Digital Platform Regulators Forum

Examination of technology: Multimodal Foundation Models

Working paper

August 2024

Important notice
This document has been prepared by the ACCC, ACMA, eSafety and OAIC in their capacity as members of the Digital Platform Regulators Forum (**member regulators**). The information in this publication is for general guidance only. It does not constitute legal or other professional advice, and should not be relied on as a statement of the law in any jurisdiction. Because it is intended only as a general guide, it may contain generalisations. You should obtain professional advice if you have any specific concern.

The member regulators have made every reasonable effort to provide current and accurate information, but do not make any guarantees regarding the accuracy, currency or completeness of that information.

Parties who wish to re-publish or otherwise use the information in this publication must check this information for currency and accuracy prior to publication. This should be done prior to each publication edition, as member regulator guidance and relevant transitional legislation frequently change. Any queries parties have should be addressed to the DP-REG@eSafety.gov.au

Contents

1. Executive summary

This working paper examines multimodal foundation models (MFMs), a type of generative artificial intelligence (AI) that can process and output multiple data types, such as text, images and audio. It is the third public paper prepared by the Digital Platform Regulators Forum (DP-REG) to understand digital platform technologies and their impact on the regulatory roles of each DP-REG member. This paper discusses some of the implications of this technology for consumer protection, competition, the media and information environment, privacy, and online safety within the digital platform context.

MFMs represent a significant advancement in generative AI. Unlike large language models (LLMs), which focus on text, MFMs can handle multiple data types. For example, MFMs could be used to create an image in response to a text prompt or an image prompt could be used to generate a video or a 3D model.

LLMs have risen to prominence since the launch of ChatGPT in November 2022. However, MFMs with capabilities in image, audio, video and 3D model generation are now increasingly being announced or publicly released. This extension from more commonly used LLMs to MFMs broadens the potential use cases for generative AI, allowing it to be used for a wider range of tasks.

An array of products and services based on MFMs have been launched or are in development. These include applications that enable users to edit images, generate video from images, translate speech or create music. While it is difficult to anticipate the range of potential future applications of this technology, it appears there is potential for widespread adoption by consumers and businesses.

MFMs present both significant opportunities and substantial risks. The combination of multiple modes of generated content can exacerbate existing risks and harms within each DP-REG member's remit that we are already working to address in other parts of the digital economy. For example, here are some specific concerns relevant to each member:

- **ACCC**: Scams and misleading conduct could be exacerbated by deepfake images and videos misrepresenting product functionalities or falsely endorsing products with celebrity likenesses.
- **ACMA**: The spread of misinformation and disinformation in Australia could be intensified by the generation of convincing and realistic images, videos and audio of individuals or events that never occurred. This includes deepfake videos and images or audio of popular figures spreading false information.
- **eSafety**: MFMs can combine images, sound and other elements to create extremely realistic but false depictions of people. This allows individuals to easily generate potentially harmful and illegal content such as non-consensual pornography or child sexual exploitation material.
- **OAIC**: MFMs may use personal information in unexpected ways that are outside the control of the individual.

While this technology has a range of implications for each DP-REG member, it also raises issues that cut across each of our individual areas of responsibility. For example, deepfakes could have implications for online safety, privacy, misinformation, consumer protection and trust in the digital economy.

Existing regulatory frameworks can be used to address harms arising from MFMs. Although this technology may develop in new ways, where frameworks apply, regulated entities across the economy using MFMs remain subject to consumer, competition, privacy, online safety and

media laws or regulations and are expected to comply with their obligations. New requirements, such as online safety codes and standards registered in 2023-24, apply to certain services deploying or providing access to MFMs.

At the same time, some proposed reforms currently under consideration by the Australian Government could further enhance the ability of regulators to address the harms associated with MFMs. The government is considering law reform in relation to consumer protection, competition, privacy, online safety and misinformation and disinformation that will strengthen protections against these harms.

There are broader Australian Government initiatives underway to address AI. For example, the government is investing in developing policies and capability to support the adoption and use of AI technology in a safe and responsible manner. This will include funding to support industry analytical capability and coordination of AI policy development, regulation and engagement activities across government, including to review and strengthen existing regulations in the areas of health care, consumer and copyright law.

This working paper aims to complement and inform broader government work on AI that is underway.

2. Background

This paper supports DP-REG's 2024-26 strategic priorities which include 'understanding, assessing and responding to the benefits, risks and harms of technology, including AI models'.[1] It also serves to enhance collaboration and capacity building among the four members while deepening our understanding of these technologies. This will support our future work, both individually and as part of DP-REG.

Our previous working paper explored the benefits and potential harms of LLMs that generate text. However, applications of generative AI are rapidly expanding into other areas, such as image, audio, and video generation as noted in section 3 below. Given this rapid evolution, it is timely to extend our exploration of these technologies beyond LLMs and consider the impacts of generative AI more holistically.

This development has the potential to bring significant benefits to some in our economy but also to exacerbate online risks related to digital products and services. This paper discusses the possible implications of this technology for consumer protection, competition, the media and information environment, privacy, and online safety. It also aims to complement and inform broader government work on AI.

3. Key insight questions

3.1 What are MFMs and how do they work?

What are multimodal foundation models (MFMs)?
Foundation models are a type of AI[2] model that are trained on broad and diverse data and that can be adapted for a wide range of tasks. Often described as 'generative' AI models, they generate new content such as text, images, audio, and code in response to prompts.[3] These models can be focused on a single data type, known as a single mode, or they can be 'multimodal'. Multimodal Foundation Models (MFMs) are a specific type of generative AI that can process and output multiple data types, such as text, images, audio.[4] LLMs, as explored in our previous working paper, are an example of a foundation model that focuses on a single data type. We recognise that these terms can be contested and difficult to define.

How do MFMs work?

MFMs generate outputs based on inputs or 'prompts'. They use algorithms trained on vast amounts of data, which could include images, audio, video, or text, depending on the specific model. For example, an image generation model might produce an image in response to a text prompt, or an image prompt might generate a video or a 3D model.

By training on vast amounts of data, MFMs learn to predict and approximate relationships between different data types, resulting in outputs that generally appear original, even though they are essentially a synthesis of the existing data used to train the model.

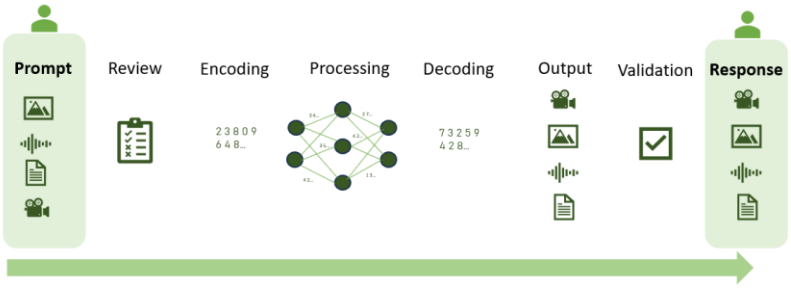
Developing MFMs

Developing MFMs involves several important steps.[5] AI developers first decide on the model's size, measured by the number of parameters, and its architecture, which is the 'topology' or structure of the network. They then gather and prepare vast amounts of training data from a range of sources, converting it into a usable format for training, such as 'tokenising' text or breaking down images into a series of image region features. This data often comes from publicly available sources, possibly gathered from web scraping (extracting data from webpages using software) or open datasets (which are freely available). Developers are also increasingly using proprietary data.[6] Unlike other generative AI models, in MFMs, the model will learn relationships between the different modes of data through this process, such as how text relates to images. Once prepared, the data can be used for 'pre-training', to build the knowledge of the model.

Fine-tuning a model is an additional process AI developers apply to pre-trained models to add particular capabilities or improvements. For example, the model can undergo additional training on specialised datasets to improve its ability to conduct specific tasks such as generating images in a particular style. Fine-tuning may also be used to reduce biased, false, or harmful outputs through human feedback.[7]

Using MFMs

Figure 1: What happens when a user enters a prompt into an MFM



Source: DP-REG analysis

Figure 1 above outlines what happens when a user enters a prompt into an MFM:

1. *Prompt:* A user enters a prompt, which could be in the form of text, image, audio, a combination of these forms or another mode.
2. *Review:* Some systems review the prompt to prevent prohibited or harmful queries.
3. *Encoding:* The model turns the prompt into a series of numbers that capture relevant information about the query (encoding and embedding). These numbers are structured so related or similar queries have similar values. In some models, this string of numbers can include other relevant inputs to provide context, such as the conversation history, user profile information or other relevant data.

4. *Processing:* The encoded numbers are processed through the MFM's mathematical equation, represented as a 'neural network' comprised of layers of 'nodes'. Each 'node' takes in numbers, applies weights (the model's parameters) and returns a result.
5. Processing: As the numbers pass through the layers of the network, they progressively transform from the input numbers to the output numbers. The numerical representations in later layers of the network typically correspond to more concrete and complex concepts, such as specific objects like faces, rather than general shapes or colours.
6. Processing: At the end of the network, a final set of numbers emerges from the final layer of nodes, known as the output layer.
7. *Decoding and output*: The numbers from the output layer are then converted into the output mode (text, image, audio, a combination thereof or another mode) through decoding.
8. *Validation:* Some systems validate the output to ensure it does not contain offensive or prohibited content. The output will be filtered or amended based on this validation.
9. *Response:* Finally, the system provides the output as a response to the user.

Notably, MFM responses are variable in nature. If a user repeats the same prompt, the model will likely produce a different output.

Emerging models and products

Increasing numbers of MFMs and associated products are being announced or publicly released, including image, audio, video and 3D model capabilities.[8] These include, for example, image or video generators such as DALL-E 3, Adobe Firefly, Jasper Art, Synthesia, Midjourney and Stable Diffusion. Audio generators include Riffusion, Suno, Udio, as well as Lyria and AudioCraft (in development by Google and Meta, respectively). Some popular chatbot services that initially operated exclusively as LLMs now offer multimodal functionality. Examples include Microsoft Copilot (formerly Microsoft Bing Chat), Google Gemini (formerly Bard) and MetaAI from Meta. Digital platforms are integrating MFMs into existing services, or releasing or planning to release MFMs, such as Open AI's GPT-4o, Google's Gemini Ultra, and Adobe Firefly.

The availability of MFMs varies on a spectrum from closed to open source. Closed source models are either kept for internal use or licensed to third parties for a fee, allowing them to develop commercial applications but not modify the underlying models. By contrast, developers of open-source models allow third parties to modify them, helping improve the models and correcting errors.[9] Open-source models might restrict certain uses by third parties, such as limiting applications to non-commercial research or imposing other commercial constraints.[10]

Developers may also create various versions of the same model to suit different tasks. For example, Google has optimised its Gemini model for three different sizes (Ultra, Pro and Nano). While larger models are better suited to dealing with novel or highly complex tasks, smaller models can be designed to operate on consumer devices, such as laptops, removing the reliance on cloud computing and reducing latency.[11]

When consumers or businesses use generative AI products and services, they are typically not engaging directly with LLMs or MFMs; rather, they are encountering new kinds of services, applications and businesses that use them, whether in the form of chatbots, enhanced applications or subscription services.[12] In some cases, the creator of the MFM deploys the model directly to create a new product or service (e.g., ChatGPT). It may also deploy an MFM to add features to existing products or services (such as Google Search Generative Experience). In other cases, the MFMs are deployed by third parties, for example, Duolingo uses OpenAI's GPT-4 in its Duolingo Max feature. Users can access a variety of models through 'model

distribution platforms'[13] such as HuggingFace's model library, Google's Model Garden, Amazon Bedrock, Microsoft's Azure Machine Learning, or OpenAI's GPT Store.[14]

3.2 What are some current and projected applications of the technology? What benefits could the technology bring?

LLMs and MFMs have the potential to be applied across the economy, providing individuals and businesses with greater capacity to generate and distribute new content.[15] These models can be integrated into existing products and services to provide new functionality or empower new products and services. As explored in our previous working paper, LLMs can create content, find and summarise information, and are increasingly being integrated into chatbots.

By incorporating more modalities, the capabilities of these models and their potential use cases can be expanded. For example:

- Image: Consumers can use image generation to edit or create new images for content or communication with friends, while businesses can leverage image generation and editing for product design, content production or creating marketing materials.
- Video: Video generation allows consumers to edit and create video content, such as dubbing languages in their videos to reach a wider audience, and businesses can generate marketing or creative content.
- Audio: Audio generation enables consumers and creative workers to produce music, provide speech translation and transcription services, offer reading assistance, and support people who are non-verbal.[16]
- 3D models: Businesses can use 3D models to help design products and prototypes.

Currently, many image, audio and video generation models produce content based on text input, but some models allow for multiple types of input data, such as using an image input to create a video output. As more modalities of input become commonplace, this will further expand the potential applications of MFMs. For example, an MFM-based service could generate audio or text summaries of video input. Models that incorporate increasing numbers of modalities are in development, such as Meta's ImageBind or OpenAI's GPT-4o.[17] As the market evolves, MFMs will be able to process more data when producing outputs, increasing the applications of this technology.

