



Productivity Commission
Level 8, Two Melbourne Quarter
697 Collins Street
Docklands Vic 3008, Australia

September 12, 2025

Re: News/Media Alliance Comments Regarding the Productivity Commission’s Interim Report on Harnessing Data and Digital Technology.

The News/Media Alliance (“N/MA”) welcomes the opportunity to provide these comments to the Productivity Commission (the “Commission”) in response to the Commission’s request for comments regarding its Interim Report on Harnessing Data and Digital Technology (the “Interim Report”), published on August 5, 2025.

N/MA is a nonprofit organization headquartered in Washington, D.C., United States, representing the newspaper, magazine, and digital media industries. N/MA represents over 2,200 diverse publishers in the United States and internationally, ranging from the largest news and magazine publishers to hyperlocal newspapers, and from digital-only outlets to papers who have printed news for centuries. In total, the Alliance’s membership accounts for nearly 90 percent of the daily newspaper circulation in the United States, over 500 individual magazine brands, and dozens of digital-only properties. While many of N/MA’s members are based in the United States, and our focus is the U.S. market, we also represent publishers based in and with significant operations and readership in Australia.

Our comment seeks to provide the Commission with information about artificial intelligence (AI) and the U.S. copyright system and correct misconceptions that a text and data mining (TDM) exception would enhance Australia’s creative and innovation ecosystem.

- ***U.S. copyright law is fundamentally based on an opt-in framework and does not provide for a blanket exception for AI-related copying.*** Fair use determinations are subject to case-by-case analyses, with the boundaries to be determined by the courts over the coming years. This careful judicial rubric supports marketplace developments and reflects the importance of protecting intellectual property. Creative and technology sectors flourish because of, and not in spite of, this fact-based approach.
- Most recently, the U.S. Copyright Office’s report on AI and copyright highlighted that the law around AI is context-dependent and still developing, and ***many uses that may cause significant market harm are likely not to be fair use under U.S. copyright law:***

- “Various uses of copyrighted works in AI training are likely to be transformative. The extent to which they are fair, however, will depend on what works were used, from what source, for what purpose, and with what controls on the outputs—all of which can affect the market. ... making commercial use of vast troves of copyrighted works to produce expressive content that competes with them in existing markets, especially where this is accomplished through illegal access, goes beyond established fair use boundaries.”¹
- “The copying involved in AI training threatens significant potential harm to the market for or value of copyrighted works. Where a model can produce substantially similar outputs that directly substitute for works in the training data, it can lead to lost sales. Even where a model’s outputs are not substantially similar to any specific copyrighted work, they can dilute the market for works similar to those found in its training data, including by generating material stylistically similar to those works.”²
- “The use of RAG is less likely to be transformative where the purpose is to generate outputs that summarize or provide abridged versions of retrieved copyrighted works, such as news articles, as opposed to hyperlinks.”³
- There are currently dozens of lawsuits challenging AI developers’ various uses of copyrighted content. The early ***decisions that have been issued do not resolve all of the disputes around unauthorized use of copyrighted material:***
 - *Kadrey v. Meta*: “In cases involving uses like Meta’s, it seems like the plaintiffs will often win, at least where those cases have better-developed records on the market effects of the defendant’s use. ... And some cases might present even stronger arguments against fair use. For instance, as discussed above, it seems that markets for certain types of works (like news articles) might be even more vulnerable to indirect competition from AI outputs. ”⁴
 - *Bartz v. Anthropic*: “This order doubts that any accused infringer could ever meet its burden of explaining why downloading source copies from pirate sites that it could have purchased or otherwise accessed lawfully was itself reasonably necessary to any subsequent fair use.”⁵
 - *Thomson Reuters v. Ross*: “The public has no right to Thomson Reuters’s parsing of the law. Copyrights encourage people to develop things that help society, like

¹ U.S. COPYRIGHT OFFICE, COPYRIGHT AND ARTIFICIAL INTELLIGENCE PART 3: GENERATIVE AI TRAINING: PRE-PUBLICATION VERSION at 107 (May 6, 2025), <https://www.copyright.gov/ai/Copyright-and-Artificial-Intelligence-Part-3-Generative-AI-Training-Report-Pre-Publication-Version.pdf>.

² *Id.* at 73.

³ *Id.* at 47.

⁴ *Kadrey v. Meta Platforms, Inc.*, 3:23-cv-03417-VC, slip op. at 39 (N.D.Cal. Jun. 25, 2025).

