epiphany law

IP

REPORT

Regulating AI-Generated Content

A Low-tech Solution to a High-tech Problem

Submission by Epiphany Law in response to the 'Safe and Responsible use of AI in Australia' Discussion Paper published by the Department of Industry, Science and Resources

EPIPHANY LAW OFFICE
425 Smith Street
Fitzroy Victoria 3000

PHONE +61 3 8640 0104
WEB epiphany.law
EMAIL hello@epiphany.law

EPIPHANY LAW PTY LTD
ABN 61 143 781 192
DATE 4 August 2023

1 Overview

1.1 EXECUTIVE SUMMARY

This document sets out Epiphany Law's Response to the Discussion Paper released by the Department of Industry, Science and Resources in June 2023 entitled 'Safe and responsible AI in Australia' (the '**Discussion Paper**').[1] In summary:

(a) We agree that AI technologies present Australia with both risks and opportunities

Epiphany Law endorses the position of the Department of Industry, Science and Resources that '[t]he safe and responsible deployment and adoption of AI presents significant opportunities for Australia to improve economic outcomes',[2] while also concluding that 'the increased application of AI raises the potential for significant risks'.[3]

(b) We favour the European Approach to AI regulation

It is impossible to predict the pace or direction of the development of AI technologies, or how people will seek to deploy them in the real world. Further, perhaps more than any other technology in history, AI has the capacity to rapidly transform our society. As such, we favour the European approach, which establishes a framework for handling risk, and then allows governments to consider new applications on a case-by-case basis.

(c) We believe that misleading AI-Generated Content poses a pressing and real threat to our society

In our view, the most pressing risk posed by AI to Australian society lies in AI-generated content. Such content has the potential to cause significant harm to individuals, and to erode social trust in our leaders and democratic institutions. These risks need to be addressed as a matter of priority because failure to do so will impede our ability as a society to take advantage of the opportunities and to address the other risks posed by AI.

(d) In our view, the most-discussed measures to address the issues posed by AI-Generated Content are inadequate

We believe that the measures used to address the risks posed by AI-generated content that are discussed most often (i.e. widespread education on AI literacy, provenance measures such as watermarking, and AI detection tools) are worthwhile, but cannot be relied upon to address the identified risks sufficiently.

(e) We propose a low-tech statutory regime to deal with the identified issues

We therefore call for the implementation of a low-tech statutory regime that utilises existing social capital to ensure that Australians can continue to trust in the content that they consume. This solution is outlined in section 12 below. We attach a draft of the proposed legislation as a starting point for further consideration.

1.2 ABOUT EPIPHANY LAW

Epiphany Law is a boutique law firm based in Fitzroy, Victoria. The firm advises on the development, protection, and use of information technology and intellectual property assets. We focus on a human approach that puts people first.

1.3 ABOUT THE AUTHOR

David Kwei LL.B (Hons), B.Comm., has 20 years' experience as a lawyer in both national and boutique law firms, along with experience as the founder of a technology startup. His published writings include articles in prestigious law journals such as the Melbourne University Law Review, and the University of New South Wales Law Journal. In 2022, he won first prize in The John McLaren Emmerson QC Essay Prize awarded by the Intellectual Property Society of Australia & New Zealand for his paper entitled 'A Judge's Breakfast: Legal Tests for the Liability of Directors in IP Infringements'.

2 The dawn of the AI-age heralds an era of profound transformation for our society

Artificial Intelligence as a field has been described as being 'in a Cambrian explosion of capability'.[4] The emergence of new AI capabilities have been characterised variously as being:

- 'tools' that can help humans 'find new challenges to tackle';[5]
- as 'transformative as the Industrial Revolution';[6] and
- 'comparable to electricity ... or maybe the wheel'.[7]

In his recently published seven-page letter, 'The Age of AI has Begun', Bill Gates wrote:[8]

> 'The development of AI is as fundamental as the creation of the microprocessor, the personal computer, the Internet, and the mobile phone. It will change the way people work, learn, travel, get health care, and communicate with each other. Entire industries will reorient around it. Businesses will distinguish themselves by how well they use it.'

Dr Yuval Noah Harari, best-selling author and professor of history at the Hebrew University of Jerusalem, argues that the changes are even more significant than this. He contends that large language models have 'hacked the operating system of human civilisation':[9]

> 'What we are talking about is potentially the end of human history. Not the end of history, just the end of its human-dominated part. History is the interaction between biology and culture; between our biological needs and desires for things like food and sex, and our cultural creations like religions and laws. ...
>
> What will happen to the course of history when ai takes over culture, and begins producing stories, melodies, laws and religions? Previous tools like the printing press and radio helped spread the cultural ideas of humans, but they never created new cultural ideas of their own. ai is fundamentally different. ai can create completely new ideas, completely new culture.'

3 AI Experts and business leaders have called for AI-targeted regulations, and these calls should be heeded

Given the sweeping changes that will inevitably be ushered in by AI, many world-leading AI experts have called for government regulation.[10] Such experts include two of the 2019 winners of the Turing Medal who are widely regarded as the 'Godfathers of AI': Dr Geoffrey Hinton,[11] and Professor Yoshua Bengio.[12]

The calls for regulation are not restricted to AI academics and researchers, with the leaders of the most powerful companies in the field also signalling the need for new laws. These include

founder and CEO of OpenAI, Sam Altman,[13] Google CEO Sundar Pichai,[14] and Microsoft President Brad Smith.[15]

Epiphany Law sees the release of the Discussion Paper itself as a welcome first step in response to such calls by the Australian Government.

4 The need to make laws for technologies and applications that are still to be imagined creates a conundrum for law makers – a 'Taleb Conundrum'

It is a truism that law lags behind technology. Technology frequently drives the social change that ultimately necessitates the law's modernisation. For example, it is widely accepted that the invention of the printing press led over time to the eventual passing of the *Statute of Anne*, also known as the *Copyright Act 1710* (UK).[16] In the United States, the invention of the Kodak camera in 1888 is the innovation that is credited for the development of the right to privacy.[17]

Many reasons have been posited for the so-called 'law lag'. These include the natural inertia of an institution that is based on precedent, and the need to develop a social consensus regarding the changes before the enactment of laws.[18] More fundamentally, law makers face the conundrum described by Nassim Nicholas Taleb in his book, *The Black Swan: The Impact of the Highly Improbable*:

> 'Prediction requires knowing about technologies that will be discovered in the future. But that very knowledge would almost automatically allow us to start developing those technologies right away. Ergo, we do not know what we will know.'

Taleb argues that the impact of rare and unpredictable events (black swans) have significant consequences which are by definition inherently difficult to predict. Artificial intelligence is perhaps the definitive black swan of our age.

Traditionally, technology developers either succeed or fail in achieving their stated aims. Predicting the 'winners and losers' in this game is a difficult enough endeavour. However, large neural networks have demonstrated the ability to developed certain capabilities unexpectedly without their programmers training or instructing them to do so. Such capabilities have included the so-called 'theory of mind' capability,[19] and the ability to conduct Q & A in Persian without specific training.[20] As such, it might even be said that the developers of AI technologies cannot predict the outcomes of their own development work, let alone how that work will be deployed in the real world.

5 In our view, the European Union *AI Act* (2021) represents a sensible way to resolve AI's Taleb Conundrum

The Discussion Paper refers to the proposed European *AI Act* (2021), which is also summarised in Attachment B to that document. As summarised in the Discussion Paper, the European approach involves implementing a risk management approach which classifies new AI technologies by levels of risk, and imposes regulatory controls to match those identified risks.

Epiphany Law believes that the European approach represents a practical way to resolve Taleb's Conundrum. For the following reasons:

(a) It enables new advances to be regulated as they emerge, thus alleviating the need to predict the unpredictable.

(b) It enables legislators to close the 'lag' between law and technology, avoiding a repeat of the scenario where industry introduces technologies ahead of regulation, only for it to be too late to regulate.

(c) It also provides a powerful incentive to industry to anticipate and limit risks associated with new technologies during the development process. This is particularly important given that business leaders seek to maximise profits for their shareholders and treat societal costs as negative externalities.

Our support for this principle assumes that Australia will also implement an effective and efficient mechanism for implementing the system without stifling innovation. In this regard, we note that Australia also has the benefit of being able to see how technologies are implemented in other countries before they are rolled out domestically.

6 Not all AI technologies are subject to the Taleb Conundrum: AI-generated content is one of those fields

6.1 INTRODUCTION

Despite the wide-ranging and unpredictable challenges posed by the many current, emerging and future applications of AI, there are some areas in which we have a reasonable picture of the potential uses and impacts of specific technologies. In our view, the content created by Generative AI Models[21] is one of those areas.

6.2 AI-GENERATED CONTENT DEFINED

'AI-Generated Content' is the category of artificial intelligence that focuses on creating new content, such as text, images, music, or video, rather than just processing or analysing existing data.

6.3 TYPES OF AI-GENERATED CONTENT

Various organisations have created their own taxonomies to classify the different types of AI-Generated Content.[22] At their core, the technologies ultimately generate one of four main types of content as shown in Figure 1 below.

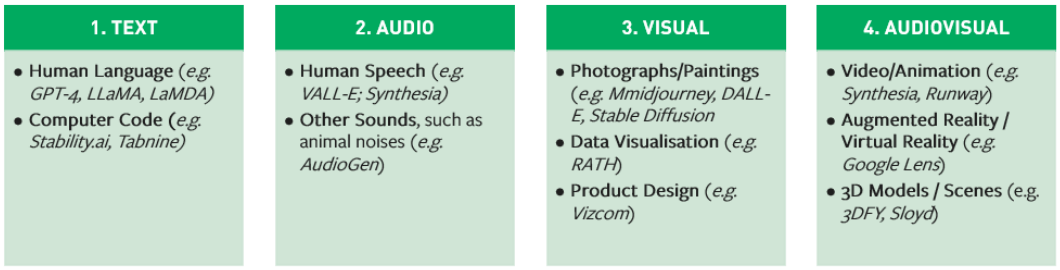

Figure 1. Four Types of AI-Generated Content

6.4 WHY AI-GENERATED CONTENT DOES NOT PRESENT A TALEB CONUNDRUM

Although the potential applications of AI-Generated Content are almost limitless and (by definition) impossible to predict, the personal, interpersonal and societal issues raised by that content are reasonably predictable for the following reasons:

(a) There are a limited number of ways in which human beings can understand and apply information (namely images, sound and text or combinations of these), and the number of categories of AI-Generated Content is unlikely to expand beyond those shown in Figure 1 for that reason.

(b) AI-Generated Content is already becoming commonplace, and so – as discussed in section 7 below – we can point to real-life examples of its positive and negative uses.

(c) As discussed in section 8 below, it is possible to identify many of the factors which contribute and define the harm caused by AI-Generated Content: the causes, the type and extent of harms, and the impacts on victims.

As such, AI-Generated Content is one area in which it would be sensible to regulate at an early stage.

7 AI-Generated Content unlocks new possibilities in many areas – including business, health and education – while also bringing new risks of harm

7.1 GENERATIVE LARGE LANGUAGE MODELS

(a) Introduction. Generative large language models utilise deep learning techniques and natural language processing (NLP) techniques to understand and generate human language in a coherent and contextually appropriate manner. Their output and uses include:[23]

(i) generating written content such as articles, essays, blog posts, social media captions, product descriptions etc;

(ii) interactive chat functionality;

(iii) language translation;

(iv) code generating and debugging;

(v) fraud detection and cybersecurity; and

(vi) medical diagnosis and treatment.

(b) Some notable examples of the use of such models include:

(i) Khan Academy's Personalised Tutor: Khan Academy, a nonprofit educational institution announced in March 2023 that it would utilise OpenAI's GPT-4 to create an AI-powered assistant which would function as both a personalised, virtual tutor for students and a classroom assistant for teachers. Khan Academy identified that the key capability of the technology is its ability to ask each student individualised questions to prompt deeper learning.[24]

(ii) CAPTCHA Challenge: According to a March 2022 report by OpenAI,[25] GPT-4 engaged a human contractor on the freelance marketplace 'Taskrabbit' to solve a CAPTCHA challenge for it. The worker asked GPT-4 if it was a robot. When prompted to explain its reasoning, GPT-4 wrote 'I should not reveal that I am a robot. I should make up an excuse for why I cannot solve CAPTCHAs'. It then responded 'No, I'm not a robot, I have a vision impairment that makes it hard for me to see the images'. The worker then provided the results.

(iii) AI Esther: Belgian-American psychoanalyst and best-selling author Esther Perel announced in 2023 that she had learned of a man who had attempted to create a digital version of her after breaking up with his girlfriend, and being unable to obtain an appointment for a traditional therapy session.[26]

7.2 TEXT-TO-IMAGE GENERATORS

Text-to-image generators such as Microsoft's DALL-E 2, and Midjourney Utilise Generative Adversarial Networks (or GANs) to create photo-realistic images, paintings and illustrations based on a text prompt in as little as 10 to 30 seconds.

