context to ensure that its interoperability with other technologies (such as metaverse technologies) is given appropriate attention.

Recommendation 6

To better understand the risks of artificial intelligence in the metaverse, the government should engage an independent statutory body to produce a report on the human rights risks of metaverse and extended reality technologies.

Recommendation 7

The government should ensure that consumer-oriented artificial intelligence chatbots have robust safeguards in place to ensure protections for users.

Recommendation 8

Safeguards in place to protect users from consumer-oriented artificial intelligence chatbots should be intensely tested with different interactions over a prolonged period to ensure such artificial intelligence products do not produce harmful responses.

Recommendation 9

The government should develop specific regulation to ensure harmful responses by consumer-oriented artificial intelligence chatbots are not provided to users.

Recommendation 10

Organisations which train and deploy artificial intelligence products should report on the environmental impact of their work.

Recommendation 11

There should be greater investment in training both government and private enterprise on the limitations of artificial intelligence products and how to better scrutinise artificial intelligence -informed decisions or recommendations.

Recommendation 12

The Australian Government should establish a permanent whole-of-government taskforce dedicated to preventing and combating interference by way of cyber-manipulation in Australia.

Recommendation 13

The Australian Government should establish clear and mandatory requirements, and pathways, for social media organisations to report suspected misinformation and disinformation. Such reports should be made to the permanent taskforce noted above in Recommendation 12, whose activities in this area must incorporate robust safeguards to protect freedom of expression.

Recommendation 14

The Commonwealth, state and territory governments should increase their investment in incorporating digital literacy into the Australian curriculum, including information about online safety, data privacy, identifying misinformation and disinformation and the role artificial intelligence algorithms play in a users' online experience.

Recommendation 15

The Australian Government should introduce a public education campaign on digital literacy and develop online digital literacy resources that are available to the general public.

Recommendation 16

The Australian Government should fund research and deployment of technologies which can detect deepfakes.

Recommendation 17

The Australian Government should work to improve digital literacy amongst Australia's population on what deepfakes are and how to spot deepfake content. This will require significant investment amongst school age children and young people as well older people and those from vulnerable groups.

Recommendation 18

The Department should review existing regulatory frameworks to assess whether they are capable of effectively combatting harmful deepfake content, and should consider introducing specific laws if regulatory gaps are identified.

Recommendation 19

Business should not utilise artificial intelligence -informed dismissal processes unless the artificial intelligence product used can provide robust and genuine reasons in accordance with the *Fair Work Act 2009* (Cth) unfair dismissal regime.

Recommendation 20

The Department has regard for recommendations 24-38 included in the Final Report in respect of people with disability.

Recommendation 21

Commonwealth, state and territory governments review relevant legislation to determine such legislation's applicability in regulating artificial intelligence. These reviews should be well resourced, consultative and conducted with urgency to ensure a timely response to the risks posed by artificial intelligence.

Recommendation 22

Australia should introduce specific legislation to address the risks of artificial intelligence, that are not already sufficiently addressed within the existing regulatory framework.

Recommendation 23

The proposed Artificial Intelligence Act should not duplicate existing regulation or create unnecessary complexities for the development and use of artificial intelligence. The government body overseeing the proposed Artificial Intelligence Act must ensure it also has oversight of legislative reviews aimed at modernising specific pieces of legislation in respect of artificial intelligence.

Recommendation 24

The Federal government establish an AI Safety Commissioner as an independent statutory office, focused on promoting safety and protecting human rights in the development and use of artificial intelligence in Australia. The AI Safety Commissioner should:

- work with regulators to build their technical capacity regarding the development and use of artificial intelligence in areas for which those regulators have responsibility
- monitor and investigate developments and trends in the use of artificial intelligence, especially in areas of particular human rights risk
- provide independent expertise relating to artificial intelligence and human rights for Australian policy makers
- issue guidance to government and the private sector on how to comply with laws and ethical requirements in the use of artificial intelligence.

Recommendation 25

Until an AI Safety Commissioner is implemented, Australia must build the capacity of existing regulators, including by increasing funding, to better respond to the human rights risks of artificial intelligence.

Recommendation 26

The AI Safety Commissioner should directly engage with the UN Guiding Principles on Business and Human Rights when providing guidance on human rights-centred artificial intelligence.

Recommendation 27

The Australian Government should not make administrative decisions using automation or artificial intelligence if the decision maker cannot generate reasons or a technical explanation for an affected person.

Recommendation 28

The Australian Government should make clear that, where a person has a legal entitlement to reasons for a decision, this entitlement exists regardless of how the decision is made. To this end, relevant legislation including s 25D of the *Acts Interpretation Act 1901* (Cth) should be amended to provide that:

- for the avoidance of doubt, the term 'decision' includes decisions made using automation and other forms of artificial intelligence
- where a person has a right to reasons the person is entitled also to a technical explanation of the decision, in a form that could be assessed and validated by a person with relevant technical expertise
- the decision maker must provide this technical explanation to the person within a reasonable time following any valid request.

Recommendation 29

The Australian Government should engage a suitable expert body to develop guidance for government and nongovernment bodies on how to generate reasons, including a technical explanation, for artificial intelligence -informed decisions.

Recommendation 30

The Australian Government should introduce legislation to create or ensure a right to merits review, generally before an independent tribunal such as the Administrative Appeals Tribunal, for any artificial intelligence -informed administrative decision.

Recommendation 31

The Australian Government should introduce legislation to require that any affected individual is notified where artificial intelligence is materially used in making an administrative decision. That notification should include information regarding how an affected individual can challenge the decision.

Recommendation 32

The Australian Government should commission an audit of all current uses of artificial intelligence informed decision making by or on behalf of Government agencies. The AI Safety Commissioner, or another suitable expert body, should conduct this audit.

Recommendation 33

Artificial intelligence-informed products which result in a legal, or similarly significant, effect on an individual must have explicability functions built into those products.

Recommendation 34

The obligation to include explicability functions into automated decision-making should fall onto the organisations which design artificial intelligence products.

Recommendation 35

When providing artificial intelligence products, which must include explicability functions, the purchasing entity must be provided with clear and understandable instructions on how reasons can be produced using the AI product. It must not become a 'set and forget' feature.

Recommendation 36

The Australian Government should introduce legislation to require that any affected individual is notified when a corporation or other legal person materially uses artificial intelligence in a decision making process that affects the legal, or similarly significant, rights of the individual.

Recommendation 37

The Australian Government should introduce legislation that provides a rebuttable presumption that, where a corporation or other legal person is responsible for making a decision, that legal person is legally liable for the decision regardless of how it is made, including where the decision is automated or is made using artificial intelligence.

Recommendation 38

The Australian Government should introduce legislation to provide that where a court, or regulatory, oversight or dispute resolution body, has power to order the production of information or other material from a corporation or other legal person:

- for the avoidance of doubt, the person must comply with this order even where the person uses a form of technology, such as artificial intelligence, that makes it difficult to comply with the order
- if the person fails to comply with the order because of the technology the person uses, the body may draw an adverse inference about the decision-making process or other related matters.

Recommendation 39

Federal, state and territory governments should conduct an audit into the use of facial recognition technologies by government agencies.

