



# **Australian Information Industry Association**

**Submission on**

**Copyright and AI Transparency**

**25 October 2024**

## Introduction

The Australian Information Industry Association (AIIA) welcomes the opportunity to contribute to the Copyright and AI Reference Group (CAIRG) and the Attorney General's Department (AGD) regarding Copyright and AI Transparency. As AI continues to play an increasingly pivotal role in the creation and dissemination of creative content, it is essential to protect Artists' Right of Action and original work of cultural significance and ensure transparency around the use of copyrighted materials in AI systems.

## Summary of Key Positions

Balanced Disclosure Requirements: The AIIA advises against indiscriminate disclosure of AI training data, stressing the importance of protecting trade secrets while fostering transparency. The EU AI Act offers a useful model for Australia by promoting a proportionate approach to AI input transparency.

Fair Use in AI Training: The AIIA supports the use of copyrighted material for AI training under fair use principles, provided the output significantly alters the original material and does not negatively impact the economic rights of creators.

Exemptions for Text and Data Mining: Drawing from international examples, the AIIA advocates for Text and Data Mining exemptions, enabling lawful access to copyrighted materials for prescribed purposes without excessive barriers.

Content Provenance Protections: The AIIA calls for regulations that ensure the preservation of content provenance metadata across platforms, to enhance transparency, accountability, and the protection of creators' rights.

Clarification of Copyright in AI-Generated Works: The AIIA highlights the need for updates to Australian copyright law to clearly define the ownership of AI-generated content, ensuring proper recognition of human contributions in co-created works.

## Context

The AIIA is keen to ensure that the Copyright regulatory framework:

- Enables AI innovation while protecting the interests of both businesses and artists
- Facilitates proportionate AI input transparency to safeguard businesses' trade secrets
- Protects original works – especially those of cultural significance – and artists' right of action
- Thwart deepfakes for nefarious applications and/or disinformation and misinformation

## Copyright Considerations of AI Outputs

- 1. Are there any copyright-related transparency requirements regarding AI inputs that have been adopted or proposed overseas that you think could serve as a model for Australia? Conversely, are there any differences between Australia's copyright law and the copyright laws of other countries that in your opinion mean transparency requirements adopted or proposed in those countries would not be suitable in Australia?**

The AIIA cautions against mandating indiscriminate disclosure of training data sources, datasets, and collection processes, whether through amendments to the Copyright Act or the introduction of AI Mandatory Guardrails. In our submission on the [Proposals for Mandatory Guardrails for High-Risk AI Settings](#), we outlined concerns related to guardrails 3 and 6, particularly with respect to the disclosure of commercially sensitive and confidential information. Any disclosure requirements should be applied thoughtfully and selectively, ensuring the protection of trade secrets and proprietary methodologies. A broad disclosure mandate could expose valuable intellectual property, which is critical for maintaining a competitive advantage in the rapidly evolving AI industry, ultimately undermining business competitiveness and discouraging innovation.

The EU AI Act provides a valuable example of such a balanced approach to disclosure requirements by recognising the challenge companies face in promoting transparency around AI inputs while safeguarding their trade secrets. The EU AI Act states that model providers are required to “draw up and make publicly available a sufficiently detailed summary of the content used for training”,<sup>1</sup> suggesting that this summary should be “generally comprehensive in its scope instead of technically detailed.”<sup>2</sup> One recommended approach is to specify the “main data collections or sets that went into training the model, such as large private or public databases or data archives, and by providing a narrative explanation about other data sources used.”<sup>3</sup> Additionally, the EU proposes that these explanations could follow a standardised format, with a template provided by the EU AI Office for model providers.

Aligning Australia’s copyright-related transparency requirements with those outlined in the EU AI Act would not only establish a sensible approach to disclosure but also foster international consistency in AI regulation. Given that AI technology operates on a global scale, it is vital for Australia to develop policies that are in harmony with international frameworks to ensure Australian developers and businesses can compete and collaborate effectively in the global market. International consistency in regulations reduces the complexity of compliance for companies working across borders, streamlining operations and reducing the costs associated with adhering to varying legal standards.