Like the digital and communications transformation in the late 90s and 00s, these models could potentially impact every part of the economy. Goldman Sachs have estimated that generative AI could lead to a 7% increase in global GDP by 2033.[18] However, some media reports have also suggested that business have been cautious in adopting generative AI services, with companies exploring use cases while being mindful of the costs of deploying models and their limitations.[19]

Given these applications are still nascent, it is difficult to assess the take-up of these services in Australia. However, research conducted in six countries between March and April 2024 found that 9% of people have used generative AI to create an image, with 4% making a video and 3% creating audio.[20] The research finds that daily use of generative AI services is rare, with many people having tried them only once or twice. Additionally, the Family Online Safety Institute (FOSI) released research in November 2023 on the emerging awareness, perceptions, and early use of generative AI tools among parents and teens in the US, Germany and Japan.[21] This research found that most parents and teens expect and accept that generative AI will become more embedded and ubiquitous in work, school, and their personal lives. Common Sense Media has also released research exploring teen and young adult perspectives on generative AI, including patterns of use, excitement, and concerns.[22]

MFMs are increasingly being integrated into widely used digital platform products and services, such as Facebook, WhatsApp, Snapchat, TikTok and Microsoft Office which may increase user uptake.

While there is potential for significant adoption of services relying on LLMs and MFMs, it is extremely difficult at this early stage to anticipate the range of potential applications and the extent of uptake.

3.3 What are some overarching limitations of the technology?

Many of the limitations and risks associated with MFMs are similar to those considered in 2023 by DP-REG members in our examination of LLMs.

These limitations should be read alongside those outlined in DP-REG's examination of LLMs and the eSafety Commissioner's August 2023 Tech Trends Position Statement on Generative AI. The position statement highlights potential drivers of risk across a range of online harms.[23] This includes the use of MFMs to produce highly personalised, emotive or potentially manipulative content that can drive specific harms. The statement also notes the risks associated with wider access to generative AI models, the speed of technology development, and the possible convergence with other emerging technologies.

Inaccurate outputs

MFMs may produce inaccurate or inappropriate outputs based on written prompts. This issue, rooted in the underlying architecture of MFMs, may not be completely resolvable.

Inaccurate responses to prompts, known as 'hallucinations', can be generated due to 'prompt engineering', which extracts inaccurate outputs from MFMs. Techniques like 'jailbreaking' or 'adversarial AI' can also generate outputs that depict illegal activities, drug use, or other sensitive content that violates providers' terms or conditions.[24]

Inaccuracy of MFM outputs can also be a consequence of limitations with the technology. Unrepresentative training data can introduce bias and generalisations into the models, which is particularly notable in image and video outputs where racial or gender biases can occur.[25] Even when MFMs are trained with representative data, outputs can still be historically inaccurate.

Inaccurate images of historical events

Hallucinations or confabulations on MFMs have real commercial consequences. In February 2024, it was reported that Google had temporarily paused its Gemini AI model's image generation capability. Media reporting stated that this was because its outputs produced inaccurate/inappropriate images of historical events.[i]

[i] Google to pause Gemini AI model's image generation | CNN Business

User awareness and the *'Uncanny Valley'*

Public reporting on popular MFM tools such as Midjourney, Stable Diffusion and DALL-E has highlighted their ability to produce realistic images of faces and landscapes. However, MFM tools have been criticised for their limited capability to produce realistic outputs of hands,[26] forming part of a broader critique about the 'uncanny valley' effect. This term describes the unsettling similarity when non-human faces resemble human features too closely. In the context of MFMs, image or video outputs that convey too much but at the same time not enough familiar 'humanness' can be unsettling.[27] While more sophisticated updates to the technology are progressively addressing this limitation,[28] it is not clear whether this issue can be completely resolved.

Deepfake detection of voices

AI-generated voices (including voice synthesis) are nearly impossible to differentiate from human speech.[i] In February 2024, the Chief Executive Officer of deepfake detection company, Reality Defender[ii], suggested that variability in voices across regions, languages, dialects and ages made detection a challenging task.[iii]

[i] Creating, Using, Misusing, and Detecting Deep Fakes | Journal of Online Trust and Safety (tsjournal.org)
[ii] Reality Defender — Deepfake Detection
[iii] Why AI-generated audio is so hard to detect (nbcnews.com)

Data poisoning
Some MFM models are trained by indiscriminately scraping online images. This has led to adversarial approaches by artists and others to deliberately contaminate data to prevent their personal styles being scraped for AI training.[29,30]

4. Potential impacts and applicable regulatory frameworks

4.1 Overview

The subsections below consider the range of potential impacts of MFMs on the remits of DP-REG members. A common theme is that MFMs could exacerbate existing and widespread harms, and the material notes how these harms may manifest within each member regulator's areas of responsibility.

Within this section, several cross-cutting issues for DP-REG members are apparent, demonstrating the importance of a collaborative approach to this technology. For example, deepfakes and scams pose issues that affect multiple areas of regulation. There are also common themes to potential harms arising in MFMs. For example, the absence of clear disclosure and labelling may make it difficult for individuals to distinguish between genuine and generated content. MFMs also enable content to be produced at scale and to use personal information to become more persuasive or increase its emotional impact, increasing the risks associated with the spread of misinformation, terrorist propaganda or scams.

At the same time, this discussion of impacts and regulatory frameworks underscores some common challenges relevant to regulators when addressing harms arising from MFMs. MFMs can involve a varied range of actors, such as developers, deployers and users, each potentially subject to different regulatory frameworks. The complex supply chains of MFMs can make it difficult to determine who is accountable and legally liable when things go wrong.

The variable nature of MFM outputs (i.e., that repeating an input prompt is likely to yield a different output) could complicate evidence gathering and the exercise of individual rights. Regulators will need to assess the authenticity of evidence for enforcement, which may be difficult as MFMs can generate false but believable content.

4.2 Australian Competition and Consumer Commission (ACCC)

4.2.1 Consumer protection

The ACCC's consumer protection role includes enforcement of the Australian Consumer Law (ACL) to ensure that consumers and small businesses are protected from misleading and deceptive conduct, unconscionable conduct, unfair terms and conditions and unsafe products, and to promote fair trading. The ACCC also operates the National Anti-Scam Centre (NASC)

and Scamwatch website which helps Australians learn how to recognise, report, and protect themselves from scams.

Misleading/deceptive conduct

The ACL applies to all products or services other than financial products and services, and contains prohibitions on misleading or deceptive conduct, and false or misleading representations.

It is relevant to consider how the use of MFMs in products and services could raise concerns about misleading or deceptive conduct. As an example of inaccurate information provided through generative AI, Air Canada was required to provide a partial refund to a grieving passenger who was misled by an airline chatbot inaccurately explaining the airline's bereavement travel policy.[31] Deepfake images and videos could misrepresent product functionalities or endorsements by celebrities, misleading consumers in their purchasing decisions. Without clear disclosure or transparency about the use of AI, consumers may struggle to distinguish between genuine and generated content. In February 2024, a Scottish Willy Wonka pop-up experience which promised an immersive experience used AI-generated images to promote the event, misleading consumers about the quality of the event.[32] It has been reported that deepfakes have also been created to depict a range of celebrities endorsing products or services, such as Oprah Winfrey appearing to endorse a US influencer's self-help course.[33]

Another potential area of concern is "AI washing", where sellers falsely claim their offerings involve AI technology. The US Federal Trade Commission (FTC) has issued a warning about these types of practices.[34] It is also important to recognise that prohibitions on misleading or deceptive conduct, and false or misleading representations are not positive requirements to be transparent or to be accurate in relation to AI use and outputs.

Scams/fake reviews

There are also a wide variety of ways scams could arise via MFMs. DP-REG's previous working paper on LLMs explored some of the ways that generated text could be used in scams, such as by increasing the volume and sophistication of 'phishing' or romance scams. MFMs can exacerbate these issues.

For example, image, video and audio generation tools could be used to create deceptive content for romance scams, and audio generation could make phone-based scams more convincing by making voices sound more realistic.[35] Scammers could tailor phone calls to sound like distressed family members in urgent need of money.[36] AI-generated fake endorsements in Australia from figures such as Dr. Karl Kruszelnicki, who has built a reputation for promoting knowledge of science, have appeared in ads for health products on Facebook and Instagram, supported by AI-generated spam websites in a coordinated campaign to deceive consumers.[37] In addition, the ability of generative AI to personalise content and influence individual decision making, as discussed below, also creates potential opportunities for scams.

The combination of multiple modes of content to create and fuel scams may exacerbate potential harms by making these schemes more convincing to victims. For example, in February 2024, it was reported that a finance worker from a multinational firm in Hong Kong lost $25 million to a scam which included a combination of an email and a deepfake video call with his colleagues.[38] Concerns have also been raised about fake AI-generated products being available for sale online. For example, media reports have raised concerns about AI-generated books and fake products being sold online.[39]

The ACMA has a cross-cutting regulatory role in relation to telecommunications, and the current development of the Scams Codes Framework discussed in section 5 could be a vehicle to respond to some of these issues.

Potential for other consumer issues

MFMs could potentially raise other consumer issues, such as lengthy and complex privacy policies that prevent consumers from understanding how their data is used. The US FTC has warned businesses against unfairly or deceptively adopting more permissive data practices to enable the use of consumer data for AI training.[40] Consumers may not understand the complexities of AI business models and how their data might be used by suppliers higher up the supply chain. This issue also intersects with the OAIC's remit on the handling of personal information.

In addition, as noted by the emergence of romantic chatbots[41], consumers may become emotionally vulnerable or attached to AI products, raising serious potential for consumer harm.[42] The capacity to iteratively experiment with auto-generated content that modulates human emotion could enable the creation of more effective targeted content, intended to modulate the emotions of a person or people, potentially inhibiting optimal consumer decision making.[43]

Some academic studies have suggested that personalised messages developed by generative AI can be more persuasive and influential than non-personalised content[44], and that individuals may respond differently to a request depending on whether it comes from another person or a generative AI tool, with consequential risks for their decision-making.[45] While these studies were not specific to MFMs, they indicate potential risks to consumers which could arise in the context of MFMs.

Proposed reforms to address unfair trading practices in the ACL, which is being considered by the government, may be critical to addressing potential harms that might arise from AI use.

Product safety

AI has the potential to enhance product safety outcomes for consumers – such as by detecting potential safety issues, improving manufacturing processes, and detecting unsafe product usage.[46] However, generative AI also introduces new safety risks, both physical and non-physical. While a 2021 report by the UK Office for Product Safety & Standards acknowledged that much of the debate about the impacts of AI on product safety is theoretical and that evidence of real-world examples is limited[47], the continued development of generative AI products could raise risks for consumers in future.

The ACCC is actively involved in discussions in international fora on how to promote safe AI design and the potential use of AI by consumer regulators. Discussions have also included challenges AI poses to allocating liability, such as when software updates make products such as smart home systems unsafe.[48]

Concluding comment on consumer issues

The ACCC's Digital Platform Services Inquiry September 2022 interim report found that existing laws do not always adequately address consumer harms online. The report recommended a range of reforms to address these harms, including the introduction of an economy-wide prohibition on unfair trading practices (which would also address similar harms offline).

It has been argued that the nature of MFMs could create enforcement challenges, such as in relation to the attribution of liability when an AI system acts on behalf of a company or when multiple actors in the supply chain have a degree of control over the risk that needs to be managed. Academics have also argued that the variable nature of MFM outputs could also pose

challenges reproducing outputs to prove events occurred.[49] Another potential challenge relates to the authenticity of evidence.[50]

With the growing use of AI in consumer products, the ACCC also notes the application of the ACL to digital products, including AI products and products using AI in their design and/or supply, could be set out more clearly.[51] In the 2024-25 Budget, the Government announced funding for the Treasury to review the application of the ACL in respect of its application to AI.[52]

4.2.2 Competition

The other key mandate of the ACCC is to promote competition by enforcing the *Competition and Consumer Act 2010* (Cth), regulating national infrastructure (such as telecommunications infrastructure), implementing the Consumer Data Right, and undertaking market inquiries as directed by the Treasurer, including in relation to digital platform services.

Effective competition encourages firms to innovate and improve the value of their offerings to consumers, leading to more choice, lower prices, and higher quality products and services. Technological advancements, such as integrating MFMs into digital platform services, can lead to innovative new products and services.

However, MFMs may have features that could result in markets tending towards concentration, as occurred in other digital platform services.[53] The UK Competition and Markets Authority has noted that depending on market developments, this emerging technology could either disrupt incumbents or exacerbate existing competition concerns and create new ones.[54]

Competition occurs at different levels of the supply chain. For example, there is competition between models (e.g., GPT-4 vs Claude) at the upstream level, and between applications of the models (e.g., Chat-GPT vs Le Chat) at the downstream level. If many downstream applications rely on one or two MFMs, the concentration upstream could still have detrimental impacts on competition among downstream applications and outcomes for consumers.