Recommendation 40

Federal, state and territory governments should introduce legislation which specifically regulates the use of facial recognition and other biometric technologies. Such legislation should:

- expressly protect human rights
- apply to the use of this technology in decision making that has a legal, or similarly significant, effect for individuals, or where there is a high risk to human rights, such as in policing and law enforcement
- be developed through in-depth consultation with the community, industry and expert bodies such as the Australian Human Rights Commission and the Office of the Australian Information Commissioner.

Recommendation 41

The government consult with community, industry and expert bodies, such as the Australian Human Rights Commission and the Office of the Australian Information Commissioner, with a view to implementing the Model Law.

Recommendation 42

Until the legislation recommended in Recommendation 40 comes into effect, Australia's federal, state and territory governments should introduce a moratorium on the use of facial recognition and other biometric technology in decision making that has a legal, or similarly significant, effect for individuals, or where there is a high risk to human rights, such as in policing and law enforcement.

Recommendation 43

The Australian Government should introduce legislation to require that a human rights impact assessment be undertaken before any department or agency uses an artificial intelligence informed decision-making system to make administrative decisions.

An HRIA should include public consultation, focusing on those most likely to be affected. A human rights impact assessment should assess whether the proposed artificial intelligence-informed decision-making system:

- complies with Australia's international human rights law obligations
- will involve automating any discretionary element of administrative decisions, including by reference to the Commonwealth Ombudsman's

Automated decision making better practice guide and other expert guidance

- provides for appropriate review of decisions by human decision makers
- is authorised and governed by legislation.

Recommendation 44

The Australian Government should convene a multi-disciplinary taskforce on artificial intelligence-informed decision making, led by an independent body, such as the AI Safety Commissioner. The taskforce should:

- promote the use of human rights by design in this area
- advise on the development and use of standards and certification schemes
- advise on the development of one or more regulatory sandboxes focused on upholding human rights in the use of artificial intelligence - informed decision making.

The taskforce should consult widely in the public and private sectors, including with those whose human rights are likely to be significantly affected by artificial intelligence-informed decision making.

Recommendation 45

The Australian Government should adopt a human rights approach to procurement of products and services that use artificial intelligence. The Department of Finance, in consultation with the Digital Transformation Agency and other key decision makers and stakeholders, should amend current procurement law, policy and guidance to require that human rights are protected in the design and development of any artificial intelligence - informed decision-making tool procured by the Australian Government.

Recommendation 46

The Australian Government should engage an expert body, such as the AI Safety Commissioner or the Australian Human Rights Commission, to issue guidance to the private sector on good practice regarding human review, oversight and monitoring of artificial intelligence-informed decision-making systems. This body should also advise the Government on ways to incentivise such good practice through the use of standards, certification schemes and government procurement rules.

Recommendation 47

The Australian Government should resource the Australian Human Rights Commission to produce guidelines for government and non-government bodies on complying with federal anti-discrimination laws in the use of artificial intelligence-informed decision-making.

Endnotes

[1] eSafety Commissioner, '*Deepfake trends and challenges — position statement*' (website) <https://www.esafety.gov.au/industry/tech-trends-and-challenges/deepfakes>.

[2] Organisation for Economic Co-operation and Development (OECD), '*Recommendation Responsible Innovation in Neurotechnology*' (OECD Legal Instrument, 2019); OECD, '*Neurotechnology and Society: Strengthening Responsible Innovation in Brain Science*' (policy papers, November 2017) 49.

[3] United Nations Educational, Scientific and Cultural Organization (UNESCO), '*Report of the International Bioethics Committee of UNESCO (IBC) on the Ethical Issues of Neurotechnology*' (Report, 2021) 5.

[4] The Neurorights Foundation, '*Market Analysis Neurotechnology*' (Report, March 2023) 3.

[5] The Neurorights Foundation, '*Market Analysis Neurotechnology*' (Report, March 2023) 3.

[6] Allan McCay, 'Neurotechnology, Law and the Legal Profession' (Horizon Report for The Law Society, August 2022) *The Law Society of England and Wales* 4; The Neurorights Foundation, '*Market Analysis Neurotechnology*' (Report, March 2023) 3.

[7] The Neurorights Foundation, '*Market Analysis Neurotechnology*' (Report, March 2023) 5.

[8] The Neurorights Foundation, '*Market Analysis Neurotechnology*' (Report, March 2023) 14.

[9] Electoral Integrity Assurance Taskforce, *Disinformation and Misinformation Factsheet* <eiat-disinformation-factsheet.pdf (aec.gov.au)>.

[10] See XR Safety Initiative, '*The Metaverse*' (website) <https://xrsi.org/definition/the-metaverse>.

[11] Office of the Australian Information Commissioner (OAIC), '*What is Privacy?*' (Website) <https://www.oaic.gov.au/privacy/your-privacy-rights/what-is-privacy>.

[12] See *Convention on the Rights of the Child* art 16; *International Convention on the Protection of the Rights of All Migrant Workers and Members of Their Families art 14*; *Convention on the Rights of Persons with Disabilities* art 22; *African Charter on the Rights and Welfare of the Child* art 10; *American Convention on Human Rights* art 11; *Convention for the Protection of Human Rights and Fundamental Freedoms* art 8.

[13] Office of the United Nations High Commissioner for Human Right, '*The Right to Privacy In the Digital Age*' (Report, A/HRC/51/17, 04 August 2022) 2.

[14] Office of the United Nations High Commissioner for Human Right, '*The Right to Privacy In the Digital Age*' (Report, A/HRC/51/17, 04 August 2022) 4.

[15] Office of the United Nations High Commissioner for Human Right, '*The Right to Privacy In the Digital Age*' (Report, A/HRC/51/17, 04 August 2022) 4.

[16] Wolflie Christl, *Corporate surveillance in everyday life* (Vienna, Cracked Lab – Institute for Critical Digital Culture, 2017).

[17] Office of the United Nations High Commissioner for Human Right, '*The Right to Privacy In the Digital Age*' (Report, A/HRC/51/17, 04 August 2022) 4.

[18] G. Bell et al., '*Rapid Response Information Report: Generative AI - language models (LLMs) and multimodal foundation models (MFMs)*' (Australian Council of Learned Academies, Report, 24 March 2023) 12.

[19] Bell, G. et al., '*Rapid Response Information Report: Generative AI - language models (LLMs) and multimodal foundation models (MFMs)*' (Australian Council of Learned Academies, Report, 24 March 2023) 12.

[20] Gil Appel, Juliana Neelbauer & David A. Schweidel, 'Generative AI Has an Intellectual Property Problem' *Harvard Business Review* (Article, 07 April 2023) <https://hbr.org/2023/04/generative-ai-has-an-intellectual-property-problem>.

[21] Office of the United Nations High Commissioner for Human Right, '*The Right to Privacy In the Digital Age*' (Report, A/HRC/51/17, 04 August 2022) 5.

[22] Office of the United Nations High Commissioner for Human Right, '*The Right to Privacy In the Digital Age*' (Report, A/HRC/51/17, 04 August 2022) 5.