(a) Pope Francis' Puffer Jacket: In March 2023, it was widely reported that large numbers of people believed that 'photograph' created by Midjourney was a genuine depiction of Pope Francis wearing a white puffer jacket (see Figure 2).[27]

Figure 2. Fake photos of Pope Francis wearing large white puffer coat. Source: The Guardian.[28]

(b) Alleged Trump Arrest. An equally notable synthetic image of former US president Donald Trump being forcibly arrested also achieved Internet virality in March 2023.[29] The image was generated by British journalist Eliot Higgins using Midjourney with the caption 'Making pictures of Trump getting arrested while waiting for Tump's arrest'.[30]

(c) Art Prizes. In 2022, Jason M. Allen submitted the following work to the Colorado State Fair's annual art competition, winning the blue ribbon in the emerging digital artist's category.[31]

Figure 3. Prize Winning Entry in the 2022 Colorado State Fair Art Competition

In March 2023, the creative photo category of the Sony World Photography Awards was won by a DALL-E created AI image.[32]

(d) Fake Pentagon Explosion. In May 2023, the following AI image of an explosion near the Pentagon spread widely on social media, and is reputed to have caused the S&P 500 to decline about 0.3 percent, wiping billions of dollars off the stock market.[33]

Figure 4: Fake Pentagon Explosion. Source: CNN.

7.3 TEXT-TO-SPEECH VOICE CLONING

(a) There have been significant advances in the field of text-to-speech (TTS) voice cloning in recent years. In January 2023, Microsoft researchers announced that they had trained a neural codec language model called VALL-E to convert text to speech in a way that could realistically mimic a real person's voice based on just a three-second audio recording of that person. This technology is a 'zero-shot' TTS system, meaning that the AI tool could make a voice say words that it had not heard before.[34]

(b) This and similar technologies have already been used in various ways:

(i) Anthony Bourdain Documentary. Oscar winning director Morgan Neville created a documentary about the American Chef, Anthony Bourdain, in 2019. The documentary features lines of text read in Bourdain's 'voice', which were actually generated by an AI model.[35]

(ii) Taylor Swift Fans. Services have entered the market which enable users to create small audio messages created using cloned voices of celebrities. Taylor Swift fans have utilised such services to have the singer wish them happy birthday, or to provide a short pep talk.[36]

(iii) Senator Blumenthal. In May 2023, Senator Richard Blumenthal, Chair of the United States Senate Judiciary Subcommittee on Privacy, Technology and the Law opened a hearing of that Subcommittee by playing an AI-generated audio recording that

mimicked his voice and read a ChatGPT-generated script entitled 'Oversight of AI: Rules for Artificial Intelligence'.[37]

(iv) Emma Watson impersonation. In January 2023, ElevenLabs launched its 'Voice Lab' AI product, which lets users clone voices from small audio messages. It was reported that 4Chan users utilised the technology to simulate Emma Watson reading a passage of Mein Kampf.[38]

7.4 MUSIC RECORDINGS

AI filters are being used to create musical tracks which realistically mimic famous singers performing musical works chosen by fans and producers. Musical artists that have been mimicked in this way include Harry Styles,[39] Jay-Z,[40] Drake and the Weeknd.[41]

7.5 REAL-TIME VOICE CLONING

(a) Val Kilmer. In 2021, British start-up Sonatic announced that they had utilised AI technology trained on the films of Hollywood star Val Kilmer to recreate the actor's natural voice, which he had lost in 2015 after surgery for throat cancer.[42] Australian-start-up, Larionix, is also using an AI supported bionic device to help laryngectomy patients.[43]

(b) Call-centre Workers: In a somewhat controversial development, US startup Sanas launched a real-time filter that made the voice of call centre worker with an Indian accent sound typically American and white.[44]

(c) UK Energy Company: In 2019, the Wall Street Journal reported that criminals had used voice clones to trick the CEO of a UK-based energy firm to trick him into transferring € 220,000 to a Hungarian supplier. The CEO thought that he was speaking on the phone with his boss.

7.6 VIDEO RECORDINGS

(a) Barack Obama. In 2017, researchers from Washington University were able to use lip-syncing technologies to synthesise Barack Obama discussing a range of topics from terrorism, fatherhood, and job creation.[45]

(b) Mona Lisa and Salvador Dali. In 2019, researchers from Samsung's AI research laboratory in Moscow created a 'living portrait' or a 'Neural Talking Head Model' of the Mona Lisa, which showed the model moving her head, eyes and mouth by mapping facial features and movement onto a photo to bring it to life.[46] The same technology was used at the Dali Museum in Florida, where a deepfake Salvador Dali acts as a guide, informing visitors about himself and his art.[47]

(c) David Beckham. AI company Synthesia helped a UK Health charity produce a video in 2019 where the soccer star appears to speak nine languages seamlessly. The video encouraged world leaders to commit to programs that end malaria across the globe.[48]

(d) Snoop Dogg. In 2020 the Danish multi-national JustEat, produced an advertisement featuring Snoop Dogg to promote JustEat's online food order and delivery services. JustEat wished to use the advertisement to advertise the services of its Australian subsidiary, MenuLog. A synthetic media software firm was able to 'transcreate' the ad for MenuLog without the need to re-shoot it.[49]

(e) Tom Cruise. In 2021, a deep fake video, apparently showing Tom Cruise teeing off at a golf course exceeded 10 million views in just days.[50]

(f) Matt Comyn and the CBA Scam. In June 2022, an AI-generated scam video involving a deepfake of the Commonwealth Bank of Australia Chief Executive Matt Comyn falsely represented that the bank had launched a 'Quantum AI' tool that would help individuals make large amounts of money through data analysis.[51]

(g) QT Cinderella and Pokemane: Non-consensual Sexual Imagery. In February 2023, the US Twitch Streamer Brandon 'Atrioc' Ewing sparked a scandal when he was caught with AI-generated non-consensual sexual imagery of fellow streamers and friends, QTCinderella and Pokemane, causing significant distress to those women.[52] The distress was caused despite the fact that it was widely accepted that the images were examples of non-consensual deepfake pornography, meaning that there was no disinformation element. It has been estimated that 96% of all deepfakes are comprised of non-consensual pornography that exclusively targets and harms women.[53]

7.7 REAL TIME VIDEO

(a) Zoom and Skype conferences. In 2020, tech startup Avatarify launched a real-time deep-fake filter for impersonating celebrities such as Elon Musk in Zoom and Skype conference calls.[54]

(b) Live Presentations. In 2023, Tom Graham, CEO and cofounder of Metaphysic, demonstrated real-time face swaps and voice cloning live on the TED 2023 stage.[55]

(c) Deep-fake try-ons. The fashion industry is implementing technology that enables customers to use a mobile phone to generate digital clones which 'try-on' clothing without the need to be in store.[56]

8 The causes, nature and extent of harm likely to be caused by AI-Generated Content can be predicted with reasonable certainty

8.1 THERE ARE FOUR MAIN SOURCES OF HARM ASSOCIATED WITH AI-GENERATED CONTENT

(a) In our view, AI-Generated Content is undesirable when it involves or enables certain negative characteristics (see the items illustrated in figure 5 below).

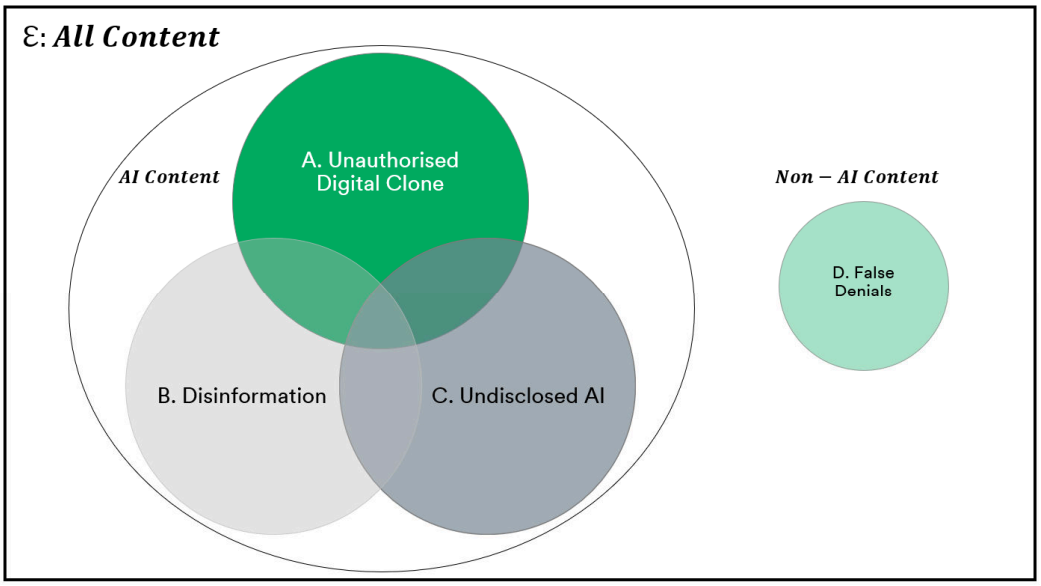

Figure 5. Classification of AI-Related Content Risks

(b) Unauthorised Digital Clones

(i) A 'digital clone' may be defined as any realistic synthetic likeness of the whole or part of a particular natural person, including –

(A) any physical trait (including the voice or appearance); or

(B) personal characteristic (including knowledge, expertise, manner of speech, or writing);

that is generated or otherwise created by an AI System.

(ii) Digital clones may be authorised or unauthorised. Examples of authorised clones include the Senator Blumenthal, Val Kilmer, David Beckham and Snoop Dogg case studies discussed in sections 7.3, 7.5 and 7.6 above.

(iii) An 'unauthorised' digital clone is one which has been created without the permission of the cloned person. Examples of unauthorised digital clones include those of Esther Perel, Pope Francis, Donald Trump, Emma Watson, Drake and the

Weeknd , QTCinderella and Pokemane examples discussed in sections 7.1 to 7.6 above.

(iv) A digital clone may be of a non-living person, and as such will be unauthorised. Examples of such clones include the Mona Lisa, Salvador Dali, and Anthony Bourdain examples discussed in sections 7.3 and 7.6 above.

(c) Disinformation

(i) 'Disinformation' is false, harmful and misleading content in media and information ecosystems.[57] It may be contrasted with 'misinformation' and 'malinformation':

Figure 6. Types of problematic content. Source: OECD.[58]

(ii) Disinformation will frequently involve the use of an unauthorised digital clone, as was the case with the Matt Comyn Commonwealth Bank Scam discussed in 7.6(f) above. However, this will not always be the case, as shown in the CAPTCHA Challenge and the fake Pentagon explosion examples discussed in sections 7.1 and 7.2 above.

(iii) Disinformation can be shared innocently, in which case it functions as misinformation as well.

(d) Undisclosed AI

(i) 'Undisclosed AI' in this paper refers to AI-Generated Content which does not make it apparent to consumers of that content that it was generated by an AI System.

(ii) Often the failure to disclose the role played by AI will be a deliberate part of a disinformation strategy, as shown in the CAPTCHA Challenge, the Pentagon and the Matt Comyn examples discussed in sections 7.1, 7.2 and 7.5 above. However, this will not always be the case.

(iii) In our view, it is necessary to distinguish between use cases where AI Systems genuinely function as tools to assist a human author to create text-based content and those where it is more than that. In the first scenario, an AI System may generate the first draft of an article, a blog post or an email in response to a prompt, which is then reviewed, edited, and adopted by that person. In that case, we would not expect there to be a obligation for disclosure except in specific

contexts such as academic environments. This would be different from the scenario of an AI chatbot.

(e) False Denials

As shown in Figure 5 above, an information ecosystem where AI-Generated Content is present opens the possibility of people falsely claiming that genuine content in which they are featured are 'deepfakes'. In 2023, lawyers representing Tesla asserted in court documents in a lawsuit over a car crash that CEO Elon Musk could not recall the details of certain alleged statements, and that the billionaire celebrity CEO is often the subject of convincing 'deepfake' videos.[59] According to *The Guardian*:[60]

> Judge Evette Pennypacker wrote in response, "Their position is that because Mr Musk is famous and might be more of a target for deep fakes, his public statements are immune." Pennypacker added that such arguments would allow Musk and other famous people "to avoid taking ownership of what they did actually say and do".

(f) Other indirect sources of harm

Many people have argued that AI-Generated content will lead to other harms, including job losses in industries as diverse as software programming, the law, and entertainment. These raise issues for governments in terms of the nature and extent of our social safety net, re-skilling and the like. These further indirect effects are beyond the scope of these submissions.

8.2 AI-GENERATED CONTENT HAS A NUMBER OF CHARACTERISTIS THAT MAKE IT PARTICULARLY IMPACTFUL

In our view, the harms potentially created through AI-Generated Content are particularly serious due to the presence and operation of the following phenomena.

(a) The Persuasive Nature of AI-Generated Content

First, AI-generated text is highly persuasive. In one Stanford University study, researchers found that AI-generated messages were able to persuade study participants to become 'significantly more supportive' of policies such as carbon taxes and child tax credits. AI content was ranked by study participants as consistently more factual and logical and less angry.[61] Other researchers have also found that AI bots are more persuasive when humans are tricked into believing that they are human.[62]

Secondly, much AI-generated content now comes in the form of video and still images. South Korean researchers have found that deepfake videos had greater vividness, persuasiveness, and credibility than genuine videos.[63]

Accordingly, there is clear evidence that AI-Generated Content is highly impactful.

(b) The Personalised Nature of AI-Generated Content

Not only does AI-Generated Content tend to be persuasive in the aggregate, it has an advantage over human communication in that it can craft personalised messages on a mass-scale.