Barriers to entry

Several factors could impact barriers to entry and the extent of concentration at the upstream level of MFM development. It will be important for competition authorities to monitor these markets and consider the extent to which these barriers to entry and expansion materialise in practice.

- **Data:** Developing MFMs typically requires exceptionally large datasets, especially in the pre-training phase. The volume and quality of data required to pre-train a generative AI model from scratch may impact the ability of new players to enter the market.[55] Existing digital platforms with large user bases may have access to large volumes of relevant data (e.g., photo, video or audio repositories or access to a web index) which could be used during pre-training.[56] Another relevant factor is the uncertainty regarding potential enforcement of copyright law relating to the scraping of data for model training and output.[57] Limitations on access to copyrighted data could increase the value of proprietary data.[58] Additionally, the performance of smaller models with high-quality data and the effective use of synthetic data are other considerations.
- **Computing resources:** Access to computational resources, including specialised hardware such as graphical processing units (chips) and supporting infrastructure, is crucial for developing and operating MFMs. These resources are currently concentrated in a small number of firms and countries, some of which are developing their own MFMs.[59] Reports have suggested these resources may be scarce, which may make it more difficult for firms to enter or expand without developing their own source of computing power.[60]
- **Economies of scale:** Economies of scale may arise with MFMs as once developed, MFMs could be used to train other subsequent models. While it may be expensive to initially train a

model, it may be comparatively cheaper to create further models.[61] However, the high computational costs of running models mean that marginal costs may not be as small as in some other digital platform services.

Strategic partnerships/investments

Recent years have seen a range of strategic partnerships between prominent digital platforms and emerging developers of foundation models, such as Microsoft/OpenAI, Google/Anthropic, Amazon/Anthropic and Microsoft/Mistral. These partnerships vary but may include a digital platform providing a developer of foundation models access to cloud compute or monetary investments. They may also enable digital platforms to deploy developers' models in their products or make them available on their model distribution platform. Microsoft also agreed a deal with AI startup Inflection AI to use its models and to hire most of its 70 staff, including its co-founders.[62] Other digital platforms such as Google and Amazon have reportedly completed similar deals to hire staff from AI startups.[63]

These partnerships can potentially bring benefits by providing developers access to compute and capital, enabling firms who operate at different levels of the supply chain to compete more effectively. However, competition authorities in the UK, EU, US and Germany have taken, or are taking, steps to consider the potential competitive impact of these partnerships and whether they could be classified as mergers.[64] Competition authorities internationally are concerned that digital platforms who already hold entrenched positions of market power in existing services may use these partnerships with foundation model developers to steer technological developments in a manner to insulate themselves from competition.[65]

Digital platform ecosystems

Large digital platforms expanding into the AI supply chain benefit from their existing ecosystems, including data, computing power, expertise, chips, and financial resources.[66]

Given that digital platform service providers generate substantial revenue from their core services, they may be able to make investments on a scale that some of their rivals may struggle to match.[67] For example, large digital platforms may be well placed to attract and retain scarce technical expertise,[68] and afford costly IP litigation and copyright licences for works used in training data or indemnification of users.[69]

Risk of anti-competitive conduct

As companies extend their reach into MFMs, where they control key inputs or adjacent markets (such as cloud computing[70] or chips), they may have opportunities to leverage positions of market power into these markets or to enhance a position in a core market.[71] Conduct such as anti-competitive self-preferencing, tying, bundling or refusal of access may have anti-competitive impacts.[72]

The US FTC has noted the potential for 'open first, closed later' tactics, where firms initially use open-source models to attract business, establish steady streams of data, and accrue scale advantages only to later close off their ecosystem to lock-in customers and lock out competition.[73]

It has been argued that data scraping practices could also have anti-competitive impacts. For example, where an AI-generated response, created using scraped data, competes directly with the content creator who produced the scraped information, the original content creator could be deprived of revenue.[74] The US DOJ has also warned AI companies to fairly compensate creators for their content.[75]

Relatedly, in a recent case, the French competition authority noted that Google made it harder for publishers to negotiate fair remuneration for their content as Google did not allow publishers

to carve out their content from Gemini without diminishing how it is displayed on Google's other services.[76] In such cases, there may be concerns that a digital platform may use its existing position of strength in other markets to enhance its position offering MFMs.

The development and deployment of MFMs are still in the early stages, with a range of important outcomes yet to be seen, such as the importance of open-source models for competition and which business models will prevail. The ACCC is considering generative AI in the context of its Digital Platform Services Inquiry. The ninth interim report of this inquiry, due to the Assistant Treasurer by 30 September 2024, will consider competition and consumer issues in relation to general search services in Australia, including the potential impact of generative AI on the competitive landscape in general search services.[77] The tenth and final report of the inquiry, due the Assistant Treasurer by March 2025, will examine potential or emerging issues related to digital platform services in Australia, including potential competition issues in generative AI.[78]

Failing to consider competition issues that may arise in relation to this technology could lead to market concentration and detrimental outcomes for competition and consumers, as previously occurred with existing digital platform services.[79] These developments also underscore the importance of the ACCC's regulatory reform proposals for digital platforms. Should concerns about anti-competitive conduct arise, Australia will need to consider how it responds to harms in a timely manner.

4.3 Australian Communications and Media Authority (ACMA)

The ACMA is the independent statutory authority that regulates broadcasting and some aspects of online content delivered by digital platform services in Australia. It oversees the voluntary Australian Code of Practice on Disinformation and Misinformation and also has powers to combat phone and SMS scams.

4.3.1 Misinformation and disinformation

In Australia, minimising the risk of harm from misinformation and disinformation on digital platforms is the subject of industry self-regulation through the Australian Code of Practice on Disinformation and Misinformation. With the rapid growth and adoption of generative AI technologies, including MFMs, ACMA has called for updates to the code to adequately address the scope of these advancements and their impacts.[80] In response, the code's independent reviewer updated the Best Practice Transparency Reporting Guidelines to seek information about the steps that code signatories were taking to address the impact of AI technologies. Multiple signatories have since reported on AI-related initiatives and policy changes in their latest May 2024 transparency reports.

MFMs have the potential to contribute to the spread of misinformation and disinformation. They can generate convincing and realistic images, videos and audio of individuals or events that never occurred, such as deepfake videos or images of political leaders or authority figures spreading false information. It can also include generating fabricated images or video of events that never occurred[81], or adding false information into depictions of real events. Such uses can reinforce divisive narratives and propagate conspiracy theories.[82]

The increasing sophistication of MFMs poses challenges for digital platforms in implementing systems and processes to detect and label AI-generated misinformation. Nevertheless, industry members are making efforts to identify and label AI-generated images and videos and to detect and address the online distribution of AI content that can negatively impact political processes.[83] For example, the Coalition for Content Provenance and Authenticity (C2PA)[84] develops technical standards to certify the source and history (or provenance) of media content, helping

publishers, creators and consumers to understand the provenance and authenticity of different types of media. Members of the C2PA include Adobe, Google and Microsoft.

4.3.2 Media and broadcasting

AI-generated images and videos can be developed easily and circulated widely. Their rapid use and adoption is having a significant influence on the media and broadcasting sector by impacting how content is created and circulated.

Identifying and combatting misinformation and disinformation is challenging for social media platforms and news organisations because of the vast volume of media, the rapid spread of information, and the often subtle or invisible nature of deceptive edits.[85] For the news sector, there has been evidence of AI-generated fake images being used in some newsrooms.[86] This may exacerbate existing concerns journalists have about unknowingly reporting false stories. In 2022, 26 per cent said they reported on a story that was later found to contain false information.[87]

This is occurring while AI-generated news and information sites are becoming more common and popular. These sites may promote or publish hallucinations or inaccurate audio, visual or video content without intention. AI-generated news and information sites may operate with little or no human oversight. This potentially increases the circulation of false and misleading information in the community.[88] While news outlets traditionally employ editorial standards and human oversight to manage this risk, it is possible that unintentional sharing of MFM generated audio, visual or video content can still occur.

In response, the industry is assessing ways to build resilience to these harms. Many news organisations are investing in experts and innovative detection technology[89] and improving digital and media literacy through responsible AI deployment in their newsrooms. For example, Guardian Australia has started[90] This may include learning techniques to identify inaccurate, inconsistent or unbelievable content produced by MFMs.

4.3.3 Phone and SMS scams

The ACMA regulates telecommunications providers to identify, trace and block scam calls and text messages, including through enforcing compliance with the Reducing Scam Calls and Scam SMS industry Codes (the Code). For example, in February 2024, ACMA took enforcement action against five telcos that sent bulk SMS for failing to comply with multiple anti-scam and public safety rules. In June 2024, a telco was directed to comply with the Code after it breached information-sharing and reporting obligations.

MFMs can create more technically sophisticated and hard-to-detect scams, as noted in the ACCC section above. They can be used with minimal technical skills and by a broad range of bad actors, including criminal organisations. For example, scammers impersonated Sunshine Coast Mayor Rosanna Natoli in fake Skype calls to solicit personal information and bank details and to organise meetings.[91]

Telecommunications companies are increasingly using AI and machine learning to detect and block scams in real time.[92] This trend is likely to continue, with human experts leveraging machine learning to assess and disrupt scam campaigns, potentially helping to counter the increased use of AI by scammers.

4.4 eSafety Commissioner (eSafety)

eSafety is Australia's independent regulator and educator for online safety, representing the Australian Government's commitment to protecting citizens from serious online harms. eSafety's functions are governed by the *Online Safety Act 2021* (Cth), which came into effect in January 2022 and is being independently reviewed in 2024.

4.4.1 Online safety risks and harms

Many online safety risks and harms associated with MFMs are not new but are amplified by this technology. Experts consulted for this paper highlighted the increased accessibility of these tools, often at a low or no cost, enabling misuse to harm others in highly personalised ways. For example, 'nudify' apps, which allow users to 'undress' or 'unclothe' individuals in images to create realistic nude or explicit images without their consent, have become more readily available.

Key online safety risks and harms are outlined below.

Synthetic or AI-generated child sexual abuse material (CSAM)

The Australian Government's interim response to DISR's 'Safe and Responsible AI' discussion paper acknowledges the creation of illegal and harmful content, including child sexual abuse material, as a harm which can be generated and spread by AI.

Through consultations in the development of this paper, academic experts highlighted that the rise of MFMs increases the risk of generating synthetic child sexual exploitation and abuse material. The WeProtect Global Alliance's Global Threat Assessment 2023 identified AI-generated child sexual abuse material as a trend exacerbating the sexual exploitation and abuse of children online. A 2023 report by the Stanford Internet Observatory and Thorn found that generative AI tools are already being used to create realistic computer-generated child sexual abuse material (CG-CSAM). Similarly, in a report on generative AI threats for 2024, online safety technology company ActiveFence highlighted the exploitation of multimodal capabilities which allow for a combination of inputs.[93]

This creates several critical risks and challenges, particularly in identifying victims. As it becomes more difficult to determine whether content is AI-generated, law enforcement agencies and hotlines will face growing challenges in determining whether certain content depicts an actual child who needs to be identified and rescued. Moreover, links have been made between offenders going from viewing imagery online to contact offending.[94] Building on what is available with a single purpose image-based generative AI system, the ability to manipulate actual photos, voices and other depictions of real children using multimodal capabilities to create material that sexualises them is an important challenge.

The resulting harm is complex; even if it becomes clear the content is synthetic or AI-generated, it can still cause immense distress for those whose images are used and shared without their consent. Whether the content is genuine or synthetic does not diminish its potential to cause humiliation, shame, harassment, and intimidation, or being used in sexual extortion.

Terrorist and violent extremist content (TVEC)

MFMs could also further enable the creation and distribution of terrorist and violent extremist content. For example, multi-modal capabilities that analyse social media posts, online interactions, and other data sources could be weaponised by terrorist groups and violent extremists to create tailored propaganda, radicalise individuals, and incite violence.[95]

According to a recent report by Tech Against Terrorism – an independent, public-private partnership working with the tech sector and supported by the United Nations – generative AI is at risk of terrorist exploitation by providing the capability to generate thousands of manipulated variants of a single image or video. These may be capable of circumventing hash-matching and automated detection mechanisms. Terrorist and Violent Extremist (TVE) actors could also repurpose old propaganda using generative AI tools to create 'new' versions which would evade mechanisms for the hash-based detection of the original propaganda. They could also use AI tools to customise messaging and media to scale up the targeted recruitment of specific

demographics, as well as generating completely artificial TVE content in multimodal formats, such as speeches, images, and even interactive environments. It is also possible that TVE actors could leverage AI tools to design variants of propaganda specifically engineered to bypass existing moderation techniques.[96]

Non-consensual intimate image and video deepfakes

Some MFMs can combine images, sound and other elements to create highly realistic but false depictions of people. In particular, ‘de-clothing' or ‘nudify’ apps present a new vector for abuse. Through consultation with academic experts in preparation for this paper, the use of MFMs to generate non-consensual sexual imagery, or deepfakes, was highlighted as a risk because less data is now required to create this content. In the past, deepfake image generators were less convincing and required large amounts of data. Now, convincing deepfakes can be generated using far fewer images of a person.