[23] Office of the United Nations High Commissioner for Human Right, '*The Right to Privacy In the Digital Age*' (Report, A/HRC/51/17, 04 August 2022) 5.

[24] Murdoch, B. 'Privacy and Artificial Intelligence: Challenges for Protecting Health Information in a New Era' 22 (2021) *BMC Med Ethics* 1–5.

[25] Brown, H., et al., 'What Does it Mean for a Language Model to Preserve Privacy?' in *2022 ACM Conference on Fairness, Accountability, and Transparency* (ACM, 2022) 2280–2292; Pan, X., et al., 'Privacy Risks of General-Purpose Language Models' in *IEEE Symposium on Security and Privacy* (2020) 1314–1331.

[26] Josh Taylor, 'Medibank Hackers Announce 'Case Closed' and Dump Huge Data File on Dark Web' *The Guardian* (Article, 01 December 2022) <https://www.theguardian.com/australia-news/2022/dec/01/medibank-hackers-announce-case-closed-and-dump-huge-data-file-on-dark-web>.

[27] Australian Human Rights Commission ('AHRC'), '*Human Rights and Technology Final Report 2021*' (Final Report, 2021) ('Final Report') 115; See David Pozen, 'The Mosaic Theory, National Security, and the Freedom of Information Act' (2005) 115 *Yale Law Journal* 628.

[28] Li Li, et al., 'I Will Only Know After Using It: The Repeat Purchasers of Smart Home Appliances and the Privacy Paradox Problem' (2023) 128 *Computers & Security* 1.

[29] Consumer Policy Research Centre, '*2020 Data and Technology Consumer Survey*' (Survey, December 2020) 33.

[30] Consumer Policy Research Centre, '*2020 Data and Technology Consumer Survey*' (Survey, December 2020) 26.

[31] Lik-Hang Lee, et al., 'All One Needs to Know about Metaverse: A Complete Survey on Technological Singularity, Virtual Ecosystem, and Research Agenda' 2021 *arXIV* 37.

[32] Lik-Hang Lee, et al., 'All One Needs to Know about Metaverse: A Complete Survey on Technological Singularity, Virtual Ecosystem, and Research Agenda' 2021 *arXIV* 37.

[33] Lik-Hang Lee, et al., 'All One Needs to Know about Metaverse: A Complete Survey on Technological Singularity, Virtual Ecosystem, and Research Agenda' 2021 *arXIV* 37.

[34] Li Li, et al., 'I Will Only Know After Using It: The Repeat Purchasers of Smart Home Appliances and the Privacy Paradox Problem' (2023) 128 *Computers & Security* 1.

[35] Australian Competition and Consumer Commission (ACCC), '*Digital Platform Services Inquiry – September 2023 Report on the expanding ecosystems of digital platform service providers*' (Commonwealth of Australia, Issues Paper, March 2023) 7 citing ACCC, *Digital Platform Services Inquiry Fifth Interim Report* (Commonwealth of Australia, Firth Interim Report, 11 November 2022) 44.

[36] ACCC, '*Digital Platform Services Inquiry – September 2023 Report on the expanding ecosystems of digital platform service providers*' (Commonwealth of Australia, Issues Paper, March 2023) 7-8.

[37] Consumer Policy Research Centre, '*In Whose Interest? Why Businesses Need to Keep Consumers Safe and Treat their Data with Care*' (Working Paper, March 2023) 4 citing Anthony Nadler & Lee McGuigan, 'An Impulse to Exploit: The Behavioral Turn in Data-driven Marketing' (2018) 35(2) *Critical Studies in Media Communication* 151-165.

[38] Consumer Policy Research Centre, '*In Whose Interest? Why Businesses Need to Keep Consumers Safe and Treat their Data with Care*' (Working Paper, March 2023) 4.

[39] Consumer Policy Research Centre, '*In Whose Interest? Why Businesses Need to Keep Consumers Safe and Treat their Data with Care*' (Working Paper, March 2023) 10 citing Jack Balkin, 'The Fiduciary Model of Privacy' (2020) 134(11) *Harvard Law Review Forum* 12.

[40] Consumer Policy Research Centre, '*In Whose Interest? Why Businesses Need to Keep Consumers Safe and Treat their Data with Care*' (Working Paper, March 2023) 4.

[41] Consumer Policy Research Centre, '*In Whose Interest? Why Businesses Need to Keep Consumers Safe and Treat their Data with Care*' (Working Paper, March 2023) 4.

[42] Bell, G. et al., '*Rapid Response Information Report: Generative AI - language models (LLMs) and multimodal foundation models (MFMs)*' (Australian Council of Learned Academies, Report, 24 March 2023) 13.

[43] Marcello Ienca & Roberto Andorno, 'Towards New Human Rights in the Age of Neuroscience and Neurotechnology' (2017) 13(5) *Life Sciences, Society and Policy* 1, 1.

[44] Oliver Whang, 'Brain Implants Allow Paralyzed Man to Walk Using His Thoughts' *The New York Times* (Article, 24 May 2023) <https://www.nytimes.com/2023/05/24/science/paralysis-brain-implants-ai.html>.

[45] Nidhi Subbaraman, 'In the Brain, Scientists Find New Clues to Treating Chronic Pain' *The Wall Street Journal* (Article, 22 May 2023) <https://www.wsj.com/articles/brain-study-finds-clues-to-treating-chronic-pain-a19be9fb?st=cm5zi978o6pwc50&reflink=desktopwebshare_permalink>.

[46] Allan McCay, 'Neurotechnology, Law and the Legal Profession' (Horizon Report for The Law Society, August 2022) *The Law Society of England and Wales* 3.

[47] Oliver Whang, 'A.I. Is Getting Better at Mind-Reading' (01 May 2023) *New York Times* <https://www.nytimes.com/2023/05/01/science/ai-speech-language.html>.

[48] See generally Jerry Tang et al., 'Semantic Reconstruction of Continuous Language from Non-invasive Brain Recordings' (2023) 26 *Nature Neuroscience*.

[49] See Jerry Tang et al., 'Semantic Reconstruction of Continuous Language from Non-invasive Brain Recordings' (2023) 26 *Nature Neuroscience*.

[50] See Jerry Tang et al., 'Semantic Reconstruction of Continuous Language from Non-invasive Brain Recordings' (2023) 26 *Nature Neuroscience*.

[51] Sjors Ligthart et al., 'Minding Rights: Mapping Ethical and Legal Foundations of Neurorights' (2023) *Cambridge Quarterly of Healthcare Ethics* 1, 4 citing Mashat MEM, Li G & Zhang D., 'Human-to-human Closed-loop Control Based on Brain-to-brain Interface and Muscle-to-muscle Interface' (2017) 7(1) *Scientific Reports* 11001.

[52] Allan McCay, 'Neurotechnology, Law and the Legal Profession' (Horizon Report for The Law Society, August 2022) *The Law Society of England and Wales* 9 citing Yasmin Anwar, 'Scientists use Brain Imaging to Reveal the Movies in our Mind' *Berkeley News* (Article 22 September 2022) <https://news.berkeley.edu/2011/09/22/brain-movies/>.