2. **Are there any copyright-related transparency issues (including, but not necessarily limited to, those raised through the CAIRG to date) that you think the Government should take action on, but that you think would not be adequately addressed if mandatory guardrails 3 and 6 were implemented as proposed in the Proposals Paper? If so, what do you see as the gaps?**
  - a. **For example, are there uses of copyright material as AI inputs, or ways in which AI systems may be used in the production of new materials, that may not fall within scope of the mandatory guardrails because they do not meet the definitions of ‘high-risk AI’ or GPAI as discussed in the Proposals Paper?**

---

<sup>1</sup> European Union, *Artificial Intelligence Act*, Article 53(1)(d).

<sup>2</sup> *Ibid*, Recital 107.

<sup>3</sup> *Ibid*.

### Fair Use

General-purpose AI models, which are designed to perform a wide variety of tasks, often require large and diverse datasets to train effectively. These datasets may include copyrighted materials, but the primary purpose of incorporating such content is to enable the AI to learn, identify patterns, and improve its performance across multiple domains. In these cases, the goal can be educational and analytical in nature, rather than the direct exploitation or reproduction of the copyrighted content. While there is a potential for economic losses in such scenarios, these risks may not meet the threshold of 'high-risk AI,' where there are severe or significant adverse impacts on an individual's human rights, physical or mental health, safety, or reputation, as required by principles a-c and f in the Proposals Paper for Introducing Mandatory Guardrails for AI in High-Risk Settings. It's important to note that not all AI types carry the same risk of copyright infringement; generative AI, which creates new content, is more likely to pose concerns than traditional or general-purpose AI systems, which primarily process existing data without generating original works.

The use of copyrighted material in this context generally aligns with fair use principles because the act of training is transformative—AI systems are not designed to simply replicate or redistribute the material, but rather to learn from it and generate new, independent outputs. The training process does not aim to exploit the original content for commercial gain, but rather to develop sophisticated AI models capable of creating unique and diverse results across different applications. This distinction is critical because it mitigates the risk of infringing upon the economic rights of the original copyright holders.

#### **Example of AI Utilisation for Copyrighted Material Benefiting Authors**

Effective AI training and application have the potential to enhance the visibility and use of original copyrighted content. For instance, a researcher could leverage an AI-powered search engine to recommend and summarise literature that contains specific references. This technology enables researchers to work more efficiently by scanning a wide range of global sources and identifying the most relevant materials for further study, ultimately benefiting both the users and the original content creators.

### Development of New Digital Services and Benefits to Consumers

A growing number of leading international jurisdictions have adopted similar approaches.<sup>4</sup>

<sup>5</sup> One such example is Singapore's computational data analysis exception, which allows both commercial and non-commercial organisations to use copyright-protected works for computational data analysis, provided they are lawfully accessed.<sup>6</sup> This exception permits the extraction of data for analytical purposes, rather than for direct commercial replication, and allows organisations to conduct such analysis for any purpose, including

<sup>4</sup> European Union, *Directive 2019/790 on copyright and related rights in the Digital Single Market*, Article 4, 2019.

<sup>5</sup> Japan, *Copyright Act*, Article 30-4, Act No. 48 of 1970, as amended.

<sup>6</sup> Singapore, *Copyright Act 2021*, Section 244.

commercial use. In a comprehensive study of the 350 submissions to two previous governmental inquiries on introducing fair use to Australian Copyright Law, the top three pro-fair-use arguments are to protect consumers rights, legalise digitisation for preservation or sharing and support the development of new services.<sup>7</sup> To ensure that Australia remains competitive on the global stage, it is essential to establish a framework that allows for the lawful and fair use of data, including copyrighted materials, during the AI training phase. Imposing restrictive access to such data would place Australian developers at a significant disadvantage compared to their international counterparts, limiting their ability to innovate and compete in the rapidly evolving AI landscape.

#### Ensuring Fair Compensation for Use of Content and Punishing Copyright Infringement

While the training phase should be protected under fair use, the potential for copyright infringement increases when considering the outputs of these AI systems. If an AI system generates content that is substantially similar to the copyrighted material used in its training and this content could serve as a direct competitor or substitute for the original, this can undermine the commercial interests of the copyright holder – an argument that has previously been raised against fair use in previous inquiries.<sup>8</sup> In such cases, existing copyright laws should be updated to punish users misusing Generative AI for commercial gains, thereby preserving the rights of the creators and rights holders.