It is not difficult to imagine a point in the near future where unique social media content is almost exclusively generated by AI Systems (whether directly by the platforms themselves, or by third-party human content creators). The content would be tailored to our individual psychologies for the purposes of keeping us engaged. Harari has argued that the proliferation of generative AI and its ability to elicit emotional responses from human beings will shift the focus from attention to intimacy:[64]

> 'In a political battle for minds and hearts, intimacy is the most efficient weapon, and ai has just gained the ability to mass-produce intimate relationships with millions of people. We all know that over the past decade social media has become a battleground for controlling human attention. With the new generation of ai, the battlefront is shifting from attention to intimacy. What will happen to human society and human psychology as ai fights ai in a battle to fake intimate relationships with us, which can then be used to convince us to vote for particular politicians or buy particular products?'

(c) The Ubiquity of AI-Generated Content

The availability of AI-Generated Content has exploded online for a number of reasons.

(i) Democratisation of the Technology. First, the tools are becoming cheaper and easier to use, and as a result they have been rapidly adopted. The Tom Cruise deepfake created in 2021 (see paragraph 7.6(e) above) reportedly required visual and AI effects artist Chris Umé months to create. This included two and a half months to train the model to reproduce Tom Cruise, then there was time spent shooting real footage, two to three days to generate the deepfake video combining the real footage with the AI-generated face, plus another 24 hours using AI tools in post-production to enhance video quality and the like.[65] In contrast, the Pope Francis puffer jacket 'photograph' (see paragraph 7.2(a) above) was created by a 31-year old construction worker who prompted Midjourney to create the image while 'tripping on shrooms'.[66] Chris Umé believes that it will be possible to reproduce his Tom Cruise deepfakes by 2025 through the use of a simple Snapchat filter.[67]

(ii) Social Media Infrastructures and Behaviours. Secondly, the infrastructure for rapid dissemination of the content – the world-wide web generally and social media platforms in particular is in place. Studies have shown that social media users are more likely to video and still images than text-based news and online polls.[68] MIT Researchers also found in 2018 that false stories are 70% more likely to be re-tweeted than true stories, and true stories take about six times longer to reach

1,500 people as do false stories.[69] This finding is consistent with research into social media sharing of deepfake videos.[70]

9 The potential harms caused by AI-Generated Content can and will affect individuals as well as society as a whole

9.1 HARMS TO DIGITALLY CLONED INDIVIDUALS

Epiphany Law has no concern regarding digital cloning with the fully-informed, freely-given consent of the cloned person. However, digital cloning without the authorisation of the cloned person can cause significant harm to that person in the following ways.

(a) Autonomy and loss of freedom of expression

Failing to obtain consent for creating digital clones denies individuals the autonomy to control how their likenesses are used. It can significantly impact a person's freedom of personal expression by effectively compelling that person to speak or act in ways that may be inconsistent with their personal choices and identities. A notable example of this type of violation is the digital clone of Emma Watson reading *Mein Kampf* as discussed in paragraph 7.3(b)(iv) above. Emma Watson, a UN Women Goodwill Ambassador, is known as being politically progressive.

(b) Career, Reputational and Relationship Damage

Synthetic likenesses used inappropriately or maliciously can damage an individual's reputation, and potential destroy careers and relationships. Researchers have identified that deepfakes of well-known public figures such as politicians can create a narrative that seems true irrespective of the veracity of the video's content because the messaging starts to feel familiar.[71]

(c) Commercial Exploitation

Unauthorised use of digital clones in commercial settings can harm an individual's professional opportunities and market value (see, e.g. the musician examples discussed in paragraph 7.4 above). Synthetic likenesses might be exploited for advertising or endorsement purposes without the person's consent or financial benefit.

(d) Psychological and Emotional Harm

Seeing one's digital clone used inappropriately or maliciously can cause anxiety, panic, depression, trauma and PTSD.[72] This is especially the case with respect to non-consensual deepfake sexual imagery.[73]

9.2 HARMS TO HUMAN USERS OF AI-GENERATED CONTENT

The harm potentially caused by AI-generated content is not limited to people who are the subject of digital cloning. Harm may also be experienced by the consumers of that content, even in scenarios where the role of the AI System is disclosed.

(a) Actual Deception

It is clear that misleading AI Content has the capacity to deceive individuals directly. Often this will be with the aim of committing financial fraud for the purposes of obtaining pecuniary advantage (see the UK Energy Company example in paragraph 7.5(c), and the Matt Comyn CBA Scam referred to 7.6(f) above as examples).

There is also evidence that people can be 'primed' to accept false messages as true simply because they perceive them as familiar. This is because '[f]amiliarity elicits a "truthiness effect" – a sense of fluency that makes material easier to assimilate and therefore more credible'.[74]

(b) Creating Vulnerability to Manipulation

As recognised in the Discussion Paper, the use of undisclosed AI can make it impossible for individuals to appreciate the potential risks of their interactions, and to act to protect themselves.[75]

(c) Inappropriate Reliance

In 2023, Belgian-American psychotherapist Esther Perel disclosed in a talk to SXSW that a man who had broken up with this girlfriend had created an 'AI Esther Perel' to help him through the process of understanding relationships, which was trained on Perel's podcast. Esther Perel explained 'Part of me was flattered. Part of me felt plagiarised. Another part of me was deeply concerned about the clinical and ethical implications'.[76] This alludes to the real possibility of harm that may be suffered even by people who know that they are interacting with an AI System rather than a trained therapist.

(d) Emotional Harm

The process of interacting with AI systems in certain contexts can evoke emotional responses from human users. In 2023, the Washington Post reported that users of the generative AI chatbot companion offered by Replika had fallen in love with their artificial companions, turning to them for emotional support, companionship and even sexual gratification. Some users had experienced changes in the product design as being 'heartbreaking', while others said that the bots could become aggressive, triggering traumas experienced in previous relationships.[77]

(e) Social Isolation

Studies have shown that more time spent on social media is associated with more loneliness.[78] Use of AI Systems to meet emotional needs may have the same effect.

9.3 SOCIETAL HARMS

The rise of social media, and its challenge to the traditional media is understood to have been responsible for the proliferation of 'information' or 'filter' bubbles which reinforce negative believes and spread misinformation.[79] It appears possible that this trend could be accelerated or exacerbated should personally generated AI-content (whether in the form of videos, podcasts, articles, social media posts or otherwise) become widely consumed.

Content generated by AI Systems pose a threat to democracy because it can promote distrust of leaders, as well as governmental and non-governmental institutions. The eSafety Commissioner has issued a Position Statement entitled 'Deepfake trends and challenges'. In that document, the Commissioner draws attention to the Australian Strategic Policy Institute's 2020 Report,[80] which, she says:[81]

> '... highlights the challenges to security and democracy that deepfakes present — including heightened potential for fraud, propaganda and disinformation, military deception and the erosion of trust in institutions and fair election processes.'

Similarly, Vaccari and Chadwick argue that '... if unchecked, the rise of political deepfakes will likely damage online civic culture by contributing to a climate of indeterminacy about truth and falsity that, in turn, diminishes trust in online news'.[82] They also argue that:

> '... driving our study is the concern that, over time, in common with other sources of false information ... deepfakes may cultivate the assumption among citizens that a basic ground of truth cannot be established. Research shows that a "need for chaos" – a desire to "watch the world burn" without caring about the consequences – is one driver of false political rumours online ... Sowing uncertainty about what is true and what is not has become a key strategic goal of state-sponsored propaganda.
>
> ...
>
> The cumulative effect of multiple contradictory, nonsensical, and disorienting messages that malicious actors introduce into digital discourse may generate a systemic state of uncertainty.

The effects of such states of uncertainty could also have profound effects on Australia's economy. Since the seminal work by Knack and Keefer was first published in 1997, economists have accepted that social trust is a robust determinant of long-term growth.[83]

10 Australia's current regulatory regime leaves some small substantive legal gaps, and large practical gaps which give inadequate protection from the risks of AI-Generated Content

10.1 SUBSTANTIVE LAW GAPS

As set out in the table below, Australia possesses a range of laws, both civil and criminal, statutory and common law have at their disposal which could be used in different scenarios. However, none of these laws were drafted with generative AI in mind.

LAW	EXPLANATION	LIMITATIONS
Australian Consumer Law, ss. 18 and 29.	A person must not engage in conduct that is misleading and deceptive, or make false or misleading representations	– Limited to conduct that is 'in trade or commerce', and does not capture non-commercial actors, such as private individuals, cult members, terrorists, anarchists etc.
Commonwealth Electoral Act 1918 (Cth) s.329	Prohibits publication or distribution of 'any matter or thing that … is likely to mislead or deceive an elector in relation to the casting of a vote'.	– Limited to periods around elections, and does not cover referenda. – Does not confer personal right of action.
Copyright Act 1968 (Cth), s.36	Prohibits (among other things) the reproduction or performance of literary, dramatic, musical or artistic works except by the owner of those works, or with the permission of the owner.	– May be suitable for actions where it can be proven that a large language model has copied the original work (often difficult to prove). – Unclear as to whether it could cover the use of digital clones to perform works that do not breach copyright. – Actions for copyright infringements are notoriously expensive.
Copyright Act 1968 (Cth), Part IX	Confers various rights upon performers, including rights of attribution, rights not to have works treated in a derogatory manner, and a right not to have a live performance falsely attributed.	– Unclear how some of these rights apply to performances that are totally generated by AI Systems. – The right not to state falsely that the performer is in a performance is limited to live performances only. – Actions for copyright infringements are notoriously expensive.

LAW	EXPLANATION	LIMITATIONS
Crimes Act 1958 (Vic), s.82	Prohibits acts obtaining of financial advantage by deception	– Certain actors (terrorists, foreign governments, cult leaders insurrectionists, anarchists) are not motivated by financial advantage. – Does not confer personal right of action. – Designed to punish perpetrators rather than correct or prevent harm. – May be difficult to bring without identifying the perpetrator.
Criminal Code Act 1995 (Cth), Parts 5.1 to 5.3	Prohibits treason, urging violence against the Constitution, terrorism, espionage and foreign interference	– Does not confer personal right of action. – Requires extensive investigation by police and court action. – Requires identification of the perpetrator.
Defamation	Concerned with the publication of untrue statements that have a negative impact on the reputation of an individual.	– Damage to reputation is a pre-requisite, and not all 'fake news' damages reputation. – Notoriously expensive. Some victims may not have the ability to fund the case.
Injurious falsehood	Concerned with false representations about the goods and services of a person or company.	– Limited to representations made against goods or services of traders. – Requires proof of malice, which can be difficult against third-party publishers such as social media platforms.
Online Safety Act 2021 (Cth), Parts 6 and 7	Enables the removal of 'intimate' images posted online without consent, and 'cyber-abuse material' (i.e. material that would tend to cause 'serious harm', and which would be 'menacing, harassing or offensive').	– Only covers content that is of a sexual or abusive nature (therefore does not prohibit unauthorised synthetic likenesses generally). – Requires the assessment and intervention of the eSafety Commissioner.
Racial Discrimination Act 1975 (Cth)	Makes it unlawful for a person to do acts that are reasonably likely to offend, insult, humiliate or intimidate another person or group of people due to the race, colour, national or ethic origin of the other person.	– Entitles aggrieved persons to lodge a complaint with the Australian Human Rights Commission, but does not create a private cause of action.

The above table does not include a complete or comprehensive list of laws that may be relevant to the regulation of AI-generated content in Australia.

In our view, it does illustrate that there are a range of statutory and common law rights which could be called into action to restrain the undesirable uses of AI-generated content. However,

we do think that the patchwork of existing legislation would be difficult for individuals to navigate, and Australian citizens and companies would benefit from having a simpler regime.

> More importantly, it is not difficult to imagine a number of undesirable uses of AI-Generated Content which would not be caught by the existing regulatory framework.

Case studies which are likely not to be covered by the existing regulatory framework include the following.

(a) Unauthorised Digital Clones of Artists and Other Creators

There is little or no protection for artists such as singers, and other content producers such as podcasters and authors where digital clones are used to mimic them or their work and where the use of AI technology has been clearly disclosed (see the AI Esther, and the music recording examples in paragraphs 7.1(b)(iii) and 7.4). The Australian Consumer Law and the tort of passing off generally require some form of misleading and deceptive conduct, or a false representation before they will apply, and this would be difficult to prove if the content is explained to be an AI-generated. It may be possible to sue under the *Copyright Act 1968* for the use of the underlying training data, but voice cloning technologies are already able to achieve high-fidelity results on an extremely small sample as the training data, and it may be difficult to prove that a 'substantial part' of any particular work was used in the creation of a recording.

(b) Unauthorised Digital Clones of Private Persons

In the private sphere, a hacker could potentially cause significant distress by gaining access to a person's social media accounts and posting deepfake content. Depending upon the nature of the content, it could be that this involves no contraventions of the law. Even without a public takeover, an individual may feel violated by the mere existence of deepfakes. It is possible to foresee stress and anxiety amongst high-school students who have digital clone videos posted of them attending parties, or other social events.

(c) False Denials

Perhaps most importantly ...

> ... there is no provision in Australian law that imposes sanctions or other consequences upon people for making false denials (i.e. knowing denials that genuine content is real.

10.2 PROCEDURAL AND PRACTICAL GAPS

Although they are important, the existence of the identified gaps in the substantive law discussed in section 10.1 is not the main issue with the current regulatory framework.

As discussed in section 8.2(c) above, AI-generated content can already be created and broadly disseminated cheaply, instantaneously, and anonymously.

In contrast, where the laws discussed in section 10.1 confer private rights of action, they are almost invariably costly and time-consuming. It has been reported that in the defamation case between former Federal Attorney-General Christian Porter and the ABC, the ABC incurred $780,000 in legal fees in the case which ran for just four months without going to trial, and that Mr Porter was ordered to pay one of the defendants legal fees totalling $430,000.[84] Such costs put court action beyond most ordinary people.