[53] Allan McCay, 'Neurotechnology, Law and the Legal Profession' (Horizon Report for The Law Society, August 2022) *The Law Society of England and Wales* 10 citing Guangye Li & Dingguo Zhang, 'Brain-Computer Interface Controlled Cyborg: Establishing a Functional Information Transfer Pathway from Human Brain to Cockroach Brain' (2016) *Plus One*.

[54] Centre for International Relations and Sustainable Development (CIRSD), '*It's Time for Neurorights*' (Webpage, 2021) <https://www.cirsd.org/en/horizons/horizons-winter-2021-issue-no-18/its-time-for-neuro--rights>.

[55] The Neurorights Foundation, '*Market Analysis Neurotechnology*' (Report, March 2023) 14; CIRSD, '*It's Time for Neurorights*' (Webpage, 2021) <https://www.cirsd.org/en/horizons/horizons-winter-2021-issue-no-18/its-time-for-neuro--rights>.

[56] Rafael Yuste, Jared Genser & Stephanie Herrmann, 'It's Time for Neuro-Rights' (2021) 18 *Horizons* 1, 157.

[57] Sjors Ligthart et al., 'Minding Rights: Mapping Ethical and Legal Foundations of Neurorights' (2023) *Cambridge Quarterly of Healthcare Ethics* 1, 6.

[58] UNESCO et al., '*The Risks and Challenges of Neurotechnologies for Human Rights*' (Report, 2023) 30.

[59] Sjors Ligthart et al., 'Minding Rights: Mapping Ethical and Legal Foundations of Neurorights' (2023) *Cambridge Quarterly of Healthcare Ethics* 1, 7.

[60] UNESCO et al., '*The Risks and Challenges of Neurotechnologies for Human Rights*' (Report, 2023) 13.

[61] UNESCO et al., '*The Risks and Challenges of Neurotechnologies for Human Rights*' (Report, 2023) 13 & 30.

[62] UNESCO et al., '*The Risks and Challenges of Neurotechnologies for Human Rights*' (Report, 2023) 29.

[63] CIRSD, '*It's Time for Neurorights*' (Webpage) <https://www.cirsd.org/en/horizons/horizons-winter-2021-issue-no-18/its-time-for-neuro--rights>.

[64] CIRSD, '*It's Time for Neurorights*' (Webpage) <https://www.cirsd.org/en/horizons/horizons-winter-2021-issue-no-18/its-time-for-neuro--rights>.

[65] Australian Institute of Health and Welfare, '*People with Disability in Australia*' (Website) <https://www.aihw.gov.au/reports/disability/people-with-disability-in-australia/contents/people-with-disability/prevalence-of-disability>.

[66] UNESCO et al., '*The Risks and Challenges of Neurotechnologies for Human Rights*' (Report, 2023) 11.

[67] UNESCO et al., '*The Risks and Challenges of Neurotechnologies for Human Rights*' (Report, 2023) 11.

[68] UNESCO et al., '*The Risks and Challenges of Neurotechnologies for Human Rights*' (Report, 2023) 11.

[69] UNESCO et al., '*The Risks and Challenges of Neurotechnologies for Human Rights*' (Report, 2023) 11.

[70] Norwegian National Human Rights Institute, '*The Metaverse and Human* Rights' (Report, December 2022) 18.

[71] Norwegian National Human Rights Institute, '*The Metaverse and Human* Rights' (Report, December 2022) 22.

[72] Yuntao Wang, et al., 'A Survey on Metaverse: Fundamentals, Security, and Privacy' (2023) 25(1) *IEEE Communications Surveys & Tutorials, Communications Surveys & Tutorials* 319.

[73] Yuntao Wang, et al., 'A Survey on Metaverse: Fundamentals, Security, and Privacy' (2023) 25(1) *IEEE Communications Surveys & Tutorials, Communications Surveys & Tutorials* 319.

[74] Yuntao Wang, et al., 'A Survey on Metaverse: Fundamentals, Security, and Privacy' (2023) 25(1) *IEEE Communications Surveys & Tutorials, Communications Surveys & Tutorials* 320.

[75] Yuntao Wang, et al., 'A Survey on Metaverse: Fundamentals, Security, and Privacy' (2023) 25(1) *IEEE Communications Surveys & Tutorials, Communications Surveys & Tutorials* 320; see also Yavuz Canbay, Anil Utku & Pelin Canbay, 'Privacy Concerns and Measures in Metaverse: A Review' *15th International Conference on Information Security and Cryptography* (Conference Paper, 2022) 82.

[76] Yavuz Canbay, Anil Utku & Pelin Canbay, 'Privacy Concerns and Measures in Metaverse: A Review' *15th International Conference on Information Security and Cryptography* (Conference Paper, 2022) 84.

[77] Yavuz Canbay, Anil Utku & Pelin Canbay, 'Privacy Concerns and Measures in Metaverse: A Review' *15th International Conference on Information Security and Cryptography* (Conference Paper, 2022) 81 citing Anniki Puura, Siiri Silm & Anu Masso, 'Identifying relationships between personal social networks and spatial mobility: A study using smartphone tracing and related surveys' (2022) 68 *Social Networks* 306-317.

[78] Yuntao Wang, et al., 'A Survey on Metaverse: Fundamentals, Security, and Privacy' (2023) 25(1) *IEEE Communications Surveys & Tutorials, Communications Surveys & Tutorials* 328.

[79] Yuntao Wang, et al., 'A Survey on Metaverse: Fundamentals, Security, and Privacy' (2023) 25(1) *IEEE Communications Surveys & Tutorials, Communications Surveys & Tutorials* 334.

[80] Yavuz Canbay, Anil Utku & Pelin Canbay, 'Privacy Concerns and Measures in Metaverse: A Review' *15th International Conference on Information Security and Cryptography* (Conference Paper, 2022) 80.

[81] Yavuz Canbay, Anil Utku & Pelin Canbay, 'Privacy Concerns and Measures in Metaverse: A Review' *15th International Conference on Information Security and Cryptography* (Conference Paper, 2022) 84.

[82] Yuntao Wang, et al., 'A Survey on Metaverse: Fundamentals, Security, and Privacy' (2023) 25(1) *IEEE Communications Surveys & Tutorials, Communications Surveys & Tutorials* 334.

[83] Yavuz Canbay, Anil Utku & Pelin Canbay, 'Privacy Concerns and Measures in Metaverse: A Review' *15th International Conference on Information Security and Cryptography* (Conference Paper, 2022) 84.

[84] Yavuz Canbay, Anil Utku & Pelin Canbay, 'Privacy Concerns and Measures in Metaverse: A Review' *15th International Conference on Information Security and Cryptography* (Conference Paper, 2022) 80.

[85] Yavuz Canbay, Anil Utku & Pelin Canbay, 'Privacy Concerns and Measures in Metaverse: A Review' *15th International Conference on Information Security and Cryptography* (Conference Paper, 2022) 82 citing Roberto Di Pietro & Stefano Cresci, 'Metaverse: Security and Privacy Issues' *Trust, Privacy and Security in Intelligent Systems and Applications Third IEEE International Conference on TPS-ISA* (Conference Paper, 2021) 281-288.