⁵ *Bartz v. Anthropic*, 3:24-cv-05417-WHA, slip op. at 18 (N.D.Cal. Jun. 23, 2025).

good legal-research tools. Their builders earn the right to be paid accordingly. ... There is nothing that Thomson Reuters created that Ross could not have created for itself or hired LegalEase to create for it without infringing Thomson Reuters's copyrights.”⁶

- Considering that many AI-related legal questions are still being adjudicated, the introduction of a broad ***TDM exception would move Australia further away from the United States***, which leads the world in innovation and productivity – including due to non-copyright factors including private markets, computer chips, energy and infrastructure policies – while retaining a balanced copyright system. The experiences of other countries suggest that a broad TDM exception would fail to spur local technology development and reduce creative and journalistic incentives. A broad TDM exception could threaten Australian creative industries, while ironically entrenching the dominance of global Big Tech firms.

This consultation comes at an important juncture, with burgeoning marketplace licensing deals and ongoing international policy discussions around how to incentivize innovation in AI while safeguarding rightsholder interests. N/MA and our members support responsible AI development, which respects publishers' intellectual property rights and takes a balanced policy approach to AI innovation and regulation. However, a new fair dealing exception for TDM would seriously undermine rightsholders' ability to control their works and turn the fundamentals of copyright upside down, violate international agreements – including the Berne Convention and the Agreement on Trade-Related Aspects of Intellectual Property Rights (TRIPS)⁷ – while also threatening Australia's thriving digital and creative economy. We strongly urge the Committee to reject such proposals outright.

1. The Commission Should Reject Any Proposals for a Text and Data Mining Exception.

High-quality, trustworthy newspapers, magazines, and digital media play an important role in the communities they serve by fostering an informed public and the public trust necessary for democracy. However, the massive proliferation of AI models, applications, and developers in the last few years presents new potential benefits and substantial risks to publishers. By producing substitute products and services that redirect audiences away from advertising and subscription-supported content – without permission or compensation – AI companies and platforms usurp value from the efforts of creators or publishers. For example, a recent study by TollBit estimated that AI search engines deliver 91 percent and chatbots 95.7 percent fewer referrals to news websites than traditional search engines.⁸ And this trend is only getting worse

⁶ *Thomson Reuters v. Ross*, No. 20-cv-613-SB, slip op. at 23 (D. Del. Feb. 11, 2025).

⁷ Article 5(2) of the Berne Convention explicitly states that the “enjoyment and the exercise of these rights shall not be subject to any formality,” which combined with the limitation of article 9(2) that any exceptions provided in “certain special cases” must “not conflict with a normal exploitation of the work and does not unreasonably prejudice the legitimate interests of the author,” renders opt-out measures, and TDM exceptions in general, problematic as a matter of international law.

⁸ TollBit, *AI Scraping Is on the Rise: TollBit AI User Agent Index - Q4 2024* (Feb. 24, 2025), <https://tollbit.com/bots/24q4/>.

– until recently, every two scrapes by Google led to one visitor to the original publisher’s website, but today that ratio is 18:1, while OpenAI’s crawl to crawler-to-visitor ratio has increased from 250:1 to 1,500:1 and Anthropic’s from 6,000:1 to 60,000:1 in just six months.⁹ Instead of readers, publishers’ websites are now full of crawlers that bring no benefits to publishers. These trends greatly undermine publishers’ ability to monetize their content through subscriptions, advertising, licensing, and other means.

The suggestion that Australia needs a TDM exception is based on a false premise. As the Interim Report notes, the Australian Law Reform Commission recommended an amendment to enable TDM due to it being a “non-expressive” use, not protected by copyright. While this may be the case for some limited use cases in non-generative contexts, N/MA has previously laid out in detail why and how generate AI developers use and copy protected publisher content specifically for its expressive qualities.¹⁰ As N/MA’s White Paper on AI and copyright, published in October 2023 and submitted to the U.S. Copyright Office to inform its study into the topic, noted, AI systems “rely on the precise grammar and word selection of original texts to best mimic the ingested materials. Thus, GAI developers use the expression from the underlying work to ensure that the LLMs better interpret queries, carry out searches, deliver responsive content, and even write articles.”¹¹

The U.S. Copyright Office later endorsed this view, noting in its report on Copyright and Artificial Intelligence that the argument that “the use of copyrighted works to train AI models is inherently transformative because it is not for expressive purposes. We view this argument as mistaken. Language models are trained on examples that are hundreds of thousands of tokens in length, absorbing not just the meaning and parts of speech of words, but how they are selected and arranged at the sentence, paragraph, and document level—the essence of linguistic expression. ... Where the resulting model is used to generate expressive content, or potentially reproduce copyrighted expression, the training use cannot be fairly characterized as ‘non-expressive.’”¹²