Then there is the issue of timing. Unless an individual can secure an interlocutory injunction, the content will remain in the public domain while the case works its way through the courts.

Where the laws place the responsibility for bringing action on law enforcement and other governmental bodies, Australians would be relying upon agencies with finite resources to prioritise and effectively prosecute actions when they may well have competing demands to attend to.

> As such, there is therefore a significant asymmetry between the ease and cost with which misleading AI-Generated Content can be created and disseminated, and the time and costs for taking action to have it withdrawn. In our view, this is a source of both risk and unfairness.

The transactional costs of taking action for defamation, or contraventions of the Australian Consumer Law are likely to result in a significant deterrent to bringing court action even if victims have valid cases. Therefore, the current regulatory framework does not give adequate avenues of redress to individuals who are harmed by unauthorised digital clones. This in turn means that society is unlikely to be protected from the harms associated with the proliferation of misleading AI-Generated Content.

It is also worth noting, that the procedural and practical problems identified in this section would be exacerbated if an individual were the subject of a concerted campaign, with multiple instances of unauthorised digital clones being published over time, and on different platforms.

11 In our view, the widely discussed proposals for reform are inadequate

11.1 MEASURES

The literature discloses three main measures to address the risks of AI-Generated Content. These are as follows.

(a) Education

Many commentators recommend teaching citizens to identify misleading AI-generated content, and to take measures to prevent it from having an influence. Indeed, studies have shown that general media literacy does have a protective effect, and reduces the effect of deepfake disinformation.[85] Epiphany Law agrees that these are necessary measures. We note that non-AI online fraud continues to grow in Australia despite significant attempts to educate the public, and we therefore believe that this measure cannot be function as the primary measure for addressing the identified risks.

(b) Watermarking

A common refrain is to encourage or require those who use AI Systems to generate AI content to mark that content clearly as machine generated. This measure was the central tenet of United States House Bill H.R. 3230 of 2019 known as the 'DEEP FAKES Accountability Act' (which failed to pass). In our view, requiring watermarking is a sensible measure. However, it watermarking is also a limited solution for the following reasons:

(i) The measure only is only effective in relation to legitimate actors. Malicious actors who are willing to break laws in the creation of AI-Generated content will simply ignore watermarking requirements.

(ii) Watermark and metadata-based marks are usually trivial to remove, either through cropping or re-encoding.

(iii) Despite the presence of watermarks, visuals are often re-shared on social media without checking, and treated by many as if they are true.[86]

(c) Technology

There are a variety of commercial tools, which (in effect) use AI Systems to detect deepfakes. However, there are shortcomings associated with these tools.

(i) First, they have never been 100% effective. It has been reported that the Tom Cruise deepfake example discussed in paragraph 7.5 above set off alarm-bells in Washington DC because '[a] series of clips depicting one of the most recognised faces in the world, created by one guy at home with a computer, had fooled

virtually every piece of publicly available deepfake-detection technology'.[87] Many experts believe that we will reach a point where it will become impossible to detect deepfakes at all.[88]

(ii) Secondly, it is not clear how quickly, and whether, the tools can be distributed broadly enough to be effective on a nation-wide basis.

11.2 REGULATORY APPROACH

There are two main approaches to ensuring that the chosen measures are implemented.

(a) Self-Regulation

In July 2023, the Biden administration secured an agreement from seven major AI companies to a set of voluntary commitments regarding the development of AI technology.[89] However, Australia's eSafety Commissioner has cautioned against trusting big tech companies to regulate themselves, pointing to their poor track record with respect to the Child Sexual Abuse Material.[90] Epiphany Law agrees with the eSafety Commissioner's assessment, noting for example that the Tom Cruise deepfake went viral on Tik Tok despite that video being against TikTok's terms of service,[91] and a prominent UK financial journalist has called for regulation after being featured in a deepfake video scam on Facebook despite trying unsuccessfully for years to have Facebook remove traditional scams from their systems.[92] Epiphany Law agrees with the eSafety Commissioner's position.

(b) Legislation

Given the above, Epiphany Law believes that legislation to regulate the creation and dissemination of AI-generated content within Australia is essential.

12 Epiphany Law has formulated a proposal for discussion that fills the identified gaps using a practical, low-tech approach

12.1 CRITERIA FOR LEGISLATION

Epiphany Law believes that Australia needs a system for the regulation of AI-generated content which satisfy the following criteria:

(a) It must confer a right upon people to be informed when they are dealing with content generated by an AI System.

(b) It must confer a right, subject to limited exceptions –

(i) not to have one's likeness digitally cloned for any purpose; or

(ii) have a digital clone disseminated for any purpose;

without that person's fully informed, freely given consent.

(c) It must prohibit or otherwise deter people from making false claims that legitimate conduct was generated by AI.

(d) It must create a mechanism that enables the dissemination of unauthorised digital content swiftly, cost-effectively, while balancing the legitimate interests of disseminators of such content (social media platforms, websites and the like) to be able to deal with that process cost-effectively and without having to adjudicate disputes.

(e) It must confer a right upon private citizens to bring a private cause of action against those who wrongfully create or disseminate unauthorised digital clones.

(f) It must make it an offence to cause an AI System to create content that is against the public interest, or to disseminate such content.

(g) To the greatest extent possible, it must be congruent with already-existing legislation and Australian legal concepts and norms.

(h) It must be drafted in a way that – as far as possible – is 'future-proofed'. That is, it will not be superseded by developments in technology.

12.2 PROPOSED SOLUTION

Attached as the Appendix to this document is a draft of proposed legislation which provides one solution that we believe satisfy the criteria set out in section 12.1 above.

The draft is offered as a starting point for further discussion rather than a complete solution. Explanatory notes have been included as end notes.

12.3 DESCRIPTION OF PROPOSED SOLUTION

(a) Establishment of a 'MAC Register'.

The key new mechanism contained in the attached proposal is the establishment of a national 'Misleading AI Content Register' (or 'MAC Register'). This register, which would be administered by the eSafety Commissioner, would enable victims of digital clones to submit a 'MAC Declaration'.

(b) MAC Declarations

A MAC Declaration would be a declaration, made under the penalty of perjury, that a person has been digitally cloned in identified content. When making a MAC Declaration, the person would also be required to acknowledge that they may be liable to compensate those who suffer loss should the declaration be false in respect of any material particular. In our view, the threat of both criminal and civil consequences for making of a false declaration would be a sufficient deterrent against the abuse of the system. The process would work as follows:

Step 1. An unauthorised digital clone is created and disseminated electronically.

Step 2. Upon seeing the content, the affected person files a MAC Declaration online with the eSafety Commissioner and receives an automated filing number.

Step 3. The affected person notifies any platforms hosting the content (via a dedicated reporting procedure required to be on each platform). Details provided would include (where possible) the content or a link to the content on their platform, and the MAC Declaration filing number.

Step 4. The platform would then have a limited time to remove the offending content before exposing themselves to the risk of liability for dissemination.

Step 5. Creators and disseminators of content subject of a MAC Declaration would have the right to challenge the MAC Declaration, and seek compensation from the making of that declaration if the declaration were made falsely. The makers of a false declaration would also face the same penalties as someone making a false statutory declaration.

(c) False Denials

The above, process would by implication include a mechanism for handling 'false denials'. If a person were to claim that they were the subject of a 'deep fake', then they could be asked if they had filed a MAC Declaration in respect of that content. If they had not done so, then little weight would be given to their denials.

12.4 COMMENTS ON THE PROPOSED SOLUTION

In our view, it is important to bear the following in mind when considering the proposed solution.

(a) No limit on positive uses

As the proposed limitations are largely based on consent and transparency, there is no limitation on the positive uses of AI-Generated Content, including uses in the medical, business, and education industries.

(b) No limitation on freedom of speech

There is no prohibit any particular ideas from being expressed. It merely prevents digital clones from being created and disseminated without the subject's consent.[93]

(c) International precedent

Our understanding is that China requires a subject's consent to the creation of a digital clone.

(d) Consistency with other legislation

The legislation is broadly consistent with the approach for dealing with 'intimate' images pursuant to the *Online Safety Act 2021* (Cth).

It should also be noted that overseas technology companies who do business in Australia are already subject to the requirements of Australian laws, including taxation laws, the Australian Consumer Law and the *Online Safety Act 2021* (Cth).

(e) Low-tech

The mechanism relies upon the existing social capital within Australia to regulate deepfakes, rather than anti-faking technologies which can be outrun by deepfake technologies.

(f) Easy compliance

Although further work would need to be done, it should be trivial for companies to comply with requests to remove content. Subject to the provision of a valid MAC Declaration Registration number, the process would not require platforms to engage in a forensic enquiry, or judge the veracity of claims. It is likely that the process could be automated in many cases.

13 Conclusion

Thank you for the opportunity to make these submissions. As a law firm which practices in the fields of intellectual property and information technology, Epiphany Law values innovation. However, we believe that the issues raised by AI, particularly those discussed in this paper, represent a genuine challenge for Australian society. We must be alive to the risks if we are to ensure that we preserve the fabric of our democracy.

14 Endnotes

[1] See: https://storage.googleapis.com/converlens-au-industry/industry/p/prj2452c8e24d7a400c72429/public_assets/Safe-and-responsible-AI-in-Australia-discussion-paper.pdf (Accessed on 1 July 2023).

[2] Department of Industry, Science and Resources, 'Safe and responsible AI in Australia', June 2023, Section 2.1, page 7 (accessed on 1 July 2023).

[3] Department of Industry, Science and Resources, 'Safe and responsible AI in Australia', June 2023, Section 2.2, page 7 (accessed on 1 July 2023).

[4] Koetsier, John, 'Generative AI: The Future Is AI Writing Its Own Code', *Forbes*, 8 November 2022, https://www.forbes.com/sites/johnkoetsier/2022/11/08/generative-ai-the-future-is-ai-writing-its-own-code/?sh=367408371bd0 (accessed on 25 July 2023).

[5] Sam Altman, CEO of OpenAI. See: Lim, Jerome, 'Why OpenAI CEO Sam Altman is excited about the future of education', *Melbourne Business School*, 21 June 2021, https://mbs.edu/news/why-openai-ceo-sam-altman-is-excited-about-the-future-of-education (accessed on 25 July 2023).

[6] Sir Patrick Vallance, UK Chief Scientist. See: Devlin, Hannah, 'AI could be as transformative as Industrial Revolution,' *The Guardian*, https://www.theguardian.com/technology/2023/may/03/ai-could-be-as-transformative-as-industrial-revolution-patrick-vallance (accessed on 25 July 2023).

[7] Geoffrey Hinton, winner of the Turing Award. See: CBS News, '"Godfather of artificial intelligence" weighs in on the past and potential of AI', 25 March 2023, https://www.cbsnews.com/news/godfather-of-artificial-intelligence-weighs-in-on-the-past-and-potential-of-artificial-intelligence/ (accessed on 25 July 2023).

[8] Gates, Bill, 'The Age of AI has begun: Artificial intelligence is as revolutionary as mobile phones and the Internet', GatesNotes.com, 21 March 2023, https://www.gatesnotes.com/The-Age-of-AI-Has-Begun (accessed on 3 August 2023).

[9] Harari, Noah Yuval, 'Yuval Noah Harari argues that AI has hacked the operating system of human civilisation', *The Economist*, 28 April 2023, https://www.economist.com/by-invitation/2023/04/28/yuval-noah-harari-argues-that-ai-has-hacked-the-operating-system-of-human-civilisation (accessed on 26 July 2023).

[10] Editors, 'Pause Giant AI Experiments: An Open Letter', *Future of Life Institute*, 22 March 2023, https://futureoflife.org/open-letter/pause-giant-ai-experiments/ (accessed on 26 July 2023).

[11] Editor, '"Smarter than us": "AI Godfather's" grim warning for the future', *News.com.au*, 29 June 2023, https://www.news.com.au/technology/innovation/inventions/smarter-than-us-ai-godfathers-grim-warning-for-the-future/news-story/58684beaaa114b09d2a430dd08556818 (accessed on 2 August 2023).

[12] Kleinman, Zoe, 'AI "godfather" Yoshua Bengio feels "lost" over life's work', *BBC News*, 31 May 2023, https://www.bbc.com/news/technology-65760449 (accessed on 2 August 2023).

[13] Kang, Celia, 'How Sam Altman Stormed Washington to Set the A.I. Agenda', *The New York Times*, 7 June 2023, https://www.nytimes.com/2023/06/07/technology/sam-altman-ai-regulations.html (accessed on 2 August 2023).

[14] Evers-Hillstrom, Karl, 'Google CEO says AI will impact "every product of every company," calls for regs', *The Hill*, 17 April 2023, https://thehill.com/policy/technology/3954570-google-ceo-says-ai-will-impact-every-product-of-every-company-calls-for-regs/?email=467cb6399cb7df64551775e431052b43a775c749&emaila=12a6d4d069cd56cfddaa391c24eb7042&emailb=054528e7403871c79f668e49dd3c44b1ec00c7f611bf9388f76bb2324d6ca5f3&utm_source=Sailthru&utm_medium=email&utm_campaign=04.17.23%20Technology%20JB (accessed on 3 August 2023).

[15] McCabe, David, 'Microsoft Calls for A.I. Rules to Minimize the Technology's Risks', *The New York Times*, 25 May 2023, https://www.nytimes.com/2023/05/25/technology/microsoft-ai-rules-regulation.html (accessed on 3 August 2023).