[86] Yavuz Canbay, Anil Utku & Pelin Canbay, 'Privacy Concerns and Measures in Metaverse: A Review' *15th International Conference on Information Security and Cryptography* (Conference Paper, 2022) 81.

[87] Yavuz Canbay, Anil Utku & Pelin Canbay, 'Privacy Concerns and Measures in Metaverse: A Review' *15th International Conference on Information Security and Cryptography* (Conference Paper, 2022) 82 citing Roberto Di Pietro & Stefano Cresci, 'Metaverse: Security and Privacy Issues' *Trust, Privacy and Security in Intelligent Systems and Applications Third IEEE International Conference on TPS-ISA* (Conference Paper, 2021) 281-288.

[88] Jamie Wise, 'ChatGPTakeover: Millions of Australian workers using AI chatbots', *Finder* (Blog Post, 14 March 2023) <https://www.finder.com.au/millions-of-australian-workers-using-ai-chatbots>.

[89] Geoffrey Fowler, 'Snapchat tried to make a safe AI. It chats with me about booze and sex', *The Washington Post* (News Article, 14 March 2023) <https://www.washingtonpost.com/technology/2023/03/14/snapchat-myai/>.

[90] Julia Turc J, 'Unconstrained Chatbots Condone Self-Harm', *Medium* (Article, 08 April 2023) <https://towardsdatascience.com/unconstrained-chatbots-condone-self-harm-e962509be2fa>.

[91] Julian De Freitas et al, 'The Dark Side of Generative AI: Chatbots and Mental Health' (2022) *Harvard Business School Research Paper No 23*.

[92] See e.g. Sangeeta Singh-Kurtz, 'The Man of Your Dreams', *The Cut* (Blog Post, 10 March 2023) <https://www.thecut.com/article/ai-artificial-intelligence-chatbot-replika-boyfriend.html>.

[93] Luca Possati, 'Psychoanalyzing artificial intelligence: the case of Replika' (2022) *AI & Society*.

[94] Luiza Jarovsky, 'AI-Based Companions Like Replika Are Harmful to Privacy and Should Be Regulated', *The Privacy Whisperer* (Blog Post, 9 February 2023) <https://www.theprivacywhisperer.com/p/ai-based-companions-like-replika>.

[95] Luca Possati, 'Psychoanalyzing artificial intelligence: the case of Replika' (2022) *AI & Society*.

[96] Elvira Pollina & Martin Coulter, 'Italy bans U.S.-based AI chatbot Replika from using personal data', *Reuters* (News Article, 4 February 2023) <https://www.reuters.com/technology/italy-bans-us-based-ai-chatbot-replika-using-personal-data-2023-02-03/>.

[97] United Nations Human Rights Council, 48th sess, UN Doc A/HRC/RES/48/13 (18 October 2021).

[98] Rita Li, 'The Environmental Impact of AI', *GRC Insights* (Blog Post, 8 May 2023) <https://insights.grcglobalgroup.com/the-environmental-impact-of-ai/#:~:text=Training%20an%20AI%20model%20can,to%20exacerbate%20existing%20environmental%20problems.>.

[99] Rita Li, 'The Environmental Impact of AI', *GRC Insights* (Blog Post, 8 May 2023) <https://insights.grcglobalgroup.com/the-environmental-impact-of-ai/#:~:text=Training%20an%20AI%20model%20can,to%20exacerbate%20existing%20environmental%20problems.>.

[100] Payal Dhar, 'The carbon impact of artificial intelligence' (2020) 2 *Nature Machine Intelligence* 423.

[101] Organisation for Economic Co-operation and Development (OECD) 'Measuring the environmental impacts of artificial intelligence compute and applications: the AI footprint' (2022) 341 *OECD Digital Economy Papers* 7.

[102] Annette Ekin, 'AI can help us fight climate change. But it has an energy problem, too', *Horizon the EU Research & Innovation Magazine* (Article, 12 September 2019) <https://ec.europa.eu/research-and-innovation/en/horizon-magazine/ai-can-help-us-fight-climate-change-it-has-energy-problem-too>.

[103] Kate Saenko, 'Is generative AI bad for the environment? A computer scientist explains the carbon footprint of ChatGPT and its cousins', *The Conversation* (Blog Post, 23 May 2023) <https://theconversation.com/is-generative-ai-bad-for-the-environment-a-computer-scientist-explains-the-carbon-footprint-of-chatgpt-and-its-cousins-204096>.

[104] Max Schemmer, et al., '*On the Influence of Explainable AI on Automation Bias*' (Working Paper, 2022) 1 quoting Kathleen Mosier & Linda Skitka, 'Automation Use and Automation Bias' in Proceedings of the Human Factors and Ergonomics Society Annual Meeting (1999) 43(3) *SAGE Publications* 344–348.

[105] See e.g. Hilary Hanson, 'GPS Leads Japanese Tourists To Drive Into Australian Bay' Huffpost (Article, 19 March 2012) <https://www.huffingtonpost.co.uk/entry/gps-tourists-australia_n_1363823>.

[106] G. Bell et al., '*Rapid Response Information Report: Generative AI - language models (LLMs) and multimodal foundation models (MFMs)*' (Australian Council of Learned Academies, Report, 24 March 2023) 12.

[107] Adam Satariano and Paul Mozur, 'The People Onscreen are Fake. The Disinformation is Real.' *New York Times* (News Article, 07 February 2023) <https://www.nytimes.com/2023/02/07/technology/artificial-intelligence-training-deepfake.html>.

[108] Joel Golby, 'I Thought I was Immune to being Fooled Online. Then I saw the Pope in a Coat' *The Guardian* (News Article, 28 March 2023) <https://www.theguardian.com/commentisfree/2023/mar/27/pope-coat-ai-image-baby-boomers>.

[109] G. Bell et al., '*Rapid Response Information Report: Generative AI - language models (LLMs) and multimodal foundation models (MFMs)*' (Australian Council of Learned Academies, Report, 24 March 2023) 12.

[110] Genroe, *Social Media Statistics for Australia* (July 2022) <Social Media Statistics for Australia (Updated July 2022) - Genroe>.

[111] Dr Jake Wallis, International Cyber Policy Centre, Australian Strategic Policy Institute ('ASPI'), *Committee Hansard*, 22 June 2020, 10.

[112] Sarah O'Connor, Fergus Hanson, Emilia Currey and Tracy Beattie, *Cyber-enabled Foreign Interference in Elections and Referendums* (ASPI, 2020) 6.

[113] Select Committee on Foreign Interference through Social Media Interim, *First Interim Report* (First Interim Report, December 2021) 23 [3.3].

[114] Brennan Weiss, 'A Russian troll factory had a $1.25 million monthly budget to interfere in the 2016 US election' *Business Insider Australia* (News Article, 17 February 2018) <https://www.businessinsider.com/russian-troll-farm-spent-millions-on-election-interference-2018-2#:~:text=A%20Russian%20troll%20factory%20had,in%20the%202016%20US%20election&text=The%20office%20of%20the%20special,nationals%20and%203%20Russia%20entities>.

[115] Hannah Smith & Katherine Mansted, '*Weaponised Deep Fakes*' (ASPI, Report No. 28, April 2020) 11-12.