Instead of imposing rigid rules, the U.S. Copyright Act is complemented by centuries of evolving judge-made law to provide guidelines and certainty around the bounds of copyright protection, based on an analysis of four fair use factors, including whether the copying will hurt the market for the copied works. The analysis does not categorically excuse otherwise infringing uses of copyrighted works. This flexible approach, while not without its own challenges, preserves the fundamental presumptive right of rightsholders to control their works, and has a track record of

⁹ Rob Thubron, *Cloudflare Tests “Pay-for-Crawl” System to Charge AI Firms for Scraping Website Content*, TECHSPOT (Jul. 1, 2025), <https://www.techspot.com/news/108521-cloudflare-tests-pay-crawl-system-charges-ai-firms.html>; Ethan Hays (@ethanhays), X (Jun. 27, 2025, 6:32 PM), <https://x.com/ethanhays/status/1938651733976310151>.

¹⁰ NEWS/MEDIA ALLIANCE, WHITE PAPER: HOW THE PERVASIVE COPYING OF EXPRESSIVE WORKS TO TRAIN AND FUEL GENERATIVE ARTIFICIAL INTELLIGENCE SYSTEMS IS COPYRIGHT INFRINGEMENT AND NOT A FAIR USE (Oct. 31, 2023), <https://www.newsmediaalliance.org/wp-content/uploads/2025/02/AI-White-Paper-with-Technical-Analysis.pdf>.

¹¹ *Id.* at 21.

¹² U.S. COPYRIGHT OFFICE, COPYRIGHT AND ARTIFICIAL INTELLIGENCE PART 3: GENERATIVE AI TRAINING: PRE-PUBLICATION VERSION at 47-8 (May 6, 2025), <https://www.copyright.gov/ai/Copyright-and-Artificial-Intelligence-Part-3-Generative-AI-Training-Report-Pre-Publication-Version.pdf>

accommodating technological expansion. As the U.S. Copyright Office notes, “[t]hroughout its history, copyright law has adapted to new technology, furthering its progress while preserving incentives for creative activity. This has enabled our nation’s creative and technology industries to become global leaders in their fields. While the use of copyrighted works to power current generative AI systems may be unprecedented in scope and scale, the existing legal framework can address it as in prior technological revolutions. The fair use doctrine in particular has served to flexibly accommodate such change. We believe it can do so here as well.”¹³

There is no established precedent in the United States that generative AI uses – training or real-time retrieval – are fair use, and the decisions in early cases have been mixed and sometimes conflicting. In February 2025, in a case addressing non-generative AI, *Thomson Reuters v. Ross*, the court ruled against the fair use defense,¹⁴ finding Ross’s use of Westlaw’s headnotes was commercial and not transformative under the first fair use factor and that the use harmed the potential market for AI training under factor four. More recently, the court in *Bartz v. Anthropic* – brought by a group of literary authors for the unauthorized use of their works for AI training – issued a decision on June 24, 2025, regarding fair use as applied to generative AI model training, digitization of print books, and the building of a library built on pirated copies of the authors’ works.¹⁵

While the *Bartz* court found for the defendant when it came to training and digitization – arguably failing to properly analyze the fourth fair use factor (market harm) and misinterpreting relevant case law, both issues likely to be corrected on appeal – it refused to dismiss the library claim involving the copying of pirated works, some of which were subsequently used for AI training. Following the decision, in a significant win for the authors, it was announced in August that Anthropic had decided to settle the suit for a proposed USD \$1.5 billion, one of the highest payouts in the history of copyright.¹⁶ The decision and the settlement sets significant precedent for AI companies who may be liable for billions in damages for creating datasets involving works copied from unauthorized sources.

Shortly following the decision in *Bartz*, the court in *Kadrey v. Meta* issued a summary judgement finding that Meta’s use of the plaintiffs’ books for AI training was “highly transformative” and amounted to fair use.¹⁷ The court emphasized that the holding was narrow due to lack of adequate evidence in this particular case, and noted that the analysis may be substantially different when it comes to the use of other works for similar AI training purposes. The court also engaged in a careful analysis of the substitutive effect of such uses, noting that “this effect also seems likely to be more pronounced with respect to certain types of works ... An LLM that could generate accurate information about current events might be expected to

¹³ U.S. COPYRIGHT OFFICE, COPYRIGHT AND ARTIFICIAL INTELLIGENCE PART 3: GENERATIVE AI TRAINING: PRE-PUBLICATION VERSION at 107 (May 6, 2025), <https://www.copyright.gov/ai/Copyright-and-Artificial-Intelligence-Part-3-Generative-AI-Training-Report-Pre-Publication-Version.pdf>.