As with synthetic or AI-generated CSAM, the resulting harm is complex and can cause immense distress for those whose images are used and shared without their consent. Whether the content is genuine or synthetic does not diminish its potential to cause humiliation, shame, harassment, intimidation, or to be used in sexual extortion.

Tech-facilitated abuse and violence, identified as a global problem, may be amplified by generative AI technologies, including those with multimodal capabilities. Recent research by UNESCO demonstrated that both open and closed AI models can modify images to depict people in non-consenting scenarios. Specifically, it showed that multimodal capabilities which allow for image and video generation can be misused to generate images of women in situations they did not consent to, creating a more realistic vector for gender-based abuse involving images.[97] While MFMs are neutral in content output, this example highlights their potential to create realistic vectors for gendered abuse, which predominantly affects women and girls.

Abuse, bullying and harassment at scale

MFMs and their outputs are vulnerable to being exploited to automate personalised online hate, bullying, abuse, and other forms of harassment and manipulation at scale. The Australian Government’s interim response to DISR’s ‘Safe and Responsible AI’ discussion paper acknowledges the potential for new or exacerbated risks to arise where AI interacts with existing harms, systems or legislative frameworks. For example, the generation and spread of online harms such as AI-generated cyber-abuse.

The capacity to generate multiple versions of content, to avoid or undermine existing moderation techniques, complicates 'notice and takedown' approaches, making it challenging to identify and remove harmful content.

In addition, various forms of generative AI-like text, audio, and image can combine to create highly personalised harassment with amplified harmful impacts. Tech-facilitated gender-based violence provides a useful case study. The UNESCO research highlights the potential of multimodal capabilities to generate cyber-harassment templates.[98] These can automate posts across various social media channels, facilitating widespread distribution of harmful content.

Age-inappropriate content

MFMs can also generate age-inappropriate content, such as violent or sexually explicit material, which may be difficult for parents or carers to spot immediately. This may include inappropriate material featuring children's cartoon characters.

4.4.2 Opportunities

While online safety risks and harms are being impacted by emerging technologies such as MFMs, it is also important to consider the online safety opportunities and benefits presented by the development of these technologies. Opportunities include:

Enhanced detection and moderation of harmful content

Emerging technologies such as MFMs offer potential for improving the detection and moderation of harmful content at scale. Academics consulted in the preparation of this paper emphasised this potential, noting that MFMs could reduce the exposure of human moderators to harmful content during review processes. For example, Tech Against Terrorism recently noted that 'generative AI also offers opportunities to augment well-developed content moderation systems.' This includes 'taking lessons from previous iterations of the internet', as 'it is now crucial for tech platforms to combine efforts to mitigate the risk.' [99]

Technical improvements in AI can also present opportunities to enhance educative prompts and nudges. Educative prompts and nudges are already used on social media, for example, to alert users to sensitive content or share additional educative information on a variety of topics. These can be adapted using generative AI technologies. For example, Meta's updates to Instagram to prevent sexual extortion[100] demonstrate how generative AI technologies can be adapted to alert users to sensitive content and provide additional educational information.

Fostering multistakeholder collaboration

The development of MFMs also offers opportunities to enhance multistakeholder collaboration across sectors and jurisdictions. The borderless nature of the internet, datasets, and models necessitates cooperation to combat online harms effectively. The recent work of the Tech Coalition, an alliance of global tech companies working to combat child sexual exploitation and abuse online, highlights this collaboration. In December 2023, the Tech Coalition convened an industry briefing about the impact of generative AI on online child sexual exploitation and abuse (OCSEA). The briefing culminated in several new multi-stakeholder efforts, including red-teaming, information-sharing, an industry classification system, and developing a process to efficiently refer cybertip reports of AI-generated child sexual abuse material to the National Centre for Missing and Exploited Children.[101]

4.4.3 Mitigations and emerging good practice

Recognising the potential for MFMs to create both online safety risks and opportunities, eSafety advocates for a Safety by Design approach, providing the tech industry with meaningful, actionable and achievable guidance. This approach aims to minimise existing and emerging harms, while promoting opportunities to enhance content moderation and detection efforts, as well as multistakeholder collaboration.

The technology industry must take a leading role in mitigating risks and harms by adopting Safety by Design , which is built on three foundational principles: service provider responsibility, user empowerment and autonomy, and transparency and accountability. Technology companies can uphold these principles by incorporating safety measures at every stage of the AI product lifecycle.

Key interventions across the three principles of Safety by Design include:

- appropriately resourced trust and safety teams
- age-appropriate design supported by robust age-assurance measures
- red-teaming and violet or purple-teaming[102] before deployment
- routine stress tests with diverse teams to identify potential harms
- informed consent measures for data collection and use

- escalation pathways to engage with law enforcement, support services, or illegal content hotlines such as eSafety
- real-time support and reporting
- regular evaluation and third-party audits.

More detailed information on these Safety by Design interventions can be found in eSafety's position statement on generative AI.

In April 2024, Thorn (a US-based non-profit organisation established in 2012 to build technology to defend children from sexual abuse) and All Tech is Human (another US-based non-profit organisation committed to solving complex tech and society issues) launched a pledge in collaboration with AI companies, including Amazon, Anthropic, Civitai, Google, Meta, Metaphysic, Microsoft, Mistral AI, OpenAI, and Stability AI, to commit to Safety by Design principles. Their pledge leverages eSafety's Safety by Design initiative and emphasises that regardless of the data modality (text, image, video, audio), Safety by Design must be considered throughout the entire AI product lifecycle.[103]

These multistakeholder efforts highlight the critical need for technology companies to adopt a Safety by Design approach and work collaboratively to enhance safeguards and take practical steps to minimise the risk of harm from MFMs.

4.4.4 eSafety's approach – prevention, protection, proactive and systemic change

Prevention

eSafety provides age-appropriate programs and resources for children, parents and carers, and the whole community. This includes professional learning for educators and supporting the delivery of best practice online safety education. eSafety promotes digital literacy, critical thinking and resilience as part of its education programs.

For example, eSafety's professional learning program for teachers and educators now includes a webinar about online safety considerations for generative AI in educational contexts. eSafety also promotes AI-related issues through community-based work.

Protection

eSafety's regulatory functions and powers are applied in a flexible and integrated way through various schemes that promote compliance and achieve good outcomes for all Australians:

- **Adult Cyber Abuse Scheme:** Under the Online Safety Act, eSafety operates a reporting scheme that gives Australian adults experiencing seriously harmful online abuse somewhere to turn if the online service providers fail to act on their reports.
- **Cyberbullying Scheme:** This world-first scheme, part of the Online Safety Act, extends protections to children being bullied in all online environments.
- **Image-Based Abuse Scheme:** eSafety has regulatory powers to remove and act against the non-consensual sharing of, or threat to share, an intimate image online.
- **Online Content Scheme:** This scheme enables eSafety to carry out a range of actions to address illegal and restricted online content.

eSafety's regulatory schemes cover both real and synthetic child sexual abuse material, deepfake image-based abuse, AI enabled-content used to target Australian children through cyberbullying, as well as adult cyber abuse. eSafety has started to receive reports from the public about AI-driven abuse and has taken regulatory action against an individual for creating and posting deepfake intimate images of Australian women without their consent.

Proactive and systemic change

The Online Safety Act 2021 provides for industry bodies representing eight sections of the online industry to develop codes and standards to regulate 'class 1' and 'class 2' material, also known as illegal and restricted online material. Class 1 and class 2 material ranges from the most seriously harmful online content, such as videos showing the sexual abuse of children or acts of terrorism, through to content which is inappropriate for children, such as online pornography. eSafety can register the industry codes if they meet certain requirements, including providing appropriate community safeguards. If a code does not meet the relevant requirements, eSafety can establish an industry standard for that section of the online industry.

In terms of eSafety's systemic regulatory powers, the codes and standards address online safety issues in eight sections of the online industry across the digital stack. While AI-generated material is treated under the legislation in the same way as 'real' class 1 and class 2 material, the unique risks associated with AI-generated material have necessitated specific requirements in relation to AI-related features. For example:

- Search Engine Services (SES) code: Registered on 12 September 2023 and effective from 12 March 2024, this code requires search engine providers to take steps to ensure AI functionality integrated into search engine services do not return search results that contain class 1 material such as child sexual abuse material (CSAM). It also requires action to reduce accessibility to AI-generated synthetic materials via the search engine service.
- Designated Internet Services (DIS) standard: A standard was prepared after the Commissioner determined an industry-drafted code failed to meet appropriate community safeguards, which is a requirement for registration. The standard includes specific obligations for certain generative AI services to prevent AI features from generating child sexual abuse material and pro-terror material. This includes regularly reviewing and testing models and promptly making adjustments. eSafety registered the DIS standard in June 2024. It is expected to come into effect in December 2024.

The Basic Online Safety Expectations (BOSE) outline the Australian Government's expectations for social media, messaging and gaming service providers, and other apps and websites to take reasonable steps to keep Australians safe online. The Minister for Communications establishes the BOSE through a legislative determination.

- To date, eSafety has issued 27 transparency notices requiring providers to report on the steps they are taking to keep Australians safe online and address unlawful and harmful material and activity. These notices have included specific questions on how they use AI to improve safety on their services, such as detecting and removing child sexual abuse material and grooming, and how they manage AI-related safety risks of AI, such as amplification via recommender systems. Findings are been published on the eSafety website. In March 2024, notices were issued covering, among other things, generative AI in relation to terrorism, extremism, and child sexual abuse, with appropriate information to be published in due course. In July 2024 the first 'periodic' notices were issued to eight providers, requiring six-monthly reports to eSafety for two years, including on generative AI.
- On 30 May 2024, the Minister for Communications amended the *Online Safety (Basic Online Safety Expectations) Determination 2022* through the *Online Safety (Basic Online Safety Expectations) Amendment Determination 2024 (BOSE Determination)* to include an explicit expectation that providers of relevant services with generative AI capabilities will take reasonable steps to:

- consider end-user safety and incorporate safety measures in the design, implementation, and maintenance of generative AI capabilities on the service
- proactively minimise the extent to which generative AI capabilities may be used to produce material or facilitate activity that is unlawful or harmful.

In addition, as part of its work as an anticipatory regulator, eSafety conducts horizon scanning and engages with subject matter experts through its Tech Trends and challenges program. This allows eSafety to identify the online safety risks and benefits of emerging technologies, and to understand the regulatory opportunities and challenges they may present.

In August 2023, eSafety published a position statement on generative AI, examining LLMs and MFMs. This position statement provides an overview of the generative AI lifecycle, examples of its use and misuse, and consideration of online safety risks and opportunities. It also details regulatory challenges and approaches, and provides specific Safety by Design interventions that industry can adopt immediately to improve user safety and empowerment.

eSafety continues to proactively assess developments and impacts of generative AI, considering safeguards through ongoing research, horizon scanning, education, and prevention activities. eSafety maintains engagement with national and international governments and industry to stay ahead of technological advancements.

eSafety also supports broader government inquiries and initiatives related to generative AI.

4.5 Office of the Australian Information Commissioner (OAIC)

The Office of the Australian Information Commissioner (OAIC) is an independent Commonwealth regulator within the Attorney-General's portfolio, established to bring together three functions: privacy (protecting the privacy of individuals under the *Privacy Act 1988* (Cth) (Privacy Act) and other legislation), freedom of information (access to information held by the Commonwealth Government in accordance with the *Freedom of Information Act 1982* (Cth)), and information management (as set out in the *Australian Information Commissioner Act 2010* (Cth)). Its purpose is to promote and uphold privacy and information access rights. Given the focus of this paper on digital platforms rather than government information handling, this subsection focuses on privacy.

4.5.1 Privacy risks

MFMs can raise several privacy risks and challenges. This subsection of the paper discusses the emerging privacy risks, regardless of whether they are within the regulatory jurisdiction of the OAIC. The OAIC's regulatory remit is discussed in more detail in subsection 4.5.2.

Loss of control over the handling of personal information

There are several ways in which MFMs can impact the control individuals have over their personal information.