[116] Stuart Thompson, et al., 'How Bots Pushing Adult Content Drowned Out Chinese Protest Tweets' *New York Times* (News Article, 19 December 2022) <https://www.nytimes.com/interactive/2022/12/19/technology/twitter-bots-china-protests-elon-musk.html>.

[117] *International Covenant on Civil and Political Rights* art 19.

[118] *International Covenant on Civil and Political Rights* art 17.

[119] *International Covenant on Civil and Political Rights* art 25.

[120] News and Media Research Centre, Submission No 8 to Senate, *Select Committee on Foreign Interference through Social Media* (2021) 2-3.

[121] Sora Park et al., *Digital News Report: Australia 2022* (News and Media Research Centre, June 2022) 71.

[122] Sora Park et al., *Digital News Report: Australia 2022* (News and Media Research Centre, June 2022) 71.

[123] Sora Park et al., *Digital News Report: Australia 2022* (News and Media Research Centre, June 2022) 71.

[124] Sora Park et al., *Digital News Report: Australia 2022* (News and Media Research Centre, June 2022) 71.

[125] V-Dem Institute, *Democracy Report 2022: Autocratization Changing Nature?* (University of Gothenburg, March 2022) 35.

[126] UN Human Rights Committee, *General Comment No 34 (Article 19: Freedom of opinion and expression)*, 102nd sess, UN Doc CCPR/C/GC/34 (12 September 2011), [21].

[127] Select Committee on Foreign Interference through Social Media Interim, *First Interim Report* (First Interim Report, December 2021) 24 [3.7].

[128] Select Committee on Foreign Interference through Social Media Interim, *First Interim Report* (First Interim Report, December 2021) 32 [4.6].

[129] Select Committee on Foreign Interference through Social Media Interim, *First Interim Report* (First Interim Report, December 2021) 45 [4.49].

[130] Responsible Technology Australia, Submission No 17 to Senate, *Select Committee on Foreign Interference through Social Media* (2021) 2.

[131] The Department of Home Affairs, Submission No 16 to Senate, *Select Committee on Foreign Interference through Social Media* (2021) 4; see also The Allens Hub for Technology, Law and Innovation, Submission No 19 to Senate, *Select Committee on Foreign Interference through Social Media* (2021) 2.

[132] Select Committee on Foreign Interference through Social Media Interim, *First Interim Report* (First Interim Report, December 2021) 24-25 [3.10].

[133] The Allens Hub for Technology, Law and Innovation, Submission No 19 to Senate, *Select Committee on Foreign Interference through Social Media* (2021) 4.

[134] Australian Electoral Commission, Submission No 120 to the Joint Select Committee on Electoral Matters, *Inquiry into and Report on All Aspects of the Conduct of the 2019 Federal Election and Matters Related Thereto* (2018) 32.

[135] eSafety Commissioner, '*Deepfake trends and challenges — position statement*' (website) <https://www.esafety.gov.au/industry/tech-trends-and-challenges/deepfakes>.

[136] Isaac Stanley-Becker & Naomi Nix, 'Fake images of Trump arrest show 'giant step' for AI's disruptive power' *The Washington Post* (News Article, 22 March 2023) <https://www.washingtonpost.com/politics/2023/03/22/trump-arrest-deepfakes/>.

[137] Paul Smith, 'Deepfakes spell deep trouble for markets' *The Australian Financial Review* (Article, 23 March 2023) <https://www.afr.com/technology/deepfakes-spell-deep-trouble-for-markets-20230523-p5daih>.

[138] Mathew Wright & Christopher Schwartz, 'Voice deepfakes are calling – here's what they are and how to avoid getting scammed' *The Conversation* (Article, 17 March 2023) <https://theconversation.com/voice-deepfakes-are-calling-heres-what-they-are-and-how-to-avoid-getting-scammed-201449>.

[139] Douglas Harris, 'Deepfakes: False pornography is here and the law cannot protect you' (2018) 17(1) *Duke Law & Technology Review* 99.

[140] Dave Lee, 'Deepfakes porn has serious consequences' *BBC News* (Article, 03 February 2018) <https://www.bbc.com/news/technology-42912529>.

[141] Hannah Smith & Katherine Mansted, '*Weaponised Deep Fakes*' (ASPI, Report No. 28, April 2020) 11-12.

[142] Hannah Smith & Katherine Mansted, '*Weaponised Deep Fakes*' (ASPI, Report No. 28, April 2020) 11-12.

[143] Hannah Smith & Katherine Mansted, '*Weaponised Deep Fakes*' (ASPI, Report No. 28, April 2020) 11-12.

[144] Hannah Smith & Katherine Mansted, '*Weaponised Deep Fakes*' (ASPI, Report No. 28, April 2020) 11-12.

[145] Hannah Smith & Katherine Mansted, '*Weaponised Deep Fakes*' (ASPI, Report No. 28, April 2020) 13.

[146] Report of the United Nations High Commissioner for Human Rights, '*The Right to Privacy in the Digital Age*' (Report, A/HRC/48/31, 2021) 8-9.

[147] Rachel Goodman, 'Why Amazon's Automated Hiring Tool Discriminated Against Women', *American Civil Liberties Union* (Article, 12 October 2018) <https://www.aclu.org/news/womens-rights/why-amazons-automated-hiring-tool-discriminated-against>.

[148] Rachel Goodman, 'Why Amazon's Automated Hiring Tool Discriminated Against Women', *American Civil Liberties Union* (Article, 12 October 2018) <https://www.aclu.org/news/womens-rights/why-amazons-automated-hiring-tool-discriminated-against>.

[149] Rachel Goodman, 'Why Amazon's Automated Hiring Tool Discriminated Against Women', *American Civil Liberties Union* (Article, 12 October 2018) <https://www.aclu.org/news/womens-rights/why-amazons-automated-hiring-tool-discriminated-against>.

[150] Irwin Mitchell, 'Computer says no: unpicking the employment risks of AI' *Lexology* (Article, 30 March 2022) <https://www.lexology.com/library/detail.aspx?g=0639439a-42a5-48ac-b7fd-5f56908bb4b8>.

[151] Jessa Crispin, 'Welcome to Dystopia: Getting Fired from your Job as an Amazon Worker by an App' *The Guardian* (Article, 05 July 2021) <https://www.theguardian.com/commentisfree/2021/jul/05/amazon-worker-fired-app-dystopia>.

[152] *Fair Work Act 2009* (Cth) s 387.

[153] AHRC, Final Report (Final Report, 2021) 13.

[154] AHRC, Final Report (Final Report, 2021) 13.

[155] Crystal Grant, 'Algorithms are Making Decisions About Health Care, Which May Only Worsen Medical Racism' *American Civil Liberties Union* (Article, 3 October 2022) <https://www.aclu.org/news/privacy-technology/algorithms-in-health-care-may-worsen-medical-racism>.

[156] Crystal Grant, 'Algorithms are Making Decisions About Health Care, Which May Only Worsen Medical Racism' *American Civil Liberties Union* (Article, 3 October 2022) <https://www.aclu.org/news/privacy-technology/algorithms-in-health-care-may-worsen-medical-racism>.