¹⁴ *Thomson Reuters v. Ross*, No. 20-cv-613-SB (D. Del. Feb. 11, 2025).

¹⁵ *Bartz v. Anthropic*, 3:24-cv-05417-WHA, slip op. at 18 (N.D.Cal. Jun. 23, 2025).

¹⁶ *Copyright Alliance CEO Issues Statement on ‘Bartz v. Anthropic’ Case Settlement*, COPYRIGHT ALLIANCE (Sep. 8, 2025), <https://copyrightalliance.org/press-releases/bartz-anthropic-case-settlement/>.

¹⁷ *Kadrey v. Meta Platforms, Inc.*, 3:23-cv-03417-VC, slip op. at 39 (N.D.Cal. Jun. 25, 2025).

greatly harm the print news market. ... [This case] involves a technology that can generate literally millions of secondary works, with a miniscule fraction of the time and creativity used to create the original works it was trained on. No other use – whether it’s the creation of a single secondary work or the creation of other digital tools – has anything near the potential to flood the market with competing works the way that LLM training does. And so the concept of market dilution becomes highly relevant.”¹⁸

In addition, broad TDM exceptions also grapple with insufficient safeguards and definitional inadequacies that render them problematic to rightsholders. Broad exceptions with opt-out provisions are unworkable as there are currently no sufficient technical rights reservation systems for rightsholders to meaningfully exercise their rights. Requiring publishers to implement imperfect and costly technical compliance systems is an unreasonable ask, while blocking all scraping is neither desirable nor practical in many instances. Meanwhile, more narrow TDM exceptions, often for non-commercial research purposes – while sometimes warranted in limited cases – present potential pitfalls if the line between commercial and non-commercial uses is not finely delineated with particular attention paid to limitations on downstream uses.

There are also concerns that a TDM regime with opt-out would be abused by dominant players in the search and emerging “answer engine” marketplace, which condition permission for indexing content on also allowing (either explicitly or in practice) the use of content for generating AI summaries. This practice has been well noted by players such as Google, whose AI Overviews and AI Mode products are tied to its indexing of content for search generally, meaning that publishers seeking to have their content discoverable on Google search, are not able to effectively opt-out of their content being exploited for AI purposes.

The proposal seems mainly to be justified by one misinformed consideration, namely that a TDM exception would enable “smaller, low compute models (such as task-specific models)” to be built and trained in Australia. Rather than facilitating small AI developers to a significant extent, a TDM exception would have the potential to seriously undermine and threaten the viability of Australian publishers and the communities they serve. Australia’s existing legal framework already supports a marketplace licensing that encourages innovation while also balancing important cultural and creative values. But licensing marketplaces can thrive only when all parties have adequate bargaining information and incentives to negotiate – something that a broad TDM exception would remove.

There is ample evidence that the licensing system is rapidly adapting for AI use purposes. The past two years have witnessed the completion of over 140 AI licensing deals announced between AI developers and all sectors of the creative industry, including in Australia. These deals evidence both the value of high-quality, vetted content for AI models and that developing innovative, trustworthy, and ethical AI is both possible and profitable. In fact, often it is smaller

¹⁸ *Id.* at 32.

actors who develop innovative business models that are based on ethical licensing.¹⁹ The fact that “large AI models are already being trained on unlicensed copyrighted materials” ignores the multiple unresolved lawsuits challenging such uses and does not support the adoption of a TDM exception – rather, it highlights the need to avoid further entrenching such practices by adopting one.²⁰

There is also no reason to believe that adopting a TDM exception would in any way incentivize domestic AI industries or lead to a renaissance of AI development, as evidenced by the seeming lack of thriving AI industries or international investment in countries such as Japan and Singapore that have adopted extremely broad TDM exceptions with limited safeguards and no ability to opt-out, while the European Union’s TDM exception has also seemingly not spurred AI innovation noticeably. Meanwhile, AI developers thrive in countries such as the United States that have no TDM exceptions. The American success is due to a mix of policy, private capital, and workforce factors that create the foundation for a vibrant AI economy.²¹ Australia should therefore retain its existing legal framework that currently balances the interests of all parties.