Data scraping: Many MFMs are trained on publicly available data scraped from the internet.[104] Although some companies attempt to remove personal information from scraped data, commentators note that this isn't industry standard and is difficult to do completely.[105] This means the datasets often contain personal information that was not intended for training MFMs. The loss of control is exacerbated when information about an individual is made publicly available by third parties, such as friends uploading group photos. Also, because scraping makes a copy of someone's data, it limits an individual's ability to remove data that was previously available.[106]

Inference and creation of personal information: MFMs can infer or create additional personal information about individuals without their involvement. This raises privacy risks, especially

when sensitive information is inferred or created. For example, the multimodal nature of MFMs allows for image manipulation, depicting individuals in situations or activities that never occurred, such as in the case of Taylor Swift in early 2024.[107] Without appropriate controls, an MFM may combine information from different datasets it has ingested to reveal information that is sensitive.[108]

Exercise of individual rights: The exercise of individual rights, such as access, correction and erasure, is challenging in the context of MFMs, further impacting an individual's control over their personal information.[109] Depending on their design, MFMs may function more like content generators than search engines, with responses generated based on learned knowledge rather than retrieved from a searchable database.[110] Individuals who are the subject of the information can only identify if their personal information is being used or disclosed by inspecting the original training dataset or prompting the model.[111]

Reusing personal information

Given the volume of data required to train models, entities building or fine-tuning models are incentivised to collect large amounts of data or use their existing data holdings for training. One way of collecting additional data is using information from prompts entered by individuals to train the model.[112] However, individuals may not expect or want their personal information to be used in this way, resulting in further loss of control if options to prevent this are not provided or if users are unaware of such options.[113] Where organisations change their terms or conditions to allow personal information to be used to train models, there may be intersections with the remit of the ACCC.

Opacity in the handling of personal information

There is limited transparency about the datasets used to train MFMs or when an individual's personal information is included in a dataset.[114] This lack of transparency means individuals may not be aware their personal information has been used, or how it has been used.

Disclosure of inaccurate personal information

Similarly to LLMs, MFMs can generate and disclose inaccurate personal information, leading to harm such as reputational damage.[115] This can occur for a variety of reasons, including inaccuracies in training data, incorrect conclusions about user intent, or inaccuracies inherent in the technology.[116]

Data breach risk

MFMs carry an increased risk of data breaches due to the size of the datasets used in training.[117] This makes companies holding these datasets attractive targets for malicious actors.

In addition to traditional methods of attack to gain access to the dataset, MFMs may be vulnerable to cyber attacks aiming to uncover the training data (model inversion), resulting in unauthorised disclosure of personal or sensitive information.[118] The multimodal nature of MFMs can create a larger potential attack surface than more limited forms of generative AI.[119] If this information is intentionally exposed online, it may amount to doxing, intersecting with the eSafety Commissioner's remit regarding the Cyberbullying Scheme and Adult Cyber Abuse Scheme.

MFMs also increase data breach risks as they can facilitate cyber attacks. For example, they can generate plausible phishing attacks or increase code vulnerability if relied on without sufficient checks.[120]

Harmful uses of personal information

MFMs can generate harmful information about individuals without their knowledge or consent, such as deepfakes.[121] The mixed media capabilities of MFMs can produce a greater impact than other generative AI technologies, such as LLMs, as the combination of audio, visuals and text can have a stronger shock value.[122] The generation of fake images and audio through MFMs intersects with the eSafety Commissioner's remit regarding the Cyberbullying Scheme and Adult Cyber Abuse Scheme, and the ACCC's remit concerning misleading or deceptive conduct.

In addition, MFMs can raise concerns about fairness because they can use personal information in ways that produce discriminatory results.[123]

4.5.2 The OAIC's regulatory remit

Introduction

The OAIC regulates the *Privacy Act 1988* (Cth) (the Privacy Act), which contains 13 Australian Privacy Principles (APPs). These principles apply across the entire personal information lifecycle, from collection through to use, disclosure, storage and destruction, including the handling of personal information in MFMs. Entities subject to the Privacy Act must comply with their obligations, regardless of their position in the MFM supply chain.

Scope of privacy protections

The Privacy Act applies to Australian Government agencies, the Norfolk Island administration, and organisations with an annual turnover of more than $3 million. It also applies to certain organisations with an annual turnover below this threshold, such as private sector health service providers, businesses that sell or purchase personal information, credit reporting bodies, and other kinds of organisations set out in the Privacy Act. It does not apply to individuals acting in a personal capacity or to state or territory government agencies.

Personal information, as defined by the Privacy Act, includes any information or an opinion about an identified or reasonably identifiable individual.[124] This encompasses names, telephone numbers, and images or videos where a person is identifiable. The definition is broad, covering situations where information can be reasonably linked with other data to identify an individual. Importantly, personal information retains its classification even if it is incorrect.[125]

Successfully de-identified data is not personal information and generally falls outside of the Privacy Act. For information to be considered de-identified, it must present a very low risk of re-identification, having regard to all the circumstances (and in particular, the context in which the information will be handled, including who will have access to the data, and what other information they might have access to).[126] Essentially, de-identified information should have no reasonable likelihood of re-identification.

Mitigating privacy risks

Several Privacy Act obligations are relevant in the context of MFMs. The application of these obligations will vary depending on the information flows and handling practices involved. Below are some privacy law considerations pertinent to fine-tuning or using MFMs, though this is not an exhaustive list and it does not cover all possible scenarios in which personal information might be handled.

Adopting a 'privacy by design' approach can significantly mitigate privacy impacts arising from MFMs. This proactive process involves integrating good privacy practices into the design specifications of technologies, business practices, and physical infrastructures.[127] It is more efficient to manage privacy risks at an early stage rather than to retrospectively alter a product or service to resolve privacy issues that come to light.

A privacy impact assessment (PIA) is a crucial tool for implementing a privacy by design approach. PIAs systematically assess a project's impact on individual privacy, offering recommendations to manage, minimise or eliminate such impacts. While PIAs focus on compliance with privacy legislation, a best practice approach also considers broader privacy implications and community acceptance of the planned use of personal information.[128]

Designing and training MFMs

This subsection includes a non-exhaustive selection of some key considerations relevant to developers or deployers of MFMs involved in designing and training MFMs. How the APPs apply depends on the flow of personal information between the relevant parties, such as developers, deployers, individuals and third parties.

A developer is an organisation or individual who designs, builds, trains, adapts or combines AI models and applications.[129] This section refers to initial developers, who train and develop an AI model from the ground up, and subsequent developers, who adapt or combine existing models.

A deployer is any individual or organisations that supplies or uses an AI system to provide a product or service.[130]

Each party has obligations under the Privacy Act in relation to their handling of personal information.

Collating the dataset for fine-tuning

Several considerations arise in the process of collecting or creating a dataset for training or fine-tuning an MFM.

- **Data scraping:** When a dataset includes scraped data, it is important for the developer to make sure any personal information within it is lawfully collected. Under the Privacy Act, organisations must not collect personal information unless it is reasonably necessary for their functions or activities.[131] While this does not prohibit the collection of personal information, organisations must consider whether they could perform the same function or activity without collecting the personal information or by collecting less of it.[132] Additionally, since scraped data is not collected directly from individuals, this method should only be used when it is unreasonable or impracticable to collect personal information directly from the individual.[133] Further protections apply to sensitive information under the Privacy Act, which cannot be collected without consent unless an exception applies.[134]
- **Use of previously collected personal information:** When a dataset includes personal information initially collected for a different purpose, it is important for the organisation that collected it to consider whether it can be retained and used to train or fine-tune an MFM. This is relevant whether the organisation is training the model itself or providing the dataset to a third-party developer. Organisations can only use or disclose personal information for the primary purpose for which it was collected unless they have consent or a relevant exception applies.[135] The ability to rely on consent or an exception will also determine whether the dataset can be retained. Organisations must take reasonable steps to destroy or de-identify information that it no longer needs for any purpose for which it may be used or disclosed under the APPs.[136]
- **Transparency obligations:** Regardless of how a dataset is collected, organisations must consider the transparency obligations under the Privacy Act to:
    - have a clearly expressed and up-to-date privacy policy about their management of personal information, including information about the kinds of personal

information collected, how it was collected, and the purposes for which it is collected, held, used and disclosed
- take reasonable steps to notify individuals of certain matters relating to the collection of their personal information.[137]

Accuracy

Accuracy is particularly relevant for MFMs, especially given concerns about presenting false or misleading information as facts. It is essential to take reasonable steps to make sure any disclosed personal information is accurate, up-to-date, complete and relevant.[138] For example, if an MFM generates an output that includes personal information, such as a photo in which an individual is reasonably identifiable, this constitutes a disclosure of personal information. The obligation to take reasonable steps to ensure accuracy may fall on the initial developer, a subsequent developer or the deployer, depending on how the service or product built on the MFM is structured. Taking these reasonable steps in the context of MFMs may involve considering measures such as using high quality training data, limiting the AI model's responses, rigorous testing and human oversight, or contractual obligations to take measures if the organisation disclosing the personal information cannot take measures themselves.[139] The reasonable steps required will vary depending on the circumstances, including the sensitivity of the personal information, the nature of the APP entity holding the personal information, and the possible adverse consequences for an individual if the information's quality is not ensured.[140]

Individual rights

MFMs present challenges for exercising individual rights. In Australia, individuals can seek access to personal information about them that is held by an organisation and request corrections if the information is inaccurate, out-of-date, incomplete, irrelevant, or misleading. An organisation holds personal information if it has possession or control of a record containing that information.[141] In the context of MFMs, this could apply to the initial developer, a subsequent developer or the deployer, depending on how the product or service is structured. Organisations must respond to individual requests and provide access unless an exception applies. They must also take such steps (if any) as are reasonable in the circumstances to correct personal information.[142] However, the transient nature of MFM outputs can make it difficult to exercise these rights, especially if it is difficult to recreate the prompts that generated an output.[143] In addition, the mechanics of how MFMs work can add complexity to addressing these requests.[144] Organisations should develop processes and procedures to enable individuals to exercise their individual rights.

<u>Using MFMs within organisations</u>

When organisations use MFM products or services within their business, the input by employees and output by AI systems of personal information raises privacy considerations. This is the case whether the product or service is developed internally or by a third party.

Inputting personal information

Inputting personal information into an MFM product or service, such as through a prompt, can constitute a use of that personal information if it remains within the organisation. It can be a disclosure if the organisation makes the information available to external parties and releases the information's subsequent handling from its effective control. For example, a disclosure may occur if a third party MFM product or service developer collects prompts to further train their

model. If the primary purpose for collecting the personal information did not include inputting it into the MFM, organisations must obtain consent or rely on an exception to permit the use or disclosure.[145] In addition, disclosing personal information overseas may entail further privacy considerations.[146]

Outputs of MFMs

Outputs of MFMs that contain personal information can enliven privacy obligations related to the collection of information. The concept of 'collects' is broad, and includes gathering, acquiring or obtaining personal information from any source by any means.[147] This includes collection by 'creation' where MFM-generated outputs contain personal information.[148] For example, if someone using the product or service in their workplace asked it to generate a video of a celebrity endorsing their product, this would constitute a collection of personal information about the celebrity. The same considerations regarding valid information collection and transparency obligations discussed earlier apply to generated personal information.[149]

The capacity of some MFM products or services to automatically generate information without human intervention can lead to novel privacy risks, especially when the information captured was not anticipated. For example, if a video call summary generated by an MFM includes discussion of a participant's health condition that was not part of the agenda, this would raise unexpected issues related to collecting sensitive information.[150]

When information generated by MFMs is collected, used or disclosed, organisations must take reasonable steps to ensure it is accurate, up-to-date and complete.[151] What is reasonable will depend on the circumstances. However, given the known risks of incorrect outputs, these steps may include measures to verify the accuracy of personal information.

5. Australian Government developments

Work is underway in a variety of areas across the Australian Government to address issues posed by AI. This brief section highlights broader developments of relevance to the remit of DP-REG members.

5.1 Regulatory initiatives / enforcement actions

Misinformation and disinformation
The ACMA highlighted in its second report on digital platform efforts under the Australian Code of Practice on Disinformation and Misinformation, that the Digital Industry Group Inc (DIGI) and its signatories should review whether the current code adequately addresses the impacts of generative AI technologies.

In 2023, the Australian Government consulted on the draft *Communications Legislation Amendment (Combatting Misinformation and Disinformation) Bill.* This proposed legislation aims to grant the ACMA new information-gathering powers and reserve code registration/standard making powers. The Government is considering amendments to the draft Bill based on the consultation feedback.

Online safety
The Minister for Communications announced in November 2023 that the Government would bring forward the legislated review of the Online Safety Act. The Terms of Reference for the review specify a broad scope, including consideration of whether additional arrangements are warranted to address online harms not explicitly captured under the existing statutory schemes – including potential harms raised by a range of emerging technologies, such as generative AI.

Public consultation for the review began in April 2024 with submissions closing on 21 June 2024. The Issues Paper and further information about the independent review are available on the Department of Infrastructure, Transport, Regional Development, Communications and the Arts (DITRDCA) website.

The Final Report of the Review will be provided to the Minister for Communications by 31 October 2024, for tabling in Parliament within 15 sitting days, as required by section 239A of the Act. Any recommendations made by the Review will be carefully considered by Government and responded to at the appropriate time.