[157] Crystal Grant, 'Algorithms are Making Decisions About Health Care, Which May Only Worsen Medical Racism' *American Civil Liberties Union* (Article, 3 October 2022) <https://www.aclu.org/news/privacy-technology/algorithms-in-health-care-may-worsen-medical-racism>.

[158] AHRC, Final Report (Final Report, 2021) 108.

[159] AHRC, Final Report (Final Report, 2021) 13.

[160] AHRC, Final Report (Final Report, 2021) 125.

[161] See generally AHRC, Final Report (Final Report, 2021) 125-135.

[162] AHRC, Final Report (Final Report, 2021) 128.

[163] AHRC, Final Report (Final Report, 2021) 129.

[164] AHRC, '*At the Crossroads: 10 years of implementing the UN Guiding Principles on Business and Human Rights in Australia*' (Report, 2021) 13.

[165] AHRC, '*At the Crossroads: 10 years of implementing the UN Guiding Principles on Business and Human Rights in Australia*' (Report, 2021) 14.

[166] AHRC, Final Report (Final Report, 2021) 66.

[167] 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).

[168] AHRC, '*Centrelink's compliance program*' submission to Senate Community Affairs References Committee regarding its inquiry into 'Centrelink's compliance program'.

[169] AHRC, '*Centrelink's compliance program*' submission to Senate Community Affairs References Committee regarding its inquiry into 'Centrelink's compliance program' 5.

[170] AHRC, Final Report (Final Report, 2021) 42.

[171] 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).

[172] Alexandria Utting, 'Kathleen Madgwick tells Robodebt royal commission about her son Jarrad and the damage the scheme caused' *ABC News* (Article, 10 March 2023) <https://www.abc.net.au/news/2023-03-10/qld-robodebt-scheme-government-royal-commission-fraud/102027838>.

[173] Brett Worthington, 'What did the Robodebt royal commission find about the people who oversaw the scheme?' *ABC News* (article, 07 July 2023) <https://www.abc.net.au/news/2023-07-07/political-reaction-to-robodebt-morrison-tudge-porter-robert/102575414>.

[174] AHRC, Final Report (Final Report, 2021) 82.

[175] AHRC, Final Report (Final Report, 2021) 65.

[176] AHRC, Final Report (Final Report, 2021) 80.

[177] AHRC, Final Report (Final Report, 2021) 80.

[178] AHRC, Final Report (Final Report, 2021) 80.

[179] AHRC, Final Report (Final Report, 2021) 83.

[180] AHRC, Final Report (Final Report, 2021) 81.

[181] AHRC, Final Report (Final Report, 2021) 81.

[182] AHRC, Final Report (Final Report, 2021) 81.

[183] AHRC, Final Report (Final Report, 2021) recommendations 2, 9 & 15.

[184] Roundtree, '*Facial Recognition Technology Codes of Ethics: Content Analysis and Review*' 211-220.

[185] AHRC, Final Report (Final Report, 2021) 113 citing Josh Bavas, 'Facial Recognition Quietly Switched on at Queensland Stadiums, Sparking Privacy Concerns,' *ABC News* (Article, 6 June 2019); Josh Bavas, 'The Facial Recognition Security Issue Police Tried to Keep Secret', *ABC News* (Article, 6 May 2019); Lisa Neville MP, 'New Police Eye in the Sky to Keep Victorians Safe' (Media Release, 10 July 2019); see, for example, Samantha Dick, 'Perth's Facial Recognition Cameras Prompt Scowls - And a Campaign to Stop 'Invasive' Surveillance', *The New Daily* (Article, 1 February 2020); Sarah Basford, 'Australian Schools Have Been Trialling Facial Recognition Technology Despite Serious Concerns About Children's Data', *Gizmodo* (Article, 10 March 2020).

[186] See e.g. Jarni Blakkarly, 'Kmart, Bunnings and The Good Guys using facial recognition technology in stores' *CHOICE* (Article, last updated 12 July 2022); There is a growing amount of literature on the use of facial recognition in a range of settings including sports stadiums, schools and retail outlets in Australia. See Brett Hutchins and Mark Andrejevcic, 'Olympian Surveillance: Sports Stadiums and the Normalization of Biometric Monitoring' (2021) 15 *International Journal of Communication*, 363; See Josh Bavas, 'Facial Recognition Quietly Switched on at Queensland Stadiums, Sparking Privacy Concerns,' *ABC News* (Article, 6 June 2019); Mark Andrejevic and Neil Selwyn, 'Facial Recognition Technology in Schools: Critical Questions and Concerns' (2020) 45(2) *Learning, Media and Technology*, 115; Sarah Basford, 'Australian Schools Have Been Trialling Facial Recognition Technology Despite Serious Concerns About Children's Data', *Gizmodo* (Article, 10 March 2020); Rick Sarre, 'Facial Recognition Technology is Expanding Rapidly Across Australia. Are Our Laws Keeping Pace?' The Conversation (Article, 10 July 2020) ; Eden Gillespie, 'Are You Being Scanned? How Facial Recognition Technology Follows You, Even as You Shop', *The Guardian* (Article, 24 February 2019).

[187] Jarni Blakkarly, 'Facial recognition technology in use at major stadiums across Australia' *CHOICE* (Article, 05 July 2023) <https://www.choice.com.au/consumers-and-data/data-collection-and-use/how-your-data-is-used/articles/facial-recognition-in-stadiums>.

[188] K.W Miller, 'Facial Recognition Technology: Navigating the Ethical Challenges' (2023) 56(1) *Computer* 76.

[189] Roundtree, '*Facial Recognition Technology Codes of Ethics: Content Analysis and Review*' 211-220.

[190] Roundtree, '*Facial Recognition Technology Codes of Ethics: Content Analysis and Review*'218.

[191] See generally Genia Kostka, Lea Steubacker & Miriam Meckel 'Under big brothers watchful eye: Cross-country attitudes towards facial recognition technology' (2023) 40(1) *Government Information Quarterly*.

[192] Genia Kostka, Lea Steubacker & Miriam Meckel 'Under big brothers watchful eye: Cross-country attitudes towards facial recognition technology' (2023) 40(1) *Government Information Quarterly* 1 citing Feldstien, S, 'The Global Expansion of AI Surveillance' *Carnegie Endowment for International Peace* (September 2019).

[193] K.W. Miller, 'Facial Recognition Technology: Navigating the Ethical Challenges' (2023) 56(1) *Computer* 76.

[194] Sanjeev Kumar, 'Delhi: Facial recognition system helps trace 3,000 missing children in 4 days' *The New India Express*, (Article, 22 April 2018) <https://www.newindianexpress.com/nation/2018/apr/22/3000-missing-children-traced-in-four-days-by-delhi-police-with-facial-recognition-system-software-1804955.html>.

[195] Julie Zaugg, 'India is trying to build the world's biggest facial recognition system' *CNN Business* (Article, 18 Oct 2019) <https://edition.cnn.com/2019/10/17/tech/india-facial-recognition-intl-hnk/index.html>.

[196] See e.g. the right to respect for the family as provided for under art 23.1 ICCPR; see also art 10 International Covenant on Economic, Social and Cultural Rights.