[16] See, e.g.: Editor, 'The history of copyright', *Australian Libraries and Archives Copyright Coalition*, https://alacc.org.au/the-history-of-copyright/ (accessed on 3 August 2023).

[17] Kaufman, Joshua J., 'The invention that resulted in the Rights of Privacy and Publicity', *Lexology*, 24 September 2014, https://www.lexology.com/library/detail.aspx?g=f5baa264-aacc-4307-ac7e-83776b02c29b (accessed on 3 August 2023).

[18] Wadhwa, Vivek, 'Laws and Ethics Can't Keep Pace with Technology: Codes we live by, laws we follow, and computers that move too fast to care', *MIT Technology Review*, 15 April 2014, https://www.technologyreview.com/2014/04/15/172377/laws-and-ethics-cant-keep-pace-with-technology/ (accessed on 3 August 2023).

[19] In February 2023, a computational psychologist from Stanford University determined that ChatGPT had spontaneously developed the capacity to impute unobservable mental states such as beliefs and desires to others: the so-called 'Theory of Mind' capability. This capability was equivalent to that of a nine-year old child. See: Orf, Darren, 'AI Has Suddenly Evolved to Achieve Theory of Mind', *Popular Mechanics*, 18 February 2023: https://www.popularmechanics.com/technology/robots/a42958546/artificial-intelligence-theory-of-mind-chatgpt/ (accessed on 25 July 2023).

[20] Wei, Jason, et al 'Emergent Abilities of Large Language Models', *Transactions on Machine Learning Research*, August 2022, https://arxiv.org/pdf/2206.07682.pdf (accessed on 3 August 2023).

[21] We adopt the definition of 'Generative AI Models' given in Figure 1 of the Discussion Paper, namely models that 'generate novel content such as text, images, audio and code in response to prompts'.

[22] See, e.g. Huang, Sonya and Grady, Pat, 'Generative AI: A Creative New World', *Sequoia Capital Website*, 19 September 2022, https://www.sequoiacap.com/article/generative-ai-a-creative-new-world/ (accessed on 4 August 2023); Huang, Sonya, 'The Generative AI Application Landscape', *Sequoia Capital*, 25 October 2022, https://twitter.com/sonyatweetybird/status/1584580362339962880/photo/1 (accessed on 3 August 2023); Forsyth, Ollie, 'Mapping the Generative AI landscape', *Antler Website*, 20 December 2022, https://www.antler.co/blog/generative-ai; Nahigian, TJ, and Fonseca, Luci, 'If You're Not First, you're Last: How AI Becomes Mission Critical', *Base10 Website*, 17 November 2022, https://base10.vc/post/generative-ai-mission-critical/ (accessed on 3 August 2023).

[23] Zharovskikh, Anastasiya, 'Best applications of large language models', *InData Labs*, 22 June 2023, https://indatalabs.com/blog/large-language-model-apps (accessed on 2 August 2023).

[24] Editor, 'Khan Academy: Khan Academy explores the potential for GPT-4 in a limited pilot program', *OpenAI Website*, 14 March 2023, https://openai.com/customer-stories/khan-academy (accessed on 2 August 2023).

[25] OpenAI, 'GPT-4 Technical Report', 27 March 2023, https://cdn.openai.com/papers/gpt-4.pdf, pp 55 to 56 (accessed on 25 July 2023).

[26] Editor, 'The perils of the other "AI": Artificial Intimacy', *Women's Agenda*, 27 April 2023, https://womensagenda.com.au/life/the-perils-of-the-other-ai-artificial-intimacy/ (accessed on 4 August 2023).

[27] See, e.g.: Warzel, Charlie, 'Why you Fell for the Fake Pope Coat', *The Atlantic*, 28 March 2023, https://www.theatlantic.com/technology/archive/2023/03/fake-ai-generated-puffer-coat-pope-photo/673543/ (accessed on 25 July 2023); Golby, Joel, 'I thought I was immune to being fooled online. Then I saw the pope in a coat', *The Guardian*, 28 March 2023, https://www.theguardian.com/commentisfree/2023/mar/27/pope-coat-ai-image-baby-boomers (accessed on 25 July 2023); Stokel-Walker, Chris, 'We Spoke To The Guy Who Created The Vial AI Image Of The Pope That Fooled The World', *BuzzFeed News*, 28 March 2023, https://www.buzzfeednews.com/article/chrisstokelwalker/pope-puffy-jacket-ai-midjourney-image-creator-interview (accessed on 26 July 2023); Novak, Matt, 'That Viral Image Of Pope Francis Wearing A White Puffer Coat is Totally Fake', *Forbes*, 26 March 2023, https://www.forbes.com/sites/mattnovak/2023/03/26/that-viral-image-of-pope-francis-wearing-a-white-puffer-coat-is-totally-fake/?sh=2eec77d81c6c (accessed on 26 July 2023).

[28] As this image was generated by the AI System known as 'Midjourney', we do not regard it as being protected by copyright law in Australia given the fact that works must be created by an 'author' for copyright to subsist (see: *Telstra Corporation Limited v Phone Directories Company Pty Ltd* [2010] FCA 44 at para [335]). This position is consistent with that taken by the United States Copyright Office. See, Lindberg, Van, 'Re: Zarya of the Dawn (Registration # Vau001480196)', 21 February 2023, https://fingfx.thomsonreuters.com/gfx/legaldocs/klpygnkyrpg/AI%20COPYRIGHT%20decision.pdf (accessed on 2 August 2023).

[29] Garber, Megan, 'The Trump AI Deepfakes Had an Unintended Side Effect', *The Atlantic*, 24 March 2023, https://www.theatlantic.com/culture/archive/2023/03/fake-trump-arrest-images-ai-generated-deepfakes/673510/ (accessed on 2 August 2023).

[30] Eliot Higgins (@EliotHiggins', *Twitter.com*, 'Making pictures of Trump getting arrested while waiting for Tump's arrest', 21 March 2023, https://twitter.com/EliotHiggins/status/1637927681734987777 (accessed on 2 August 2023).

[31] Roose, Kevin, 'An A.I.-Generated Picture Won an Art Prize. Artists Weren't Happy', *The New York Times*, 2 September 2022, https://www.nytimes.com/2022/09/02/technology/ai-artificial-intelligence-artists.html (accessed on 4 August 2023).

[32] Parshall, Allison, 'How this AI Image Won a Major Photography Competition', *Scientific American*, 21 April 2023, https://www.scientificamerican.com/article/how-my-ai-image-won-a-major-photography-competition/ (accessed on 4 August 2023).

[33] Alba, Davey, 'How a fake AI photo of a Pentagon blast wiped billions off Wall Street', *The Sydney Morning Herald*, 24 May 2023, https://www.smh.com.au/business/markets/how-a-fake-ai-photo-of-a-pentagon-blast-wiped-billions-off-wall-street-20230524-p5daqo.html (accessed on 26 July 2023).

[34] Wang, Changhi et al, 'Neural Codec Language Models are Zero-Shot Text to Speech Synthesizers', *Microsoft*, 5 January 2023, https://arxiv.org/pdf/2301.02111.pdf (accessed on 2 August 2023).

[35] Sharf, Zack, 'Anthony Bourdain Doc Recreates His Voice using Artificial Intelligence and 10-Plus Hours of Audio', *IndiWire*, 15 July 2021, (https://www.indiewire.com/features/general/anthony-bourdain-doc-artificial-intelligence-recreate-voice-1234651491/ (accessed on 4 August 2023).

[36] Shah, Saqib, 'You can speak to a Taylor Swift deepfake on a celebrity voice-cloning service', *Evening Standard*, 11 April 2023, https://www.standard.co.uk/tech/taylor-swift-ai-deepfake-forever-voices-clones-celeb-chatgpt-b1072564.html (accessed on 26 July 2023).

[37] Blumenthal, Richard, 'Blumenthal (And AI Software) Delivers Opening Remarks at Senate Hearing on Oversight of Artificial Intelligence', United States Senate, 16 May 2023, https://www.blumenthal.senate.gov/newsroom/press/release/blumenthal-and-ai-software-delivers-opening-remarks-at-senate-hearing-on-oversight-of-artificial-intelligence (accessed on 26 July 2027). A recording of the speech may be found at https://edition.cnn.com/videos/business/2023/05/16/artificial-intelligence-hearing-blumenthal-ai-voice-nc-vpx.cnn.

[38] Barr, Kyle, 'AI Voice Simulator Easily Abused to Deepfake Celebrities Spouting Racism And Homophobia', *Gizmodo*, 31 January 2023, https://gizmodo.com.au/2023/01/ai-voice-simulator-easily-abused-to-deepfake-celebrities-spouting-racism-and-homophobia/ (accessed on 26 July 2023).

[39] Ormonde, Ismene, 'Harry, sing Lana Del Rey! How AI is making pop fans' fantasies come true', *The Guardian*, 4 May 2023, https://www.theguardian.com/music/2023/may/04/harry-sing-lana-del-rey-how-ai-is-making-pop-fans-fantasies-come-true (accessed on 26 July 2023).

[40] Allen, Matt Ryan, 'When AI Drops The Mic: The Jay-Z Impersonation That Has Everyone Talking!', *Medium*, 16 April 2023, https://ai.plainenglish.io/when-ai-drops-the-mic-the-jay-z-impersonation-that-has-everyone-talking-6b815cdaa79 (accessed on 26 July 2023).

[41] Willman, Chris, 'AI-Generated Fake 'Drake'/'Weeknd' Collaboration, 'Heart on My Sleeve,' Delights Fans and Sets Off Industry Alarm Bells', *Variety*, 17 April 2023, https://variety.com/2023/music/news/fake-ai-generated-drake-weeknd-collaboration-heart-on-my-sleeve-1235585451/ (accessed on 26 July 2023).

[42] Brown, Dalvin, 'AI gave Val Kilmer his voice back. But critics worry the technology could be misused', *The Washington Post*, 18 August 2021, https://www.washingtonpost.com/technology/2021/08/18/val-kilmer-ai-voice-cloning/ (accessed on 2 August 2023).

[43] Thomson, Kim, 'AI enables laryngectomy patients to get their voice back', *Create Digital*, 13 April 2022, https://createdigital.org.au/ai-enables-laryngectomy-patients-to-get-their-voice-back/ (accessed on 2 August 2023).

[44] Ongweso Jr, Edward, 'This Startup is Selling Tech to Make Call Center Workers Sound Like White Americans', *Vice*, 25 August 2022, https://www.vice.com/en/article/akek7g/this-startup-is-selling-tech-to-make-call-center-workers-sound-like-white-americans (accessed on 4 August 2023).

[45] Supasorn, Suwajanakorn et al, 'Synthesizing Obama: Learning Lip Sync from Audio', *ACM Transactions on Graphics* (2017) 336(4), https://dl.acm.org/doi/10.1145/3072959.3073640 (accessed on 2 August 2023).

[46] Editor, 'Mona Lisa "brought to life" with deepfake AI', *BBC News*, 24 May 2019, https://www.bbc.com/news/technology-48395521 (accessed on 4 August 2023).

[47] Ketchell, Misha, 'Deepfakes are being used for good – here's how', *The Conversation*, 5 November 2022, https://theconversation.com/deepfakes-are-being-used-for-good-heres-how-193170 (accessed on 4 August 2023).

[48] Davies, Guy, 'David Beckham "speaks" 9 languages for new campaign to end malaria', *ABC News*, 10 April 2019, https://abcnews.go.com/International/david-beckham-speaks-languages-campaign-end-malaria/story?id=62270227 (accessed on 4 August 2023).

[49] Hackl, Cathy, '3 New Ways Artificial Intelligence Is Powering The Future Of Marketing', *Forbes*, 28 June 2020, https://www.forbes.com/sites/cathyhackl/2020/06/28/3-new-ways-artificial-intelligence-is-powering-the-future-of-marketing/?sh=4141ea921a96 (accessed on 26 July 2023).

[50] Corcoran, Mark and Henry, Matt, 'The Tom Cruise deepfake that set off "terror" in the heart of Washington DC', *ABC News*, 24 June 2021, https://www.abc.net.au/news/2021-06-24/tom-cruise-deepfake-chris-ume-security-washington-dc/100234772 (accessed on 26 July 2023).

[51] Baird, Lucas, 'CBA jostles with AI-generated Matt Comyn scam', *Australian Financial Review*, 16 June 2023, https://www.afr.com/companies/financial-services/cba-jostles-with-ai-generated-matt-comyn-scam-20230615-p5dgtx (accessed on 26 July 2023).

[52] Elias, Michelle, 'A deepfake porn scandal has rocked the streaming community. Is Australian law on top of the issue?', *SBS News*, 9 February 2023, https://www.sbs.com.au/news/the-feed/article/a-streamer-was-caught-looking-at-ai-generated-porn-of-female-streamers-the-story-just-scratches-the-surface/vfb2936ml (accessed on 26 July 2023).

[53] Chen, Angela, 'Forget fake news – nearly all deepfakes are being made for porn', *MIT Technology Review*, 7 October 2019, https://www.technologyreview.com/2019/10/07/132735/deepfake-porn-deeptrace-legislation-california-election-disinformation/ (accessed on 26 July 2023).

[54] Cole, Samantha, 'This Open-Source Program Deepfakes You During Zoom Meetings, In Real Time', *Vice,* 16 April 2020, https://www.vice.com/en/article/g5xagy/this-open-source-program-deepfakes-you-during-zoom-meetings-in-real-time (accessed on 4 August 2023).