Privacy

The Government Response to the Privacy Act Review will uplift privacy protections, both in relation to automated decision–making and more generally.

The Government has agreed to proposals for privacy policies to set out the types of information used in substantially automated decisions which have a legal, or similarly significant effect on an individual's rights and that individuals should have the right to request meaningful information about how such automated decisions are made. These proposals would enhance protections for substantially automated decision making, including where it is underpinned by MFMs.[152]

In addition to these technology-specific proposals, the Government has agreed-in-principle to proposals aimed at improving personal information handling standards across the economy. This includes establishing a positive obligation on organisations to collect, use, and disclose personal information fairly and reasonably. This will require entities to proactively consider whether their personal information handling activities within MFMs are appropriate.[153] In addition, expanded individual rights will also be introduced, allowing people to request explanations about how their personal information is used or held and to seek deletion or de-identification through a right to erasure.[154]

Competition and consumer protection

On 31 August 2023, The Treasury released a consultation paper on possible reforms to the Australian Consumer Law to address currently unregulated unfair trading practices. The consultation closed in November 2023.

On 30 November 2023, the Government announced a public consultation on the proposed Scams Code Framework. Feedback was sought on the proposed features of the mandatory industry codes outlined in the discussion paper, which would introduce obligations for banks, digital communications platforms and telecommunications providers to combat scams. This consultation closed in January 2024.

On 8 December 2023, the government announced its in-principle support for the recommendations made by the ACCC in its Regulatory Reform report to address competition and consumer harms on digital platforms. The government will undertake further work to implement the recommendations, including consulting on the development of a new ex-ante digital competition regime.

In May 2024, the Government announced funding to support industry analytical capability and coordination of AI policy development, regulation and engagement activities across government, including to review and strengthen existing regulations in the areas of health care, consumer and copyright law.

Other relevant developments across government

- **Safe and responsible AI:** In January 2024, the Australian Government published its interim response to the Department of Industry, Science and Resources' (DISR) discussion paper on 'safe and responsible AI in Australia'. In the May 2024 Federal Budget, the Government

announced it will provide funding over five years from 2023–24 for the development of policies and capability to support the adoption and use of AI technology in a safe and responsible manner. This will include funding to support industry analytical capability and coordination of AI policy development, regulation and engagement activities across government, including to review and strengthen existing regulations in the areas of health care, consumer and copyright law.

- **Productivity Commission research:** On 1 February 2024, the Productivity Commission released three research papers considering how governments can best harness AI for productivity, while anticipating and limiting any associated risks. The second research paper addresses the need for regulation and the kind of regulation and accountability required. It also highlights issues for AI regulation.
- **Senate Select Committee on Adopting AI:** In March 2024, the Senate Select Committee on Adopting Artificial Intelligence (AI) was established to report on the opportunities and impacts for Australia arising from the uptake of AI technologies. The Committee's report is due to the Parliament by 19 September 2024. In May 2024, DP-REG provided a joint submission to this Select Committee.
- **Generative AI in schools:** The Australian Parliament House Standing Committee on Employment, Education and Training is considering the use of AI in education settings. Separately, the Framework for Generative Artificial Intelligence (AI) in Schools is providing guidance to students, teachers, schools and community members on the responsible and ethical use of generative AI tools (including MFMs). One of the principles of the framework is 'Privacy, Security and Safety'.
- **Australian Signals Directorate (ASD) guidance on engaging with AI:** ASD published guidance for organisations on how to use AI systems securely. This guidance summarises important threats related to AI systems and suggests steps organisations can take to engage with AI while managing risk. It also provides mitigations to assist organisations that use self-hosted and third-party hosted AI systems. This could help to inform what are reasonable steps to protect personal information used in AI systems.
- **Copyright and AI Reference Group:** In December 2023, it was announced that the Government is establishing a copyright and AI reference group to better prepare for future copyright challenges emerging from AI. The reference group will be a standing mechanism for ongoing engagement with stakeholders on copyright issues, including the material used to train AI models, transparency of inputs and outputs, the use of AI to create imitative works, and whether and when AI-generated works should receive copyright protection.

6. Overseas developments

In recent times, there has been significant focus on AI and generative AI by regulators and policymakers around the world. Key developments include the EU's AI Act, Canada's Artificial Intelligence and Data Act, the Bletchley Declaration, and the US Executive Order on safe, secure and trustworthy AI.

While these developments are relevant for the broader development of MFMs, in this section, we focus on some examples of approaches taken internationally to address concerns arising within the remit of DP-REG members. This is not intended to be an exhaustive account of international work in these areas.

6.1 Regulatory initiatives/enforcement actions

Consumer protection and competition

Consumer protection and competition authorities worldwide are considering the potential risks and harms to competition and consumers posed by generative AI. For example, the US FTC is

inquiring into whether investments and partnerships pursued by dominant companies risk distorting and undermining fair competition. The UK Competition and Markets Authority recently released its second paper on AI foundation models and has noted a range of risks, such as powerful incumbents exploiting their position to distort choice and restrict competition.[155]

The US FTC has also issued statements to warn businesses against making false or unsubstantiated claims about their products' AI capabilities or unfairly or deceptively adopting more permissive data practices to enable the use of consumer data for AI training.[156] Authorities in a range of other jurisdictions, such as the European Union, Canada, India and France are also considering these issues.[157] G7 competition authorities have also recognised the potential competitive harm that could arise from generative AI and note they are prepared to address the risks that the development and use of AI become dominated by a few players with the market power to prevent the full competitive benefits of AI.[158] In July 2024, the European Commission, the UK Competition and Markets Authority, the US FTC and Department of Justice released a joint statement on competition in generative AI foundation models and AI products.[159]

Misinformation and disinformation

Information integrity remains a concern for Australia and likeminded nations. Several jurisdictions are assessing the impact of MFM technologies on information integrity. For example, the European Union's *AI Act*, *Digital Services Ac*t, and the 2022 *Strengthened Code of Practice on Disinformation* require providers to address the risks posed by MFMs to democratic and electoral processes. On 8 February 2024, the US Federal Communications Commission (FCC) banned robocalls that use voices generated by artificial intelligence[160]. In August 2024, the FCC announced a settlement to resolve enforcement action against US-based telecommunications company Lingo Telecom, who agreed to pay a $1 million (USD) fine for its role in transmitting robocalls that used generative AI voice cloning technology to spread disinformation in connection with a presidential primary election in New Hampshire[161].

Online safety

Internationally, various regulatory approaches are being considered in response to the online safety impacts of generative AI. These include voluntary principles and governance frameworks, application of existing regulations, pledges around self-regulatory principles, dedicated AI legislation, and considerations for synthetic material in online safety frameworks (e.g., Canada's *Online Harms Bill* and the United Kingdom's *Online Safety Act 2023*).

The European Commission has formally sent requests for information under the *Digital Services Act* (DSA) to Bing and Google Search (Very Large Online Search Engines, or VLOSEs), as well as to Facebook, Instagram, Snapchat, TikTok, YouTube, and X (Very Large Online Platforms, or VLOPs).

The Commission is requesting these services to provide more information on their respective mitigation measures for risks linked to generative AI, such as AI 'hallucinations' where AI provides false information, the viral dissemination of deepfakes, and the automated manipulation of services that could mislead voters.

The Commission is also requesting information and internal documents on the risk assessments and mitigation measures linked to the impact of generative AI on electoral processes, dissemination of illegal content, protection of fundamental rights, gender-based violence, protection of minors, mental well-being, protection of personal data, consumer protection, and intellectual property. The questions relate to both the dissemination and the creation of generative AI content. Platforms were required to report to the Commission by 26 April 2024.

In addition, the Global Online Safety Regulators Network (GOSRN) is a forum dedicated to supporting collaboration between independent online safety regulators, with members from

Australia, Fiji, France, Ireland, Netherlands, Republic of Korea, Slovakia, South Africa, and the United Kingdom. GOSRN has established a Technology Working Group to consider the risks and benefits of various technologies, including AI.

Privacy

Data protection authorities worldwide have released or updated guidance to reflect how existing privacy laws apply to generative AI.[162] Several data protection authorities have also commenced regulatory actions relating to generative AI products, such as OpenAI's ChatGPT, or raised concerns about training generative AI on user data.[163]

Some data protection authorities have also taken steps to address practices relevant to generative AI. For example, the OAIC and eleven other data protection authorities globally published a joint statement calling for the protection of people's personal data from unlawful data scraping taking place on social media sites.[164]

7. Conclusion

MFMs have the potential to exacerbate a number of risks and harms relevant to the remit of each DP-REG member. They also raise cross-cutting issues and common challenges. As digital regulators, we are mindful of the intersections between the risks and harms arising within our remits and the benefits of co-operation.

Some aspects of DP-REG members' existing regulatory frameworks can address the harms arising from MFMs. Where these frameworks apply, regulated entities across the economy using MFMs remain subject to consumer, competition, privacy, online safety and media laws or regulations. These entities are expected to comply with their obligations under these frameworks. In some cases, there are also new requirements, such as online safety codes and standards registered in 2023-24, which apply to certain services deploying or providing access to MFMs.

At the same time, some proposed reforms under government consideration could further strengthen protections against these harms. The Australian Government is currently considering potential reforms in relation to consumer protection, competition, privacy, online safety and misinformation and disinformation. The government is also progressing work through a range of other processes, including its work on Safe and Responsible AI. DP-REG members will continue to apply our existing frameworks and engage with government on these issues to ensure the digital economy is a safe, trusted, fair, innovative and competitive space.

8. Acknowledgements

DP-REG members acknowledge the contribution made by experts from the ARC Centre of Excellence for Automated Decision-Making and Society. Their insights on generative AI have been instrumental in preparing this paper. In particular, we thank the following experts:

- Dr. Aaron Snoswell, Queensland University of Technology
- Dr. Ariadna Matamoros Fernandez, Queensland University of Technology
- Dr. Hao Xue, University of New South Wales
- Dr. Jake Goldenfein, University of Melbourne
- Dr. Tegan Cohen, Queensland University of Technology
- Professor Christine Parker, University of Melbourne
- Professor Flora Salim, University of New South Wales
- Professor Jean Burgess, Queensland University of Technology
- Professor Kimberlee Weatherall, University of Sydney

- Professor Mark Sanderson, Royal Melbourne Institute of Technology
- Professor Nicolas Suzor, Queensland University of Technology

9. Endnotes

[1] Digital Platform Regulators Forum, Digital Platform Regulators Forum 2024 communique, 25 July 2024, accessed 1 August 2024.
[2] AI refers to an engineered system that generates predictive outputs such as content, forecasts, recommendations or decisions for a given set of human-defined objectives or parameters without explicit programming. AI systems are designed to operate with varying levels of automation. Department of Industry, Science and Resources, Discussion paper- Safe and responsible AI in Australia, June 2023.
[3] Department of Industry, Science and Resources, Discussion paper- Safe and responsible AI in Australia, June 2023.
[4] Department of Industry, Science and Resources, Discussion paper- Safe and responsible AI in Australia, June 2023.
[5] CMA, AI Foundation Models: Full Report, 18 September 2023, p 10-12.
[6] For example, digital platforms developing models have made agreements to access data from other companies to train their models. For example, this includes Meta and Shutterstock, Google and Reddit, OpenAI and Shutterstock and OpenAI and the Financial times. Google, An expanded partnership with Reddit, 22 February 2024, accessed 1 August 2024; Shutterstock, Shutterstock expands long-standing relationship with Meta, 12 January 2023, accessed 1 August 2024; Shutterstock, Shutterstock expands partnership with OpenAI, signs new six-year agreement to provide high-quality training data, 11 July 2023, accessed 1 August 2024; Financial Times, The Financial Times and OpenAI strike content licensing deal, 29 April 2024, accessed 1 August 2024.
[7] One common approach is called reinforcement learning from human feedback (RLHF). This approach trains the model by using a reward function which punishes 'bad behaviour'. Contractors can be used to outsource this service.
[8] See for example the following list of models in the public domain from the Stanford Center for Research on Foundation Models.
[9] C Carugati, Working paper- Competition in Generative AI Foundation Models, Social Science Research Network, 18 September 2023.
[10] C Carugati, Working paper- Competition in Generative AI Foundation Models, Social Science Research Network, 18 September 2023.
[11] Google, Introducing Gemini: our largest and most capable AI model, 6 December 2023, accessed 1 August 2024.
[12] Australia's Chief Scientist, Rapid response information report- Generative AI: Language models and multimodal foundation models, 24 March 2023.
[13] Please note, 'model distribution platforms' are covered by the eSafety Commissioner's Designated Internet Services (DIS) Industry Standard. The DIS standard was registered in June 2024. More information can be found on the Fact sheet: Registration of the Designated Internet Services Standard
[14] Google, Model Garden on Vertex AI, accessed 1 August 2024.
[15] K Weatherall et al, ADM+S submission to the Safe and responsible AI in Australia discussion paper, 4 August 2023.
[16] Open AI, Navigating the challenges and opportunities of synthetic voices | OpenAI), accessed 1 August 2024.
[17] Meta, ImageBind:Holistic AI learning across six modalities, 9 May 2023, accessed 1 August 2024; OpenAI, Hello GPT-4o, 13 May 2024, accessed 1 August 2024.