As AI continues to advance, it is crucial to find a balance between fostering innovation and protecting the rights of copyright holders. This approach was reinforced by the UK Government Chief Scientific Advisor, who recommended that “if the government’s aim is to promote an innovative AI industry [...], it should enable mining of available data, text, and images (the input) and utilise existing protections of copyright and IP law on the output of AI.”<sup>9</sup> This balanced approach supports innovation by allowing widespread access to data for AI development while ensuring that copyright owners’ rights are protected when it comes to how their works are used beyond the input stage. Imposing excessive restrictions on the use of copyrighted materials for AI training could hinder technological development, particularly for smaller companies and start-ups that may lack the resources to license large datasets. This could lead to monopolisation in AI development, where only large corporations have the means to develop sophisticated models, stifling competition and limiting access to advanced AI technologies. To encourage an inclusive and competitive AI industry, policies should allow for the fair use of copyrighted material during the training phase while maintaining strong protections for creators regarding AI-generated outputs.

#### **b. Is there in your view a case for copyright-related AI transparency issues to be regulated differently to the approach proposed through the Proposals Paper? If so, why?**

---

<sup>7</sup> Patricia Aufderheide and Dorian Davis, Contributors and Arguments in Australian Policy Debates on Fair Use and Copyright: The Missing Discussion of the Creative Process, *International Journal of Communication* (2017) 532 ([link](#)).

<sup>8</sup> Ibid 534.

<sup>9</sup> Government Chief Scientific Adviser. (2023, March). *Pro-innovation regulation of technologies review: Digital technologies* (p. 9). HM Government. Retrieved from <https://www.gov.uk/government/publications>.

Not all data are created equal so not all data requires the same level of transparency

Mandatory guardrails 3 and 6 refer to data in a simplistic manner, focusing on the good and bad (biased) types of data. In principle, the AIIA emphasises the need for Governments to fundamentally reconsider the classification of ‘data’ – including the different types of copyrighted information - and the development of a strategy that provides the appropriate and proportionate transparency requirements and safeguards for their use in training AI. The different types of copyrighted information include:

- licensed content
- content where permission is not required by law
- content that does not require a license because copyright has expired
- content with a license that permits AI training to train an AI model

In these examples, limited, descriptive disclosure would be appropriate and proportionate AI transparency approach.

**3. Are there specific provisions in the Copyright Act, or in any other Act, that you think the Government should consider as a model for implementing transparency requirements in the Copyright Act (whether to implement the proposed Guardrails 3 or 6, or otherwise)?**

Exemptions for Text and Data Mining (TDM)

Many jurisdictions have granted exemptions for Text and Data Mining based on the types of copyrighted information involved and the reasonable use cases for such activities: <sup>10</sup>

- when there are lawful access and for the sole purpose of research for a non-commercial purpose (s 29A of the Copyright, Designs and Patents Act 1988 in the United Kingdom)
- for the purposes of scientific research (EU Directive on Copyright in a Single Digital Market (2019/790/EU))
- for the purpose of retention of the copies for archiving and backup (Art 24d of the Copyright Act in Switzerland)
- for computation data analysis and where communications to the public occurs only in limited circumstances (ss 243-4 of the Copyright Act 2021 in Singapore)

In these cases, these jurisdictions only require the proof of permission granted and sufficient acknowledgement.

**Copyright Considerations of AI Outputs**

**4. Are you aware of specific methods or techniques for providing transparency about AI outputs that are being adopted in, or could in future be appropriate for, creative industries’ value chains?**

Protecting Content Provenance

In the digital age, content provenance metadata serves as a vital tool for protecting the rights of creators and ensuring transparency in the use of digital works. Metadata contains key

---

<sup>10</sup> Intellectual Property Office of Singapore and Singapore Management University, When Code Creates: A Landscape Report on Issues at the Intersection of Artificial Intelligence and Intellectual Property Law 88 - 91 ([link](#)).

information about the origin, ownership, and licensing terms of a piece of content, helping to prevent copyright infringement and misuse. Without these embedded details, the risk of unauthorised replication, distribution, or manipulation increases significantly. Provenance metadata is already being embedded by creators in both synthetic and non-synthetic digital content. However, many online platforms regularly strip this crucial information from the digital content they display, depriving users of important context regarding the origins and ownership of the material. The removal of this metadata undermines transparency and diminishes the ability to ensure proper attribution and accountability in the digital environment.