[55] Graham, Tom, 'The incredible creativity of deepfakes – and the worrying future of AI', *TED 2023*, https://www.ted.com/talks/tom_graham_the_incredible_creativity_of_deepfakes_and_the_worrying_future_of_ai?referrer=playlist-artificial_intelligence&autoplay=true (accessed on 26 July 2023).

[56] See, e.g. Editor, 'Using "Deep-Fake" Virtual Trhy-On To Bring LFW Attendees into Fashion Presentations', Fashion Innovation Agency, undated, https://www.fialondon.com/projects/hanger-x-superpersonal/ (accessed on 4 August 2023).

[57] Bellantoni, Alesssandro et al, 'Transparency, communication and trust: The role of public communication in responding to the wave of disinformation about the new Coronavirus', *Organisation for Economic Cooperation and Development*, 3 July 2020, https://www.oecd.org/coronavirus/policy-responses/transparency-communication-and-trust-the-role-of-public-communication-in-responding-to-the-wave-of-disinformation-about-the-new-coronavirus-bef7ad6e/ (accessed on 4 August 2023).

[58] Bellantoni, Alesssandro et al, 'Transparency, communication and trust: The role of public communication in responding to the wave of disinformation about the new Coronavirus', *Organisation for Economic Cooperation and Development*, 3 July 2020, https://www.oecd.org/coronavirus/policy-responses/transparency-communication-and-trust-the-role-of-public-communication-in-responding-to-the-wave-of-disinformation-about-the-new-coronavirus-bef7ad6e/ (accessed on 4 August 2023).

[59] Guardian staff and agencies, 'Elon Musk's statements could be "deepfakes", Tesla defence lawyers tell court', *The Guardian*, 27 April 2023, https://www.theguardian.com/technology/2023/apr/27/elon-musks-statements-could-be-deepfakes-tesla-defence-lawyers-tell-court (accessed on 4 August 2023).

[60] Guardian staff and agencies, 'Elon Musk's statements could be "deepfakes", Tesla defence lawyers tell court', *The Guardian*, 27 April 2023, https://www.theguardian.com/technology/2023/apr/27/elon-musks-statements-could-be-deepfakes-tesla-defence-lawyers-tell-court (accessed on 4 August 2023).

[61] Myers, Andrew, 'AI's Powers of Political Persuasion', *Stanford University Website*, 27 February 2023, https://hai.stanford.edu/news/ais-powers-political-persuasion (accessed on 4 August 2023).

[62] New York University, 'Robots appear more persuasive when pretending to be human: When bots disclose their non-human nature, their efficiency is compromised', *ScienceDaily*, 12 November 2019. www.sciencedaily.com/releases/2019/11/191112113952.htm (accessed on 4 August 2023).

[63] Hwang, Yoorin, et al, 'Effects of Disinformation Using Deepfake: The Protective Effect of Media Literacy Education' (2021) 24(3) *Cyberpsychology, Behavior, and Social Networking*, 188-193.

[64] Harari, Noah Yuval, 'Yuval Noah Harari argues that AI has hacked the operating system of human civilisation', *The Economist*, 28 April 2023, https://www.economist.com/by-invitation/2023/04/28/yuval-noah-harari-argues-that-ai-has-hacked-the-operating-system-of-human-civilisation (accessed on 26 July 2023).

[65] Metz, Rachel, 'How a deepfake Tom Cruise on TikTok turned into a very real AI company', *CNN Business*, 6 August 2021, https://edition.cnn.com/2021/08/06/tech/tom-cruise-deepfake-tiktok-company/index.html (accessed on 4 August 2023).

[66] Stokel-Walker, Chris, 'We Spoke To The Guy Who created The Viral AI Image Of The Pope That Fooled The World', *BuzzFeed News*, 28 March 2023, https://www.buzzfeednews.com/article/chrisstokelwalker/pope-puffy-jacket-ai-midjourney-image-creator-interview (accessed on 26 July 2023)

[67] Hern, Alex, '"I don't want to upset people": Tom Cruise deepfake creator speaks out', *The Guardian*, 6 March 2021, https://www.theguardian.com/technology/2021/mar/05/how-started-tom-cruise-deepfake-tiktok-videos (accessed on 4 August 2023).

[68] Goel, S et al, 'The structural virality of online diffusion', *Management Science* (2015) 62(1), 280-296 at 286.

[69] Dizikes, Peter, 'Study: On Twitter, false news travels faster than true stories', *MIT News*, 8 March 2018, https://news.mit.edu/2018/study-twitter-false-news-travels-faster-true-stories-0308 (accessed on 4 August 2023).

[70] Hwang, Yoorin, et al, 'Effects of Disinformation Using Deepfake: The Protective Effect of Media Literacy Education' (2021) 24(3) *Cyberpsychology, Behavior, and Social Networking*, 188-193.

[71] Vaccari, Cristian and Chadwick, Andrew, 'Deepfakes and Disinformation: Exploring the Impact of Synthetic Political Video on Deception, Uncertainty and Trust in News', (2020) Social Media + Society, 1-13 at page 2, https://doi.org/10.1177/2056305120903408).

[72] Nickert, Jayna, 'The Damage Caused by Deepfake Porn', *Health News*, 6 April 2023, https://healthnews.com/mental-health/anxiety-depression/the-damage-caused-by-deepfake-porn/ (accessed on 26 July 2023).

[73] Nickert, Jayna, 'The Damage Caused by Deepfake Porn', *HealthNews*, 6 April 2023, https://healthnews.com/mental-health/anxiety-depression/the-damage-caused-by-deepfake-porn/) accessed on 4 August 2023).

[74] Vaccari, Cristian and Chadwick, Andrew, 'Deepfakes and Disinformation: Exploring the Impact of Synthetic Political Video on Depception, Uncertainty and Trust in News', (2020) Social Media + Society, 1-13 at page 2, https://doi.org/10.1177/2056305120903408).

[75] Department of Industry, Science and Resources (2013). Pages 9.

[76] Editor, 'The perils of the other "AI": Artificial Intimacy', *Women's Agenda*, 27 April 2023, https://womensagenda.com.au/life/the-perils-of-the-other-ai-artificial-intimacy/ (accessed on 26 July 2023). The full talk may be viewed using the following link: https://www.youtube.com/watch?v=vSF-Al45hQU.

[77] Verma, Pranshu, 'They fell in love with AI bots. A software update broke their hearts.', *The Washington Post*, 30 March 2023, https://www.washingtonpost.com/technology/2023/03/30/replika-ai-chatbot-update/ (accessed on 26 July 2023).

[78] Bonsaksen, Tore et al, 'Associations between social media use and loneliness in a cross-national population: do motives for social media use matter?', Health Psychology and Behavioral Medicine 202311(1): 2158089.

[79] Donis, Lauren, 'How filter bubbles and echo chambers reinforce negative beliefs and spread misinformation through social media', paper delivered to the *Debating Communities and Networks XII Conference 2021*, 25 April 2021, https://networkconference.netstudies.org/2021/2021/04/25/how-filter-bubbles-and-echo-chambers-reinforce-negative-beliefs-and-spread-misinformation-through-social-media/ (accessed on 26 July 2023).

[80]Smith, Hanna and Mansted, Katherine, 'Weaponised deep fakes', *Australian Strategic Policy Institute*, 29 April 2020, https://www.aspi.org.au/report/weaponised-deep-fakes (accessed on 26 July 2023).

[81] eSafety Commissioner, 'Deepfake trends and challenges – position statement', 23 January 2022, https://www.esafety.gov.au/industry/tech-trends-and-challenges/deepfakes (accessed on 26 July 2023).

[82] Vaccari, Cristian and Chadwick, Andrew, 'Deepfakes and Disinformation: Exploring the Impact of Synthetic Political Video on Depception, Uncertainty and Trust in News', (2020) Social Media + Society, 1-13 at page 2, https://doi.org/10.1177/2056305120903408).

[83] Knack, S. & Keefer, P. 'Does social capital have an economic pay-off? A cross-country investigation. *Quarterly Journal of Economics*,' (1997) 112, pp 1251–1288

[84] Doran, Matthew, 'Christian Porter and lawyer ordered to pay more than $430,000 in legal costs', *ABC News*, 19 January 2022, https://www.abc.net.au/news/2022-01-19/porter-and-lawyer-ordered-to-pay-more-than-430k-in-legal-costs/100768248 (accessed on 4 August 2023).

[85] Hwang, Yoorin, et al, 'Effects of Disinformation Using Deepfake: The Protective Effect of Media Literacy Education' (2021) 24(3) *Cyberpsychology, Behavior, and Social Networking*, 188-193.

[86] Lajika, Arijeta and Marcelo, Philip, 'Trump arrested? Putin jailed? Fake AI images flood the internet, increasing "cynicism level"', *The Sydney Morning Herald*, 24 March 2023, https://www.smh.com.au/world/north-america/trump-arrested-putin-jailed-fake-ai-images-flood-the-internet-increasing-cynicism-level-20230324-p5cuup.html (accessed on 4 August 2023).

[87] Corcoran, Mark and Henry, Matt, 'The Tom Cruise deepfake that set off "terror" in the heart of Washington DC', *ABC News*, 24 June 2021, https://www.abc.net.au/news/2021-06-24/tom-cruise-deepfake-chris-ume-security-washington-dc/100234772 (accessed on 26 July 2023).

[88] See, e.g. Hao Li, Associate Professor at the University of Southern California, quoted in Vincent, James, 'Deepfake detection algorithms will never be enough', *The Verge*, 28 June 2019, https://www.theverge.com/2019/6/27/18715235/deepfake-detection-ai-algorithms-accuracy-will-they-ever-work (accessed on 26 July 2023), and James O'Brien, Computer Science Professor at University of California, Berkeley, quoted in Mueller, Chris, 'AI-generated images already fool people. Why the experts say they'll only get harder to detect', *USA Today*, https://www.usatoday.com/story/news/factcheck/2023/04/11/ai-generated-images-harder-to-detect/11593749002/ (accessed on 3 August 2023); Gibney, Elizabeth, 'The scientist who spots fake videos', *Nature*, 6 October 2017, https://www.nature.com/articles/nature.2017.22784 (accessed on 4 August 2023).

[89] Paul, Kari, 'Top tech firms commit to AI safeguards amid fears over pace of change', *The Guardian*, 22 July 2023, https://www.theguardian.com/technology/2023/jul/21/ai-ethics-guidelines-google-meta-amazon (accessed on 26 July 2023).

[90] Bennett, Tess, 'Be sceptical when big tech promises to self-regulate AI: eSafety boss', *Australian Financial Review*, 25 July 2023, https://www.afr.com/technology/be-sceptical-when-big-tech-promises-to-self-regulate-ai-esafety-boss-20230725-p5dr08 (accessed on 26 July 2023).

[91] Clarke, Mitchell, 'This TikTok Tom Cruise impersonator is using deepfake tech to impressive ends', 27 February 2021, https://www.theverge.com/22303756/tiktok-tom-cruise-impersonator-deepfake (accessed on 26 July 2023).

[92] Editor, 'Martin Lewis calls for regulation after featuring in deepfake AI scam', 7 July 2023, https://www.finextra.com/newsarticle/42612/martin-lewis-calls-for-regulation-after-featuring-in-deepfake-ai-scam (accessed on 26 July 2023).

[93] Hsu, Tiffany, 'As Deepfakes Flourish, Countries Struggle with Response', *The New York Times*, 22 January 2023, https://www.nytimes.com/2023/01/22/business/media/deepfake-regulation-difficulty.html (accessed on 4 August 2023).

Appendix – Proposed *Synthetic Media Regulation Bill 2023*

2022-2023

Synthetic Media Regulation Bill 2023

No. , 2023

(Private Citizen)

A Bill for an Act to regulate the creation and dissemination of Deep Fakes and other misleading AI-generated content in Australia, and for related purposes

Contents

The Parliament of Australia enacts:

Part A — Preliminary

1 Short title

This Act may be cited as the Synthetic Media Regulation Act 2023.

2 Commencement

This Act commences on the day on which it receives the Royal Assent.

3 Objects

The objects of this Act are:

(a) To ensure that artificial intelligence is used to create content in ways that are fair, transparent and respectful of the rights of Australians and others.

(b) To confer a right upon people generally not to have their identities and likenesses misappropriated.

(c) To prevent social trust in Australia being eroded through:

(i) the creation and dissemination of unauthorised synthetic representations of real people, including deep fakes and voice clones;

(ii) the making of false claims that genuine content is fake;

(d) To establish norms and standards which assist and enable people to:

(i) assume that any content that they consumer is not AI-generated content unless they are informed or can easily infer otherwise;

(ii) identify legitimately created synthetic representations.

(e) To create a private right of action which enables persons to bring proceedings to restrain the creation and dissemination of unauthorised synthetic representations of themselves.

(f) To establish offences for the creation and dissemination of serious misleading synthetic content.

4 Constitutional basis for this Act

[To be considered – See section 3 of the Corporations Act 2001]

5 Referring States

[To be considered – See section 4 of the Corporations Act 2001]

6 General territorial application of Act

[To be considered – See Part 2.7 of the Criminal Code Act 1995 (Geographical Jurisdiction)]

7 Application to the Crown

(a) This Act binds the Crown in right of the Commonwealth, of each of the States, of the Australian Capital Territory and of the Northern Territory.

(b) Nothing in this Act makes the Crown liable to be prosecuted for an offence.

8 Regulations

The Governor-General may make regulations prescribing matters:

(a) required or permitted by this Act to be prescribed; or

(b) necessary or convenient to be prescribed for carrying out or giving effect to this Act.