[18] Goldman Sachs, The Potentially Large Effects of Artificial Intelligence on Economic Growth, 26 March 2023, accessed 1 August 2024.
[19] The Information, Amazon, Google quietly tamp down generative AI expectations, 12 March 2024, accessed 1 August 2024.
[20] Dr. R Fletcher and Prof R Kleis Nielsen, What does the public in six countries think of generative AI in news?, *Reuters Institute*, 28 May 2024, accessed 1 August 2024.
[21] Family Online Safety Institute (FOSI), 2023 Research Report - Generative AI: Emerging Habits, Hopes and Fears, November 2023.
[22] Common Sense Media, Teen and Young Adult Perspectives on Generative AI: Patterns of Use, Excitements, and Concerns, June 2024.
[23] eSafety Commissioner, Tech Trends Position Statement – Generative AI, August 2023, p 10-12.
[24] A Rizzoli, Jailbreaking ChatGPT's image generator, *Medium*, 12 December 2023, accessed 1 August 2024.
[25] M Zhou et al, Bias in Generative AI, *Cornell University,* (2024).
[26] K Chayka, The Uncanny Failure of A.I.-Generated Hands, *The New Yorker,* 10 March 2023, accessed 1 August 2024.
[27] J O'Meara and C Murphy, Aberrant AI creations: co-creating surrealist body horror using the DALL-E Mini text-to-image generator, *Convergence: the International Journal of Research into News Media Technologies*. 29:4 (2023), p 1082.
[28] B Marr, The Uncanny Valley: Advancements And Anxieties Of AI That Mimics Life, Forbes, 7 February 2024, accessed 1 August 2024.
[29] TJ Thomson and D Angus, Data poisoning: how artists are sabotaging AI to take revenge on image generators, *The Conversation,*18 December 2023, accessed 1 August 2024.
[30] M Heikkila, This new data poisoning tool lets artists fight back against generative AI, *MIT Technology Review,* 23 October 2023, accessed 1 August 2024.
[31] A Belanger, Air Canda must honor refund policy invented by airline's chatbot, *Arstechnica*, 17 January 2024, accessed 1 August 2024.
[32] E Karpen, Willy Wonka event 'so bad people called cops', *News.com.au*, 28 February 2024, accessed 1 August 2024.
[33] J Kelly and L Jones, Piers Morgan and Oprah Winfrey 'deepfaked' for US influencer's ads, *BBC*, 24 February 2024, accessed 1 August 2024.
[34] FTC, Keep your AI claims in check, 27 February 2023, accessed 1 August 2024
[35] A Patty, Mum, help: Nina made three bank transfers before realizing she had been scammed, *The Sydney Morning Herald*, 19 March 2023, accessed 1 August 2024.
[36] A Patty, Mum, help: Nina made three bank transfers before realizing she had been scammed, *The Sydney Morning Herald*, 19 March 2023, accessed 1 August 2024; ABC News, Mom warns of hoax using AI to clone daughter's voice, 13 April 2023, accessed 1 August 2024.
[37] J Purtill, Scammers are using a fake, AI-generated Dr Karl to sell health pills to Australians, *ABC News*, 17 April 2024, accessed 1 August 2024.
[38] CNN, Finance worker pays out $25 million after video call with deepfake 'chief financial officer', 4 February 2024, accessed 1 August 2024; The Guardian, CEO of world's biggest ad firm targeted by deepfake scam, 17 June 2024, accessed 1 August 2024.
[39]C Grady, The AI grift that can literally poison you, *The Vox*, 29 April 2024, accessed 1 August 2024.
[40] FTC, Keep your AI claims in check, 27 February 2023, accessed 1 August 2024; FTC, AI (and other) companies: Quietly changing your terms of service could be unfair or deceptive, 13 February 2024, accessed 1 August 2024.

[41] J Purtill, Replika users fell in love with their AI chatbot companions, then they lost them, *ABC News*, 1 March 2023, accessed 1 August 2024.
[42] K Weatherall et al, ADM+S submission to the Safe and responsible AI in Australia discussion paper, 4 August 2023.
[43] K Weatherall et al, ADM+S submission to the Safe and responsible AI in Australia discussion paper, 4 August 2023.
[44] SC Matz et al, The potential of generative AI for personalised persuasion at scale, *Scientific Reports,* 14:4692 (2024).
[45] T Kim, Hey Siri, how persuasive is AI?, *University of Technology Sydney,* accessed 1 August 2024.
[46] Office for Product Safety & Standards, Study on the Impact of Artificial Intelligence on Product Safety, December 2021.
[47] Office for Product Safety & Standards, Study on the Impact of Artificial Intelligence on Product Safety, December 2021.
[48] Digital Platform Regulators Forum, Joint submission to Department of Industry, Science and Resources' AI discussion paper, 26 July 2023.
[49] K Weatherall et al, ADM+S submission to the Safe and responsible AI in Australia discussion paper, 4 August 2023.
[50] N Raymond, US judicial panel wrestles with how to police AI-generated evidence, *Reuters*, 20 April 2024, accessed 1 August 2024.
[51] Digital Platform Regulators Forum, Joint submission to Department of Industry, Science and Resources' AI discussion paper, 26 July 2023.
[52] The Department of the Treasury, Budget 2024-25: Budget Measures, 14 May 2024.
[53] Digital Platform Regulators Forum, Working paper 2: Examination of technology – large language models, 23 November 2023.
[54] CMA, AI Foundation Models: Full Report, 18 September 2023.
[55] FTC, Generative AI Raises Competition Concerns, 29 June 2023.
[56] C Hogg and D Westrik, Generating Concerns? Exploring Antitrust Issues in the Generative AI Sector, TechREG Chronicle, December 2023.
[57] For example, there are copyright cases in the US in relation to scraping of data which may have implications for copyright law. For example, in a scenario where it is a requirement for MFM developers to acquire licenses to offer services, larger companies may be better resourced to complete licensing deals. Another scenario that could arise is where laws are passed in some jurisdictions which limit the *future* scraping of copyrighted content (but which do not apply retrospectively), this could further entrench the position of providers who already have access to data and raise barriers to entry for potential competitors.
[58] CMA, AI Foundation Models: Initial Report, 18 September 2023.
[59] Australia's Chief Scientist, Rapid response information report- Generative AI: Language models and multimodal foundation models, 24 March 2023.
[60] For example, see AINowInstitute, Computational power and AI, 27 September 2023, accessed 1 August 2024; P Dave, Nvidia chip shortages leave AI startups scrambling for computing power, *Wired*, 24 August 2023, accessed 1 August 2024; CMA, AI Foundation Models - Technical update report, 16 April 2024, p 18; CMA, Cloud services market investigation – Competitive landscape working paper, 23 May 2024, p 143-154; CNBC, Microsoft says cloud AI demand is exceeding supply even after 79% surge in capital spending, 25 April 2024, accessed 1 August 2024.
[61] CMA, AI Foundation Models: Full Report, 18 September 2023.

[62] The Information, Microsoft agreed to pay Inflection $650 million while hiring its staff, 21 March 2024, accessed 1 August 2024.
[63] K Cai, Google hires top talent from startup Character.AI, signs licensing deal, *Reuters,* 3 August 2024, accessed 5 August 2024.
[64] CMA, CMA seeks views on Microsoft's partnership with OpenAI, 8 December 2023, accessed 1 August 2024; CMA, CMA seeks views on AI partnerships and other arrangements, 24 April 2024, accessed 1 August 2024; Bundeskartellamt, Cooperation between Microsoft and OpenAI currently not subject to merger control, accessed 1 August 2024; FTC, FTC launches inquiry into generative AI investments and partnerships, 25 January 2024, accessed 1 August 2024.
[65] For example, the US FTC is inquiring into whether investments and partnerships pursued by dominant companies risk distorting and undermining fair competition while the UK Competition and Markets Authority has requested comments on a range of AI partnerships in recent months; FTC, FTC launches inquiry into generative AI investments and partnerships, 25 January 2024, accessed 1 August 2024; CMA, CMA seeks views on AI partnerships and other arrangements, 24 April 2024, accessed 1 August 2024.
[66] C Caffara, The global AI conundrum for antitrust agencies, *Tech Policy.Press*, 8 February 2024, accessed 1 August 2024.
[67] ACCC, Digital Platform Services Inquiry Seventh Interim Report, 27 November 2023. p27
[68] CMA, AI Foundation Models Initial Report, 18 September 2023. p 39
[69] For example, Microsoft has announced it will pay to defend any customers of its Copilot AI software against copyright lawsuits, as well as the amount of any adverse judgments. Microsoft, Microsoft announces new Copilot Copyright Commitment for customers, 7 September 2023, accessed 1 August 2024.
[70] A range of competition authorities internationally have noted high levels of concentration and potentially anti-competitive conduct in cloud markets among large digital platforms which are also actively developing foundation models and generative AI products and services. For example, several reports have highlighted barriers to switching for users of cloud services. For example, ACM, Market study cloud services, 5 September 2022; Ofcom, Cloud services market study, 5 October 2023.
[71] FTC, Generative AI Raises Competition Concerns, 29 June 2023, accessed 1 August 2024; See also ACCC, Digital Platform Services Inquiry Seventh Interim Report 7, September 2023.
[72] ACCC, Digital Platform Services Inquiry Fifth Interim Report, September 2022.
[73] FTC, Generative AI Raises Competition Concerns, 29 June 2023, accessed 1 August 2024.
[74] C Hogg and D Westrik, Generating Concerns? Exploring Antitrust Issues in the Generative AI Sector, TechREG Chronicle, December 2023.
[75] US Department of Justice, Assistant Attorney General Jonathan Kanter delivers remarks at the promoting competition in artificial intelligence workshop, 30 May 2024, accessed 1 August 2024.
[76] Autorite de la Concurrence, Related rights: the Autorite fines Google 250m for non-compliance with some of its commitments made in June 2022, 20 March 2024, accessed 1 August 2024.
[77] ACCC, Digital Platform Services Inquiry Ninth Interim Report Issues paper, 18 March 2024.
[78] ACCC, Digital Platform Services Inquiry 2020-25, March 2025 final report - Issues paper, 25 July 2025,
[79] C Caffarra, The Global AI Conundrum for Antitrust Agencies, *Tech Policy.Press*, 8 February 2024, accessed 1 August 2024..
[80] ACMA, Digital platforms' efforts under the Australian Code of Practice on Disinformation and Misinformation: Second report to government, July 2023, p 15.

[81] P Marcelo, Fact Focus: Fake image of Pentagon explosion briefly sends jitters through stock market, *The Associated Press*, 24 May 2024, accessed 1 August 2024.
[82] M Goldin, Osama bin Laden was digitally added to a photo of a post-9/11 Pentagon meeting, *The Associated Press*, 23 January 2024, accessed 1 August 2024.
[83] Munich Security Conference, A Tech Accord to Combat Deceptive Use of AI in 2024 Elections, accessed 1 August 2024.
[84] Coalition for Content Provenance and Authenticity, Overview, accessed 1 August 2024.
[85] TJ Thomson et al, Visual Mis/disinformation in Journalism and Public Communications: Current Verification Practices, Challenges and Future Opportunities, *Journalism Practice,* 16:5 (2020).
[86] C Wilson, Adobe is selling fake AI images of war in Israel-Palestine, *Crikey*, 1 November 2023, accessed 1 August 2024.
[87] J Gottfried et al, Journalists highly concerned about misinformation, future of press freedoms, *Pew Research Center,* 14 June 2022, accessed 1 August 2024.
[88] N Sadeghi et al, Tracking AI-enabled Misinformation: Over 800 'Unreliable AI-Generated News' Websites (and Counting), Plus the Top False Narratives Generated by Artificial Intelligence Tools, *NewsGuard*, 10 June 2024, accessed 1 August 2024.
[89] T Johnson, CBS News Launches Venture To Identify AI Deepfakes And Misinformation, *Deadline,* 6 November 2023, accessed 1 August 2024.
[90] M Attard, M Davis and L Main, GEN AI and Journalism, *UTS Centre for Media Transition*, 2023, p 61.
[91] B Dwyer, Scammers use artificial intelligence to impersonate Sunshine Coast mayor as experts warn of video call cybercrime tactic, *ABC News*, 2 May 2024, accessed 1 August 2024.
[92] Telstra, How we use artificial intelligence and machine learning, accessed 1 August 2024.
[93] N Tamari, These Are the Top Generative AI Dangers to Watch for in 2024, *ActiveFence*, 10 January 2024, accessed 1 August 2024.
[94] Internet Watch Foundation, How AI is being abused to create child sexual abuse imagery, accessed 1 August 2024.
[95] ActiveFence, The exploitation of GenAI is innovative and dangerous, accessed 1 August 2024.
[96] Tech against terrorism, Terrorist Use of Generative AI, accessed 1 August 2024.
[97] R Chowdhury and D Lakshmi, "Your opinion doesn't matter, anyway": exposing technology-facilitated gender-based violence in an era of generative AI, *UNESCO*, 2023.
[98] R Chowdhury and D Lakshmi, "Your opinion doesn't matter, anyway": exposing technology-facilitated gender-based violence in an era of generative AI, *UNESCO*, 2023.
[99] Tech against terrorism, Terrorist Use of Generative AI, accessed 1 August 2024.
[100] Instagram, New Tools to Help Protect Against Sextortion and Intimate Image Abuse, 11 April 2024, accessed 1 August 2024.
[101] Tech Coalition, Tech Coalition Hosts Generative AI Briefing for Key U.S. Stakeholders, 11 December 2023, accessed 1 August 2024.
[102] 'Purple teaming' involves combining both offensive and defensive tactics (sometimes referred to as red and blue teaming) to identify, assess, and mitigate risks.
[103] Thorn, Thorn and All Tech is Human Forge Generative AI Principles with AI Leaders to Enact Strong Child Safety Commitments, 23 April 2024, accessed 1 August 2024.
[104] Electronic Privacy Information Centre (EPIC), Generating Harms: Generative AI's Impact & Paths Forward, May 2023, p 25; Office of the Privacy Commissioner of Canada, Principles for responsible, trustworthy and privacy-protective generative AI technologies, 7 December 2023, accessed 1 August 2024; UK ICO, Generative AI first call for evidence: The lawful basis for web