2. The Competition Commission Should Encourage Voluntary Collective Licensing.

Rather than an overly burdensome and fundamentally unfair TDM exception, the Commission should recommend policies to facilitate healthy markets through consensus between rightsholders, developers, and the Government around protection and partnerships. Rightsholders are by and large not against responsible AI development – indeed, many of N/MA’s members use AI in their daily operations in one way or another – and there is a potential for a symbiotic, mutually beneficial relationship between AI developers and rightsholders. Such a relationship, built on respect for intellectual property, would provide developers with legal certainty and access to the high-quality content needed to keep their systems reliable and rightsholders with licensing revenue that supports the development of high-quality content.

Voluntary licensing, of which based on N/MA’s calculations there are over 140 examples just in the AI context, has always been the cornerstone of copyright policy, and these marketplace

¹⁹ See, e.g., BRIA, <https://bria.ai/> (“Visual Gen AI Platform for Developers, Trained with 100% Licensed Data”); ProRata, <https://prorata.ai/> (“Our goal is nothing less than making sure that there is a thriving ecosystem of creators, businesses, and consumers that benefit from the emerging world of zero-click search.”).

²⁰ PRODUCTIVITY COMMISSION, INTERIM REPORT: HARNESSING DATA AND DIGITAL TECHNOLOGY at 28 (Aug. 5, 2025), <https://www.pc.gov.au/inquiries/current/data-digital/interim/data-digital-interim.pdf>.

²¹ See, e.g., Brad Smith, *The Golden Opportunity for American AI*, MICROSOFT ON THE ISSUES (Jan. 3, 2025), <https://blogs.microsoft.com/on-the-issues/2025/01/03/the-golden-opportunity-for-american-ai> (“America’s technological strength has always been rooted in the private sector. Today, the United States leads the global AI race thanks to the investment of private capital and innovations by American companies of all sizes, from dynamic start-ups to well-established enterprises.”); Sandy Carter, *5 Top Reasons Why U.S. And EU Could Spearhead AI Development*, FORBES (Feb. 16, 2024), <https://www.forbes.com/sites/digital-assets/2024/02/16/will-us-innovation-or-eu-regulation-win-the-race-to-dominate-ai/> (“Microchips aren’t the only thing that puts the US in the lead. According to a research report from Macro Polo, 60% of top AI researchers are American-based, with the remaining 40% spread around jurisdictions worldwide. The fact is there simply isn’t nearly as much talent concentrated anywhere on Earth as there is in the US.”).

solutions – including voluntary collective licensing where appropriate – should be encouraged. This is in addition to numerous services providing dynamic licensing solutions facilitating easy and large-scale content licensing for AI use purposes, such as TollBit, Cloudflare, Created by Humans, CCC, and others.²² Marketplace licensing solutions are particularly important as AI technologies are so new, the use cases so unpredictable, and the economics so unknown that voluntary licensing is needed to provide the necessary flexibility and leverage to respond to new challenges, business developments, and evolving technologies. TDM exceptions would significantly threaten these free-market approaches, and we therefore urge the Commission to propose measures to facilitate marketplace licensing to the fullest extent possible, including voluntary third-party aggregation and brokering services.

3. Conclusion.

Publishers rely on copyright, their ability to control the uses of their works, and to enforce their rights against those who misappropriate their content without licensing or compensation. TDM exceptions for commercial uses reverse the basis of the copyright system that has served all stakeholders well for centuries. While limited TDM exceptions for non-commercial research may be appropriate in some circumstances, even these need to be carefully and narrowly drafted to prevent misuses and circumvention. We strongly urge the Commission to reject any suggestion that a TDM exception is warranted in Australia or, at the very least, limit such an exception to non-commercial research uses with meaningful safeguards.

Thank you for the opportunity to provide these comments.

²² See, e.g., TollBit, <https://tollbit.com/>; Will Allen et al., *The Next Step for Content Creators in Working with AI Bots: Introducing AI Crawl Control*, CLOUDFLARE BLOG (Aug. 28, 2025), <https://blog.cloudflare.com/vi-vn/introducing-ai-crawl-control/>; Created by Humans, <https://www.createdbyhumans.ai/>; CCC Announces AI Systems Training License for the External Use of Copyrighted Works Coming Soon, COPYRIGHT CLEARANCE CENTER (Mar. 4, 2025), <https://www.copyright.com/media-press-releases/ccc-announces-ai-systems-training-license-for-the-external-use-of-copyrighted-works-coming-soon/>.