[197] Alexandra Ulmer & Zeba Siddiqui, 'India's use of facial recognition tech during protests causes stir' *Reuters* (Article, 17 February 2020) <https://www.reuters.com/article/us-india-citizenship-protests-technology-idUSKBN20B0ZQ>.

[198] Genia Kostka, Lea Steubacker & Miriam Meckel 'Under big brothers watchful eye: Cross-country attitudes towards facial recognition technology' (2023) 40(1) *Government Information Quarterly* 1 citing Feldstien, S, 'The Global Expansion of AI Surveillance' *Carnegie Endowment for International Peace* (September 2019).

[199] James Leibold, 'Surveillance in China's Xinjiang region: Ethnic sorting, coercion, and inducement' *Journal of Contemporary China* (2020) 29(121) 46-60.

[200] Office of the United Nations High Commissioner for Human Rights, *OHCHR Assessment of human rights concerns in the Xinjiang Uyghur Autonomous Region, People's Republic of China* (31 August 2022), [96] <https://www.ohchr.org/en/documents/country-reports/ohchr-assessment-human-rights-concerns-xinjiang-uyghur-autonomous-region>.

[201] Genia Kostka, Lea Steubacker & Miriam Meckel 'Under big brothers watchful eye: Cross-country attitudes towards facial recognition technology' (2023) 40(1) *Government Information Quarterly* 13.

[202] Genia Kostka, Lea Steubacker & Miriam Meckel 'Under big brothers watchful eye: Cross-country attitudes towards facial recognition technology' (2023) 40(1) *Government Information Quarterly* 13.

[203] Genia Kostka, Lea Steubacker & Miriam Meckel 'Under big brothers watchful eye: Cross-country attitudes towards facial recognition technology' (2023) 40(1) *Government Information Quarterly* 13.

[204] Jarni Blakkarly, 'Kmart, Bunnings and The Good Guys using facial recognition technology in stores' *CHOICE* (Article, last updated 12 July 2022) <https://www.choice.com.au/consumers-and-data/data-collection-and-use/how-your-data-is-used/articles/kmart-bunnings-and-the-good-guys-using-facial-recognition-technology-in-store>.

[205] Office of the Australian Information Commissioner, *'OAIC Opens Investigations into Bunnings and Kmart'* (Media Statement, 12 July 2022) <https://www.oaic.gov.au/newsroom/oaic-opens-investigations-into-bunnings-and-kmart>.

[206] Jarni Blakkarly, 'Kmart, Bunnings and The Good Guys using facial recognition technology in stores' *CHOICE* (Article, last updated 12 July 2022) <https://www.choice.com.au/consumers-and-data/data-collection-and-use/how-your-data-is-used/articles/kmart-bunnings-and-the-good-guys-using-facial-recognition-technology-in-store>; see also Amy Periera, '*Complaint OAIC on Use of Facial Recognition*' (CHOICE, Submission, June 2022).

[207] See generally Human Technology Institute, '*Facial Recognition Technology Towards a Model Law*' ('Model Law') (University of Technology Sydney, Report, September 2022).

[208] AHRC, Final Report (Final Report, 2021) 119.

[209] Human Technology Institute, *Model Law* (University of Technology Sydney, Report, September 2022).

[210] AHRC, Final Report (Final Report, 2021) 114.

[211] AHRC, Final Report (Final Report, 2021) 114.

[212] AHRC, Final Report (Final Report, 2021) 114; Pete Fussey and Daragh Murray, '*Independent Report on the London Metropolitan Police Service's Trial of Live Facial Recognition Technology*' (Human Rights, Big Data and Technology Project, July 2019) 36.

[213] AHRC, Final Report (Final Report, 2021) 114.

[214] AHRC, Final Report (Final Report, 2021) 115; See David Pozen, 'The Mosaic Theory, National Security, and the Freedom of Information Act' (2005) 115 *Yale Law Journal* 628.

[215] AHRC, Final Report (Final Report, 2021) 115; *Report of the Special Rapporteur on the Right to Privacy*, UN Doc A/HRC/40/63 (27 February 2019) 3.

[216] AHRC, Final Report (Final Report, 2021) 115; See e.g. Human Rights Watch, '*China's Algorithms of Repression: Reverse Engineering a Xinjiang Police Mass Surveillance App*' (Report, 1 May 2019).

[217] Bobby Allyn, 'The Computer Got It Wrong': How Facial Recognition Led To False Arrest Of Black Man' *NPR* (Article, 24 June 2020) <https://www.npr.org/2020/06/24/882683463/the-computer-got-it-wrong-how-facial-recognition-led-to-a-false-arrest-in-michig#:~:text=Bobby%20Allyn-,'The%20Computer%20Got%20It%20Wrong'%3A%20How%20Facial%20Recognition%20Led,False%20Arrest%20Of%20Black%20Man&text=Police%20in%20Detroit%20were%20trying,estimated%20%243%2C800%20worth%20of%20merchandise>.

[218] See e.g. Joy Buolamwini and Timinit Guru, 'Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification' (2018) 81 *Proceedings of Machine Learning Research* 1; KS Krishnapriya, Kushal Vangara, Michael C King, Vitor Albiero and Kevin Bowyer, 'Characterizing the Variability in Face Recognition Accuracy Relative to Race' (Conference Paper, IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, 2019); Inioluwa Deborah Raji and Joy Buolamwini, 'Actionable Auditing: Investigating the Impact of Publicly Naming Biased Performance Results of Commercial AI Products' (Conference on Artificial Intelligence, Ethics, and Society, 2019).

[219] Larry Magid, 'IBM, Microsoft And Amazon Not Letting Police Use Their Facial Recognition Technology' *Forbes* (Article, 12 June 2020) <https://www.forbes.com/sites/larrymagid/2020/06/12/ibm-microsoft-and-amazon-not-letting-police-use-their-facial-recognition-technology/?sh=15a148191887>.

[220] Toby Walsh, '*Machines Behaving Badly: The Morality of AI*' (Black Inc, Melbourne, 2022) 195.

[221] Ryan Mac, et al., 'Surveillance Nation' *Buzzfeed News* (Article, April 6 2021) <https://www.buzzfeednews.com/article/ryanmac/clearview-ai-local-police-facial-recognition>.

[222] Ryan Mac, et al., 'Surveillance Nation' *Buzzfeed News* (Article, April 6 2021) <https://www.buzzfeednews.com/article/ryanmac/clearview-ai-local-police-facial-recognition>; Ed Markey, '*Senators Markey, Merkley lead colleagues on legislation to ban government use of facial recognition, other biometric technology*' (Press Release, 15 June 2021).

[223] AHRC, Final Report (Final Report, 2021) 98.

[224] See generally AHRC, Final Report (Final Report, 2021).

[225] OECD, '*OECD Guidelines for Multinational Enterprises on Responsible Business Conduct*' (Guidelines, 08 June 2023) 46.

[226] AHRC, Final Report (Final Report, 2021) 55.

[227] AHRC, Final Report (Final Report, 2021) 55.

[228] AHRC, Final Report (Final Report, 2021) 98.