scraping to train generative AI models, 15 January 2024, accessed 1 August 2024; Vox, The tricky truth about how generative AI uses your data, 27 July 2023, accessed 1 August 2024.
[105] Electronic Privacy Information Centre (EPIC), Generating Harms: Generative AI's Impact & Paths Forward, p 24.
[106] Electronic Privacy Information Centre (EPIC), Generating Harms: Generative AI's Impact & Paths Forward, May 2023, p 25.
[107] J Weatherbed, X is being flooded with graphic Taylor Swift AI images, The Verge, 26 January 2024, accessed 1 August 2024.
[108] Electronic Privacy Information Centre (EPIC), Generating Harms: Generative AI's Impact & Paths Forward, May 2023, p 25-27.
[109] UK ICO, Generative AI first call for evidence: The lawful basis for web scraping to train generative AI models, 15 January 2024, accessed 1 August 2024; IBM Ethics Board, Foundation models: Opportunities, risks and mitigations, February 2024, p 10; Confederation of European Data Protection Organizations, Generative AI: The Data Protection Implications, 16 October 2023, p 14-17.
[110] D Zhang et al, Right to be Forgotten in the Era of Large Language Models: Implications, Challenges, and Solutions, 22 September 2023, p 8.
[111] D Zhang et al, Right to be Forgotten in the Era of Large Language Models: Implications, Challenges, and Solutions, 22 September 2023, p 9-10.
[112] IBM Ethics Board, Foundation models: Opportunities, risks and mitigations, February 2024, p 11.
[113] Electronic Privacy Information Centre (EPIC), Generating Harms: Generative AI's Impact & Paths Forward, May 2023, p 25-27.
[114] IBM Ethics Board, *Foundation models: Opportunities, risks and mitigations*, February 2024, p 10.
[115] IBM Ethics Board, Foundation models: Opportunities, risks and mitigations, February 2024, p 19; Confederation of European Data Protection Organizations, Generative AI: The Data Protection Implications, 16 October 2023, p 4; D Zhang et al, Right to be Forgotten in the Era of Large Language Models: Implications, Challenges, and Solutions, 22 September 2023, p 6.
[116] Confederation of European Data Protection Organizations, Generative AI: The Data Protection Implications, 16 October 2023, p 4; D Zhang et al, Right to be Forgotten in the Era of Large Language Models: Implications, Challenges, and Solutions, 22 September 2023, p 6; E Salvaggio, The Hypothetical Image – The Aestheticization of Algorithmic Ideologies, 12 November 2023.
[117] Australia's Chief Scientist, Rapid response information report- Generative AI: Language models and multimodal foundation models, 24 March 2023.
[118] Confederation of European Data Protection Organizations, Generative AI: The Data Protection Implications, 16 October 2023, p 12.
[119] See Zhou et al, Revisiting the Adversarial Robustness of Vision Language Models: a Multimodal Perspective, 30 April 2024; Chang et al, Adversarial Testing for Visual Grounding via Image-Aware Property Reduction, 2 March 2024; Yin et al, VLAttack: Multimodal Adversarial Attacks on Vision-Language Tasks via Pre-trained Models, 5 February 2024.
[120] Electronic Privacy Information Centre (EPIC), Generating Harms: Generative AI's Impact & Paths Forward, May 2023, p 30-32; Hannah Murphy, 'AI: a new tools for cyber attackers – or defenders?', *Financial Times*, 21 September 2023, accessed 1 August 2024; IBM Ethics Board, Foundation models: Opportunities, risks and mitigations, February 2024, p 21.
[121] Electronic Privacy Information Centre (EPIC), Generating Harms: Generative AI's Impact & Paths Forward, May 2023, p 9-17; H Mort, I felt numb – not sure what to do. How did deepfake

images of me end up on a porn site?, *The Guardian*, 28 October 2023, accessed 1 August 2024; IBM Ethics Board, Foundation models: Opportunities, risks and mitigations, February 2024, p 12, 20.
[122] L Weidinger et al, Sociotechnical Safety Evaluation of Generative AI Systems, 31 October 2023, p 13.
[123] The Alan Turing Institute, Data Protection, AI and Fairness; IBM Ethics Board, Foundation models: Opportunities, risks and mitigations, February 2024, p 9,12.
[124] Privacy Act 1988 (Cth) s 6(1).
[125] Privacy Act 1988 (Cth) s 6(1).
[126] See OAIC, Guide to data analytics and the Australian Privacy Principles, 21 March 2018, Part 1.6.
[127] OAIC, Privacy by design, accessed 1 August 2024.
[128] OAIC, Guide to undertaking privacy impact assessments, 2 September 2021.
[129] Department of Industry, Science and Resources, *Safe and responsible AI in Australia: Proposals paper for introducing mandatory guardrails for AI in high-risk settings*, September 2024.
[130] Department of Industry, Science and Resources, *Safe and responsible AI in Australia: Proposals paper for introducing mandatory guardrails for AI in high-risk settings*, September 2024.
[131] *Privacy Act 1988* (Cth) APP 3.2; OAIC APP guidelines [3.17]-[3.21].
[132] *Privacy Act 1988* (Cth) APP 3.1; APP 3.2.
[133] *Privacy Act 1988* (Cth) APP 3.6.
[134] *Privacy Act 1988* (Cth) APP 3.3. Sensitive information' is a subset of personal information and is defined as: information or an opinion (that is also personal information) about an individual's racial or ethnic origin, political opinions, membership of a political association, religious beliefs or affiliations, philosophical beliefs, membership of a professional or trade association, membership of a trade union, sexual orientation or practices, or criminal record; health information about an individual; genetic information (that is not otherwise health information); biometric information that is to be used for the purpose of automated biometric verification or biometric identification; or biometric templates – see *Privacy Act 1988* (Cth) s 6(1).
[135] *Privacy Act 1988* (Cth) APP 6.1.
[136] *Privacy Act 1988* (Cth) APP 11.2; OAIC, APP Guidelines, Chapter 11 [11.22]-[11.45].
[137] *Privacy Act 1988* (Cth) APP 1; APP 5.
[138] *Privacy Act 1988* (Cth) APP 10.
[139] IBM, What are AI hallucinations?, accessed 1 August 2024.
[140] OAIC, APP guidelines – Chapter 10, [10.6]-[10.7]
[141] *Privacy Act 1988* (Cth) s 6(1), definition of holds.
[142] *Privacy Act 1988* (Cth) APP 12, APP 13.
[143] See UK ICO, How do we ensure individual rights in our AI systems?, chapter in Guidance on AI and Data Protection; Dawen Zhang et al, *Right to be Forgotten in the Era of Large Language Models: Implications, Challenges, and Solutions*, 22 September 2023; Confederation of European Data Protection Organizations, *Generative AI: The Data Protection Implications*, 16 October 2023, p 14-17; Dawen Zhang et al, *Right to be Forgotten in the Era of Large Language Models: Implications, Challenges, and Solutions*, 22 September 2023.
[144] See UK ICO, How do we ensure individual rights in our AI systems?, chapter in Guidance on AI and Data Protection; Dawen Zhang et al, *Right to be Forgotten in the Era of Large Language Models: Implications, Challenges, and Solutions*, 22 September 2023; Confederation of European Data Protection Organizations, *Generative AI: The Data Protection Implications*, 16

October 2023, p 14-17; Dawen Zhang et al, *Right to be Forgotten in the Era of Large Language Models: Implications, Challenges, and Solutions*, 22 September 2023.
[145] Privacy Act 1988 (Cth) APP 6.1.
[146] Privacy Act 1988 (Cth) APP 8.
[147] OAIC, APP guidelines, Chapter B: Key concepts, [B.30].
[148] See similar concepts in OAIC, Guide to data analytics and the Australian Privacy Principles, section 2.2.
[149] See Privacy Act 1988 (Cth) APP 1, APP 3, APP 5.
[150] See similar examples in OVIC, *Public Statement: Use of Microsoft 365 Copilot in the Victorian public sector*, OVIC, October 2023.
[151] Privacy Act 1988 (Cth) APP 10.
[152] Attorney-General's Department, Government response to the Privacy Act Review report, 28 September 2023, p 11.
[153] Attorney-General's Department, Government response to the Privacy Act Review report, 28 September 2023, p 8.
[154] Attorney-General's Department, Government response to the Privacy Act Review report, 28 September 2023, p 18.
[155] FTC, FTC launches inquiry into generative AI investments and partnerships, 25 January 2024, accessed 1 August 2024; CMA, AI Foundation Models: Update paper, 16 April 2024; CMA, CMA seeks views on AI partnerships and other arrangements, 24 April 2024, accessed 1 August 2024
[156] FTC, Keep your AI claims in check, 27 February 2023, accessed 1 August 2024; FTC, AI (and other) companies: Quietly changing your terms of service could be unfair or deceptive, 13 February 2024, accessed 1 August 2024.
[157] European Commission, Commission launches calls for contributions on competition in virtual worlds and generative AI, 9 January 2024, accessed 1 August 2024; Competition Bureau Canada, Artificial intelligence and competition – Discussion Paper, 20 March 2024; Competition Commission of India, Competition Commission of India (CCI) invites proposal for launching Market Study on Artificial Intelligence and Competition in India, 22 April 2024; Autorite de la Concurrence, Generative artificial intelligence: the Autorité starts inquiries ex officio and launches a public consultation open until Friday, 22 March, 8 February 2024.
[158] G7 Hiroshima Summit, G7 Competition authorities and policymakers summit digital competition communique, 8 November 2023.
[159] CMA, Joint Statement on competition in generative AI foundation models and AI products, 23 July 2024, accessed 1 August 2024.
[160] See Federal Communications Commission, FCC Makes AI-Generated Voices in Robocalls Illegal, 8 February 2024, USA.
[161] See Federal Communications Commission, FCC Settles Spoofed AI-Generated Robocalls Case, 21 August 2024, USA.
[162] See Confederation of European Data Protection Organizations, Generative AI: The Data Protection Implications, 16 October 2023; UK ICO, Guidance on AI and Data Protection, 15 March 2023; UK ICO; ICO consultation series on generative AI and data protection; Office of the Privacy Commissioner of Canada, Principles for responsible, trustworthy and privacy-protective generative AI technologies, OPC (Canada), 7 December 2023; Office of the Privacy Commissioner of New Zealand, Artificial Intelligence and the IPPs, OPC (NZ), 21 September 2023; Commission Nationale Informatique & Libertés, *AI how-to-sheets*, CNIL, 2024.
[163] See Future of Privacy Forum, How Data Protection Authorities are de facto Regulating Generative AI, 12 September 2023, Garante per la protezione dei dati personali, ChatGPT:

Italian DPA notifies breaches of privacy law to OpenAI, 29 January 2024; Chee FY, 'Meta pauses AI models launch in Europe due to Irish request', *Reuters*, 15 June 2024, accessed 1 August 2024; Mascellino A, 'Meta Faces Suspension of AI Data Training in Brazil', *Infosecurity Magazine*, 4 July 2024, accessed 1 August 2024.

[164] OAIC et al, *Global expectations of social media platforms and other sites to safeguard against unlawful data scraping*, OAIC website, 24 August 2023.