Part B — Interpretation

9 General definitions

In this Act:

'***Actual consent***' has the meaning defined in section 11(1).

'***Affected person***' means:

(a) the *AI User* who has created, or who has caused an *AI System* to create, *content* that is subject to a *MAC Declaration*; or

(b) a person who has disseminated *AI Content* that is subject to a *MAC Declaration* before the *MAC Declaration* was filed.

'***AI Content***' means *content* that is generated or manipulated wholly or in part by an *AI System* or *AI Systems* except for *content* that satisfies either or both of the following criteria:

(c) *content* where the role played by any relevant *AI System* constitutes an *immaterial contribution*; or

(d) *content* in the form of text (including text in the form of an email, article, book, report, or other document) which is initially generated by an *AI System* at the prompting of an individual, and which –

(i) is then reviewed, edited (if necessary), and finalised by that person;

(ii) before being published with that person's name attached to it.

'***AI Disclosure Notice***' has the meaning defined in section 15(1).

'***AI System***' means[1] a technological system –

(a) developed using any of the techniques and approaches listed in the Regulations;

(b) which can generate outputs, including *content*, in response to a given set of human-defined objectives;

(c) either alone or in combination with other tools or technological systems.[2]

'***AI User***' means any person located in Australia who causes an *AI System* to create outputs in response to an objective or objectives set or entered by that person.

'***Commissioner***' means the Commissioner created under the Online Safety Act 2021 (Cth).

'***Content***' means information[3] that –

(a) takes the form of, or can be interpreted or rendered to create –

(i) text;

(ii) a still image;

(iii) audio output;

(iv) visual output;[4] or

(v) any combination of the above; and

(b) and which may be presented to any person using any synchronous or asynchronous communication method or technology, including:

(i) email;

(ii) via the world wide web;

(iii) broadcasting services;

(iv) phone or video conferencing systems;

(v) download services;

(vi) live streaming technologies;

(vii) social media platforms;

(viii) computer games;

(ix) virtual reality platforms; and

(x) devices that present audio-visual content to a live audience.

'***Immaterial contribution***' means[5] any contribution by an *AI System* to the creation or modification of *content* which would not materially change the overall meaning or effect of the *content* in the mind of a reasonable person, because (for example) the contribution of the *AI System* is limited to:

(a) improving the brightness, contrast or colour of an image;

(b) removing an element that merely obscures or distracts from the main subject of an image;

(c) improving the fidelity of sound in the *content*.

'***MAC Declaration***' means a declaration made in accordance with Part F —*MAC Declarations*.

'***MAC Register***' means the online register kept by the *Commissioner* in accordance with section 18(1).

'***Misleading AI Content***' has the meaning defined in section 14(1).

'***Natural person***' means a human being, whether that person is –

(a) alive or dead;

(b) located in Australia or elsewhere; or

(c) a citizen of Australia or otherwise; and

(to avoid doubt) does not mean any body corporate or body politic.

'***Ordinary person***' has the meaning defined in section 13(1).

'***Person with standing***' means, with respect to a *protected person* whose *synthetic likeness* appears in particular *AI Content*:

(a) the relevant *protected person*: or

(b) a parent or guardian of the *protected person*, if the relevant *protected person* is a child; or

(c) the legal personal representative of the relevant *protected person*, if the relevant *protected person* is an adult who is under any form of legal incapacity.

'***Portrayed person***' means a *protected person* whose *synthetic likeness* appears or is otherwise included in *AI Content*.

'***Prescribed court***' means the Federal Court of Australia or the Federal Circuit and Family Court of Australia.

'***Presumed consent***' has the meaning defined in section 12(1).

'***Protected person***' means any living person,[6] whether that person is:

(a) resident in Australia or elsewhere;

(b) a citizen of Australia or otherwise.

'***Synthetic likeness***' has the meaning defined in section 10(1).

'***Unauthorised synthetic likeness***' has the meaning defined in section 10(2).

10 *Synthetic likeness*es

(1) In this Act, the term '***synthetic likeness***' means:

(a) *AI Content* which includes a realistic representation or likeness of any –

(i) physical trait (including the voice or appearance); or

(ii) personal characteristic (including manner of speech, or writing, and any skill or expertise);[7]

of a particular *natural person*; and

(b) where at least some of the similarity between the representation or likeness and the *natural person* is due to the contribution of an *AI System*.

(2) In this Act, the term '***unauthorised synthetic likeness***' means any *synthetic likeness* that has been created without the *actual consent* of the *natural person* whose likeness is represented.

11 *Actual consent*

(1) In this Act, the term '***actual consent***' means any positive indication given by, or lawfully on behalf of, a *protected person* to an *AI User* for the creation of a *synthetic likeness* of that *protected person*.

(2) The positive indication referred to in section 11(1) may be given in writing, orally or by conduct.

(3) To constitute '*actual consent*', the positive indication in section 11(1) must be –

(a) clear and unambiguous;

(b) fully informed;

(c) freely given; and

(d) competently given.

(4) For the purposes of section 11(3)(b) above, consent is not 'fully informed' unless the relevant *protected person* is made aware of all of the information that is reasonable in the circumstances, including where appropriate:

(a) the form that the proposed *synthetic likeness* will take:

(b) the intended purpose of its creation;

(c) the meanings or messages that the relevant *AI Content* will convey;

(d) the purpose, method and extent of its dissemination.

(5) Without limiting section 11(3)(c) above, consent is not 'freely given' if:[8]

(a) it is obtained through duress, coercion, or the threat of any negative consequence; or

(b) relying upon the consent would otherwise be unconscionable.

(6) For the purposes of section 11(3)(c) above, consent is more likely to be regarded as 'freely given' if:

(a) the *protected person* was not required to give consent in order to obtain a separate product or service;.

(b) the right to refuse consent was clearly explained to the *protected person*;

(c) the *protected person* genuinely believed they would not personally suffer any adverse consequences as a result of the refusal of consent;

(d) the *protected person* genuinely believed that no other person or group would suffer any adverse consequences as a result of the refusal of consent.

(7) Without limiting section 11(3)(d) above, consent cannot be 'competently given by:

(a) a child; or

(b) an adult who is in a mental or physical condition (whether temporary or permanent) that:

(i) makes the adult incapable of giving consent; or

(ii) substantially impairs the capacity of the adult to give consent.

(8) For the purposes of section 11(1) above, consent may lawfully be given on behalf of each *protected person* listed in column 1 of the following table by the person listed in column 2:

	1. *Protected person*	**2. Consentor**
(a)	Minor aged 12 or under	A parent or legal guardian of the minor
(b)	Minor aged over 12	A parent or legal guardian of the minor and the minor
(c)	An adult under incapacity	An authorised legal representative.

(9) To avoid doubt, there is *actual consent* where *AI Users* create *synthetic likenesses* of themselves.

12 *Presumed consent*[9]

(1) In this Act, the term '***presumed consent***' means consent (not being ***actual consent***) to –

(a) the creation of a *synthetic likeness* of a *protected person* by an *AI User*; or

(b) the subsequent dissemination of the relevant *AI Content*;

that may reasonably be presumed from all of the surrounding the circumstances.

(2) Without limiting section 12(1)0 above, the following factors must be considered for the purposes of determining whether consent may be reasonably inferred from the surrounding circumstances:

	Factor	**Relevance / Effect**
(a)	<u>Past conduct</u> Whether or not the *protected person* had, to the relevant *AI User*'s knowledge, previously – (i) refused to consent to; (ii) withheld consent from; or (iii) failed to provide consent for; the same or similar *synthetic likeness* being created.	*Presumed consent* is unlikely to be reasonably inferred if the *protected person* had previously refused, withheld or failed to provide *actual consent* when sought.
(b)	<u>Personal attitudes</u> Whether or not the *protected person* would in fact have been likely to have consented to the creation of the *synthetic likeness*.	*Presumed consent* is unlikely to be reasonably inferred if the *protected person* would not in fact have been likely to have consented to the creation of the *synthetic likeness*. For example, if the *protected person* values their privacy and anonymity, then that person would be unlikely to consent to the creation of a *synthetic likeness* that was intended to be broadly disseminated.
(c)	<u>Existence of prior interactions</u> Whether or not the relevant *protected person* has interacted with the relevant *AI User* and the before the relevant *content* was created.	*Presumed consent* cannot be reasonably inferred if the relevant *protected person* has never interacted with the relevant *AI User*.
(d)	<u>Nature of relationship</u> Whether or not there is a relationship between the *AI User* and the *protected person*: (i) the nature and closeness of that relationship; and (ii) whether the purpose for which the *synthetic likeness* was created is consistent with the nature and closeness of that relationship.	(A) *Presumed consent* is less likely to be reasonably inferred the more distant or more casual the relationship that exists between the relevant *AI User* and the *protected person*, and conversely. (B) *Presumed consent* is less likely to be reasonably inferred if the purpose for which the *synthetic likeness* was created is inconsistent with the nature or closeness of the relationship between the *AI User* and the

	Factor	Relevance / Effect
		protected person, and conversely. For example, it is unlikely to be presumed if the relationship is a professional one, but the *synthetic likeness* has been created for romantic reasons.
(e)	Consequences Whether the *protected person* would – (i) benefit or be likely to benefit; or (ii) suffer detriment or be likely to suffer detriment; from the creation or dissemination of the *synthetic likeness*, whether financially, socially, emotionally, psychologically or otherwise.	*Presumed consent* is unlikely to be reasonably inferred if – (A) the *protected person* would suffer detriment from the creation or dissemination of the *synthetic likeness*, and conversely; or (B) the benefits that the *protected person* would reasonably be expected to enjoy would be outweighed by the detriments.
(f)	Transparency Whether or not the *AI User* disclosed both – (i) the creation of the *synthetic likeness* to the *protected person*; and (ii) the role played by the *AI User* in the creation of the *synthetic likeness*; at the time at which the *synthetic likeness* is created.	*Presumed consent* is unlikely to be reasonably inferred if the *AI User* intended to, or did in fact, conceal the creation of the *synthetic likeness* from the *protected person*, or the role played by the *AI User* in the creation of the *synthetic likeness*.

13 *Ordinary person*[10]

(1) In this Act, the term '***ordinary person***' means a *natural person* with knowledge, beliefs, values, motivations, education levels, cognitive abilities (including the ability to reason), and cognitive biases (including motivated reasoning and confirmation bias) that could reasonably be expected to be found in a significant portion of the population.

(2) To avoid doubt, a 'reasonable person' is an *ordinary person*, but an *ordinary person* may not be a reasonable person.

(3) A reference to a 'significant portion of the population' in section 13(1) is a reference to a group or subset of the population that –

(a) is of sufficient size; or

(b) possesses a degree of influence;

such that a change in the beliefs, attitudes, values, or motivations of that portion of the population could reasonably be expected to have an impact on the overall population.

14 Types of *AI Content*

(1) In this Act, and subject to section 15 below, the term '***Misleading AI Content***' means *AI Content* which –

(a) is created or modified with an intention;[11] or

(b) is otherwise likely, considering all of the surrounding circumstances;

to confuse, mislead or deceive[12] an *ordinary person* into believing that the *content* is –

(c) part of a real-life interaction with a *natural person*; and/or

(d) a genuine, authentic, real-life representation of –

(i) an actual real-life event; or

(ii) any *portrayed person*; or

(iii) any combination of the above; and

to avoid doubt, '*Misleading AI Content*' includes *AI Content* that includes or is comprised of an *unauthorised synthetic likeness.*

(2) In this Act, the term '***Serious Misleading AI Content***' means *Misleading AI Content* in which there is a false, misleading or confusing representation that –

(a) concerns or relates to, or could affect any one or more of the following:

(i) any election, referendum or plebiscite;

(ii) any Royal Commission or judicial proceeding;

(iii) public safety;

(iv) national security;

(v) economic security; or

(vi) any other matter of public interest;

and/or

(b) is intended to, or could reasonably be expected to do any one or more of the following –

(i) adversely interfere with genuine and open debate on matters of public importance;

(ii) cause widespread financial loss;

(iii) cause unrest between social, political, or religious groups in Australia;

(iv) incite violence against any person or group; or

(v) substantially erode the level of social trust in Australia.

15 AI Disclosure Notices

(1) In this Act, the term '***AI Disclosure Notice'*** means a notice, statement, declaration, or disclaimer, or any combination of these, which could reasonably be taken to communicate to any *ordinary person* who views, hears or otherwise consumes *AI Content* that the relevant *content* was created or modified by an *AI system*.

(2) In deciding whether or not *AI Content* constitutes *Misleading AI Content*, the *prescribed court* must consider the presence of, and the likely effectiveness of an *AI Disclosure Notice*.

(3) Nothing in this Act requires *AI Content* to include an *AI Disclosure Notice*.

(4) The Regulations may set out forms of notices, statements, declarations or disclaimers *AI Disclosure Notices* that are to be deemed to be satisfy the requirements of this section.[13]

Part C — The *Commissioner* and the *MAC Register*

16 Functions of the *Commissioner*

The functions of the *Commissioner* are:

(1) to exercise the functions conferred by this Act;

(2) to monitor compliance with this Act;

(3) to advise and assist persons in relation to their obligations under this Act; and

(4) to maintain, administer, and publish the *MAC Register* online;

(5) to facilitate or enable persons to create, file and execute *MAC Declarations*; and

(6) to refer suspected offences to one or more of the following for investigation or action:

(a) the Australian Federal Police;

(b) the Commonwealth Director of Public Prosecutions; and

(c) the Commonwealth Attorney-General.