We recommend the implementation of regulations requiring online platforms to maintain and display content provenance metadata embedded in digital content on their systems. Prohibiting the deliberate removal of provenance metadata will safeguard the rights of content creators, prevent unauthorised use, and ensure proper attribution of digital works.

**5. Are there other aspects of how copyright law applies to AI-generated outputs that need to be clarified or strengthened before the best way to implement related transparency measures will become clear? (e.g. regarding the subsistence and ownership of copyright in certain AI-generated content)?**

Clarifications to the Copyright Act for AI-Generated Works

As AI continues to influence the creation of content, there is an urgent need to clarify certain aspects of copyright law to ensure legal certainty and protection for both creators and rights holders. Australian copyright law requires a work to be the product of independent human intellectual effort, with that effort being directed to the creation of the material form of the work.<sup>11 12</sup> However, as AI increasingly becomes a tool for human creators, the distinction between human contributions and AI-generated content becomes blurred. Copyright law must evolve to reflect this changing landscape, ensuring that human creators retain their rights over works involving AI while providing transparency around the role AI played in the creation process.

Legal recognition of human-AI co-creation will help protect the rights of human authors while addressing the complexities of AI-assisted content. As the creative process becomes more intertwined with AI technologies, it is critical that the law adapts to clearly define the boundaries between human intellectual effort and machine-generated assistance. Without such clarification, creators may face uncertainty over ownership, leading to potential conflicts regarding the rights to the final work. This could undermine the protection that copyright law traditionally provides to human creators, while also complicating the enforcement of rights for those who rely on AI to augment their creative process.

In industries where the licensing and syndication of content are common, clear and practical guidelines are required to manage the rights associated with AI-generated works. Copyright law

---

<sup>11</sup> *IceTV Pty Limited v Nine Network Australia Pty Limited* [2009] 239 CLR 458.

<sup>12</sup> *Telstra Corp Ltd v Phone Directories Co Pty Ltd* [2010] FCAFC 149.

should address the management and ownership of these works, particularly with respect to licensing, reproduction, and distribution.

#### Protecting Original Work of Cultural Significance and Artists' Right of Action

The AIIA supports updating the Copyright Act to protect original work by First Nations people, respecting their culture and expressions of their spirituality, where relevant. We recognise that this is an emerging issue and is committed to working with the government accordingly.

#### **Conclusion**

As AI continues to reshape the landscape of content creation and dissemination, Australia faces the critical challenge of adapting its copyright laws to reflect the evolving technological environment. It is essential to strike a balance between fostering innovation in AI development and safeguarding the rights of creators and copyright holders. By implementing thoughtful, forward-looking policies, Australia can lead in promoting an inclusive and innovative AI industry, while upholding the fundamental principles of copyright law that protect creators and encourage responsible use of digital content.

Should you require further information, please contact Ms Siew Lee Seow, General Manager, Policy and Media, at [siewlee@aiia.com.au](mailto:siewlee@aiia.com.au) or 0435 620 406, or Mr David Makaryan, Advisor, Policy and Media, at [david@aiia.com.au](mailto:david@aiia.com.au).

Thank you for considering our submission.

Yours sincerely  
Simon Bush  
CEO, AIIA

#### **About the AIIA**

The AIIA is Australia's peak representative body and advocacy group for those in the digital ecosystem. Since 1978, the AIIA has pursued activities to stimulate and grow the digital ecosystem, to create a favourable business environment for our members and to contribute to Australia's economic prosperity. We are a not-for-profit organisation to benefit members, which represents around 90% of the over one million employed in the technology sector in Australia. We are unique in that we represent the diversity of the technology ecosystem from small and medium businesses, start-ups, universities, and digital incubators through to large Australian companies, multinational software and hardware companies, data centres, telecommunications companies and technology consulting companies