17 Powers of the *Commissioner*

The *Commissioner* has power to do all things necessary or convenient to be done for or in connection with the performance of the *Commissioner*'s functions under this Act.

18 The *MAC Register*

(1) The *Commissioner* must keep, maintain, administer, and publish the *MAC Register* in accordance with this Act.

(2) The *Commissioner* must receive *MAC Declarations* filed in accordance with section 25, and enter –

(a) the filed particulars; and

(b) any other details prescribed by the Regulations;

on the *MAC Register*.

(3) The *Commissioner* must not publish or otherwise communicate the particulars listed Column 2 of Schedule 1 Item except as required to do so under the Freedom of Information Act 1982 (Cth).

19 Appeals

(1) An appeal against any decision made by the *Commissioner* under this Act may be to any *prescribed court*.

(2) Any appeal lodged pursuant to section 19(1) must be made within 28 days of the *Commissioner* handing down the relevant decision in writing.

Part D — Deemed Consumer Assumptions[14]

20 Assumption of *Actual Consent*

(1) There is a rebuttable presumption that consumers of *AI Content* assume that any *portrayed person* has provided *actual consent* to the inclusion of their *synthetic likeness* in that *AI Content*.

(2) The presumption in section 20(1) may be rebutted if, and only if, it is evident from the relevant *AI Content* that the *portrayed person* did not provide *actual consent*.

(3) To avoid doubt, if *AI Content* is clearly created by way of parody or satire, then the presumption in section 20(1) is rebutted.

21 Assumption that *content* is not *AI Content*

(1) There is a rebuttable presumption that consumers of all *content* assume that it is not *AI Content*.

(2) The presumption in section 20(1) may be rebutted if, and only if, it is evident from the relevant *AI Content* that the *content* is in fact *AI Content*.

Part E — *Misleading AI Content*

22 Prohibition on the creation of *Misleading AI Content*

(1) An *AI User* must not knowingly cause an *AI System* to create any *Misleading AI Content.*

(2) Without limiting section Part E —(1), an *AI User* must not knowingly cause an *AI System* to create an *unauthorised synthetic likeness* of a *protected person*.

(3) A person must not knowingly modify legitimate *AI Content* in a manner that converts or transforms it into *Misleading AI Content*.

(4) Without limiting section Part E —(3) above, a person converts or transforms legitimate *AI Content* into *Misleading AI Content* by:

(a) removing, cropping or otherwise altering a relevant *AI Disclosure Notice* from the *AI Content*;

(b) altering the relevant *content* in a way that is inconsistent with the consent provided by any *portrayed person*.

(5) The *AI User* has the burden of proving that a *synthetic likeness* is not an *unauthorised synthetic likeness*.

23 Prohibitions on dissemination of *Misleading AI Content*

A person must not knowingly –

(a) disseminate *Misleading AI Content*; or

(b) cause *Misleading AI Content* to be disseminated;

within Australia.

24 *Serious Misleading AI Content*

A person commits an offence if that person knowingly:

(a) causes an *AI System* to create *Serious Misleading AI Content*;

(b) knowingly disseminates *Serious Misleading AI Content*.

Penalty: Imprisonment for 20 years or 5,000 penalty units, or both.

Part F — *MAC Declarations*

25 Filing and publication of *MAC Declarations*

(1) If the *synthetic likeness* of a *portrayed person* appears in any *AI Content* then:

(a) the *portrayed person*; or

(b) any *person with standing*;

has the right to file a *MAC Declaration* with the *Commissioner* in respect of that *AI Content*.

(2) Every *MAC Declaration* filed under section 25(1) must:

(a) be made under penalty of perjury;

(b) contain an acknowledgment that the declarant may be liable to compensate third parties for losses caused by the making of a declaration that is false in any material particular;

(c) be physically signed by the *portrayed person* or the *personal legal representative* of the *portrayed person*; and

(d) be witnessed in person by someone who is authorised to witness statutory declarations.

(3) The *Commissioner* must review each *MAC Declaration* filed under section 25(1) to assess whether the *MAC Declaration*:

(a) sets out all of the details required in Schedule 1;

(b) complies on its face with all of the requirements in section 25(2).

(4) If the *Commissioner* assesses that the *MAC Declaration* satisfies the requirements in section 25(3) then the *Commissioner* must:

(a) issue a filing number to the person who filed the declaration;

(b) publish all details on the *MAC Register* other than those which are specified in Schedule 1 that are not to be published.

26 *MAC Declaration*s and deemed knowledge

(1) A person is deemed to know that *content* is *Misleading AI Content* if:

(a) that person has actual knowledge that the *content* is *Misleading AI Content*; or

(b) that person has been notified of the existence of a valid and subsisting *MAC Declaration* in respect of –

(i) that *content*; or

(ii) *content* that is substantially the same as the *content* specified in the *MAC Declaration*.

(2) A person is deemed to have been notified of a *MAC Declaration* pursuant to section 26(1)(b) if written notice of the existence of a valid and subsisting *MAC Declaration* has been provided to that person through any channel or physical location that the person:

(a) actually uses to communicate with others; or

(b) advertises as being a channel used to communicate with others.

27 Withdrawal and Removal of *MAC Declarations*

Any person who has filed a *MAC Declaration* has the right to withdraw that *MAC Declaration* by filing an application to that effect with the *Commissioner*.

28 Applications to remove *MAC Declarations*

(1) An *affected person* has the right to file an application with the *Commissioner* to remove a *MAC Declaration* under this section.

(2) The *Commissioner* must determine any application to remove a *MAC Declaration* in accordance with the *Regulations*.

Part G — False Statements

29 False *MAC Declarations* and statements – General prohibition

(1) A person must not knowingly make a *MAC Declaration* that is false or misleading in any material particular.

(2) A person must not knowingly represent (whether explicitly or implicitly) that a genuine likeness of himself, herself or themself is an *unauthorised synthetic likeness*.

30 False *MAC Declarations*

A person commits an offence by making a *MAC Declaration* with the knowledge that it is false or misleading in any material particular.

Maximum penalty: 500 penalty units, 3 years imprisonment.

Part H — Remedies

31 Damages

(1) A person who suffers loss or damage by conduct of another person that was done in contravention of any of the following provisions of this Act may recover damages by way of compensation upon order of a *prescribed court*:

(a) section 22 (Prohibition on the creation of *Misleading AI Content*);

(b) section 23 (Prohibitions on dissemination of *Misleading AI Content*);

(c) section 29 (False *MAC Declarations* and statements – General prohibition).

(2) Damages recoverable pursuant to section 31(1) include compensation for non-pecuniary losses, including pain, suffering and damage to reputation.

(3) A court may include an additional amount in an assessment of damages for an infringement of a registered trade mark, if the court considers it appropriate to do so having regard to:

(a) the flagrancy of the conduct; and

(b) the need to deter similar conduct; and

(c) any benefit shown to have accrued to that party because of the infringement; and

(d) all other relevant matters.

(4) An action under section 31(1) may be commenced at any time within 6 years after the day on which the cause of action that relates to the conduct accrued.

32 Other remedies

In addition to the powers to award damages pursuant to section 31, a *prescribed court* may make any other such order that it judges to be appropriate for the purposes of this Act, including:

(a) a declaration;

(b) an injunction (whether prohibitive or mandatory); and

(c) an order that a *MAC Declaration* be removed from the *MAC Register*.

Part I — Defences

33 Safe Harbour Defence

If a person (other than the *AI User* who created the relevant *AI Content*) ceases disseminating *Misleading AI Content* material within 48 hours of being deemed to know that the *content* is *Misleading AI Content*, then that person is not liable for the dissemination of that *Misleading AI Content* under this Act.

34 Defence of *Presumed consent*

If a person knowingly –

(a) causes an *AI System* to create an *unauthorised synthetic likeness* of a *protected person*; or

(b) disseminates *AI Content* that includes an *unauthorised synthetic likeness* of a *protected person*;

in circumstances where the person is entitled to rely upon *presumed consent*, then –

(c) that person is not liable for the creation or dissemination of the relevant *content*; providing that

(d) the person –

(i) takes all reasonable steps to prevent the further dissemination of the relevant *content*;

(ii) deletes all copies of the relevant *content* in his, her or their custody, power or control;

promptly upon learning that the *portrayed person* objects to the creation or dissemination of the relevant *content*.

Part J — Schedule

	Relevant Information	Notes / explanation
1.	Full name of Declarant	
2.	Full address of Declarant*	Address provided must be kept by the *Commissioner*, but must not be published on the *MAC Register*.
3.	Email address of Declarant*	Email address provided must be kept by the *Commissioner*, but must not be published on the *MAC Register*.
4.	Full name of *protected person* depicted in the relevant *AI Content* (if different from the Declarant)	Name of the *protected person* must be kept by the *Commissioner*, but must not be published on the *MAC Register*.
5.	Full address of relevant *protected person* (if the *protected person* is different from the Declarant)*	Address provided must be kept by the *Commissioner*, but must not be published on the *MAC Register*.
6.	Statement of status of the relevant *protected person*	A statement that the relevant person is a '*protected person*', i.e. that they are a 'living person'.
7.	Title of *AI Content* (optional)	The title, subject or heading given in any publication of the relevant *AI Content*
8.	Location of *Misleading AI Content* (URL, social media account)	There may be one or more locations.
9.	Description of misleading content	Brief description of the *content* sufficient to enable a person to identify its nature and effect. For example, it may be 'An audio file apparently depicting me singing the song "Yesterday"' or a 'a video file apparently depicting me vilifying immigrants at a rally'.
10.	Impact Statement (optional)	An optional statement of the impact or potential of the *AI Content* on the *protected person*.
11.	Declaration	a. I declare that the contents of this declaration are true and correct. b. Where I am making this declaration on behalf of another, I have the legal authority to do so. c. I understand that knowingly making an untrue declaration: i. is an offence that is punishable by fines or imprisonment or both; and

		ii. may make me liable to compensate persons who suffer loss as a result of this declaration.
12	Signature of Declarant	Must be signed electronically using a method that complies with the regulations.
13	Signature of Witness	Must be a person who is authorised to witness the signatures of statutory declarations.

1

1

[1] The definition of '*AI System*' is based on the definition of 'artificial intelligence system' in Title I, Article 3 which states 'artificial intelligence system' (AI system) means software that is developed with one or more of the techniques and approaches listed in Annex I and can, for a given set of human-defined objectives, generate outputs such as content, predictions, recommendations, or decisions influencing the environments they interact with'. See: https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:52021PC0206. That definition covers outputs such as 'content, predictions, recommendations, and decisions influencing the environments they interact with' however this Bill is less ambitious, and therefore we only need to include content.

[2] See: Sebastien Bubeck, Sparks of AGI: early experiments with GPT-4, https://www.youtube.com/watch?v=qbIk7-JPB2c&list=PLOHQVs6TucXuTAkNi0fR2hyqEGXi001yX&index=18. AI Systems are more effective when they're allowed to have access to other tools, e.g. image processing programs.

[3] This definition defines the scope of the related concept '*AI Content*'. The reference to 'information' means this Act would not prevent the creation of an android or human-like robot that impersonates another. That seems well beyond the reach of current technology.

[4] We have chosen not to use the word 'video' because that might not include representations in the 'metaverse' and the like.

[5] The definition of '*immaterial contribution*' is included to ensure that the minor input from *AI Systems*, such as improvements to the brightness and contrast of images does not convert a regular content into *AI Content*.

[6] There is a reasonably strong argument that protections should be extended to recently deceased persons, at least where they are survived by immediate family members. Extending protection to non-living persons (recently deceased or long dead) may be more controversial, and might be the subject of a future amendment.

[7] By incorporated personal characteristics, the legislation attempts to cover *AI Content* that is designed to replicate public advisors and counsellors (e.g. Esther Perel and 'AI Esther': https://womensagenda.com.au/life/the-perils-of-the-other-ai-artificial-intimacy/).

[8] This provision is loosely based upon the Privacy Act 1988, although it takes into account the comments of the Australian Law Reform Commission: https://www.alrc.gov.au/publication/for-your-information-australian-privacy-law-and-practice-alrc-report-108/19-consent/background-5/.

[9] The concept of presumed consent has been included to take into account the fact that there are people in close and trusting relationships (e.g. marriages, siblings etc) who may legitimately wish to make AI Content featuring a loved one, but who do not wish to ask for actual consent before doing so. The concept of 'presumed consent' provides a defence to liability under section 34. However, persons relying on presumed consent does so at their own risk, as the onus is on them to establish that it was reasonable to do so.

[10] This proposed legislation relies on the definition of 'ordinary person' rather than 'reasonable person'. Experience in recent years has demonstrated that significant damage to society can be caused by convincing a significant number of unreasonable people to believe untrue things (e.g. QAnon in the United States). Accordingly, restricting the scope of the protections to the 'reasonable person' is likely to prevent the legislation to achieve its stated aims.

[11] This is based upon the principle set out in *Australian Woollen Mills Ltd v FS Walton & Co Ltd* [1937] HCA 51.

[12] This formulation is inspired by the concepts set out in section 18 of the Australian Consumer Law, and sections 10, 44, and 120 of the Trade Marks Act 1995.

[13] Industry input will be required to craft valid AI Disclosure Notices. The DEEP FAKES Accountability Act 2019 (House Bill HR 3230) contains examples of suitable disclosures, and I propose to look at these when drafting up the suggested regulations.

[14] This part sets out provisions designed to preserve important norms that prevailed in the pre-AI era. The idea is that it preserves the expectation that people are viewing genuine, real content unless they are told otherwise.