

SUBMISSION on the Interim Report of the “Harnessing data and digital technology” inquiry: Promoting microeconomic reform for a sustainable AI ecosystem

12 September 2025



DAVID KENNEDY

MANAGING DIRECTOR, VENTURE INSIGHTS

david.kennedy@ventureinsights.com.au

Introduction.....	2
Executive Summary.....	2
The impracticability of a TDM exception.....	3
Rights unenforceable without AI operator transparency.....	3
A history of resistance by AI operators.....	3
Lessons from the Media Bargaining Code.....	3
Opt-out is no easy solution.....	4
The undesirability of a TDM exception.....	4
International examples need careful interpretation.....	4
Private gain is not the same as public interest.....	6
“Non-expressive use” claim rests on an irrelevant distinction.....	7
Risks of market disruption.....	9
Supporting market development is the answer.....	10
Generative AI companies need to step up.....	10
Role of collecting societies will grow.....	11
Initial steps.....	11
Next steps.....	12
Other players may have a role.....	13

Introduction

Venture Insights is a research and consulting firm based in Melbourne, producing high-quality, locally relevant research, forecasting, and consulting for the Technology, Media, and Telecommunications (TMT) sectors in Australia and New Zealand:

- Delivering insights into TMT sector dynamics, emerging trends, and competitive landscapes to inform strategic decision-making.
- Maintaining robust, data-driven 5-10 year market forecasts of TMT markets based on consumer and enterprise surveys and leverage unique data assets.
- Conducting bespoke consulting, briefings, workshops, and direct strategic counsel to investment committees and executive teams, facilitating informed leadership.

Executive Summary

The Productivity Commission's Interim Report on "Harnessing data and digital technology" floated a proposal to create a Text and Data Mining (TDM) exception to copyright law to fuel AI development.

This submission specifically addresses this TDM exception proposal. We argue two separate points:

- That implementation of a TDM exception is impracticable.
- That implementation of a TDM exception is undesirable.

While framed as a move to foster innovation, this proposal is misguided and inequitable. It overlooks the established principles of copyright, the demonstrated behaviour of AI developers, and the practical issues of enforcement.

More fundamentally, a TDM exception is a solution to the wrong problem. The problem is not simply that AI platforms need more access to copyright resources, and therefore should be granted this access. The real problem is how to develop institutional and technological arrangements that support a long-term market solution to the AI platforms' need for training content, but in a way that maintains incentives for investment in creative content. In short, the real issue is one of microeconomic reform.

A TDM exception does not meet this challenge. It is a short-term "sugar hit" for the AI industry that is self-defeating because it will undermine the development of these market solutions, leading to creative underproduction. This would result in a suboptimal long term outcome for AI platforms, content creators and the wider economy alike.

The key takeaway from this submission is that the focus of any recommendations from the Inquiry should be for microeconomic reform of copyright markets. Where appropriate, this should address any new technical and institutional arrangements needed to support content provider participation in the digital economy. These solutions are still emergent, and their development could easily be disrupted if property rights in these markets are not maintained.

Any policy solution that is both practicable and principled needs to build on the existing licensing structures and the markets they currently sustain, not undermine them. What is required at this time

is not short-term thinking, but a steady policy framework and a commitment to positive, incremental change that is respectful of content holders' rights.

The impracticability of a TDM exception

Rights unenforceable without AI operator transparency

We recognise that the Commission has proposed limitations on any TDM exception. But the most critical practical flaw in the TDM exception proposal ("The TDM proposal") is that any limits placed on an exception would be unenforceable. How could a creator or publisher prove their work was used improperly if the AI model's training dataset is a secret?

AI companies aggressively protect their training data, labelling it a trade secret. Without transparency, rights holders have no way to audit a model to see if its training complied with the law. An exception that cannot be policed is not a limited exception at all; it is a blank cheque. This opacity leaves creators with theoretical rights but no practical way to enforce them.

This is no hypothetical. In ongoing litigation, AI companies have consistently refused to disclose the full contents of their training datasets. OpenAI, for example, has argued in court filings that its training data and methods are proprietary trade secrets. This resistance to transparency is a core part of their business and legal strategy, making any regulatory framework that relies on it inherently unworkable.

A history of resistance by AI operators

Further, there is no reason to believe that global AI platforms would respect such limitations. Their history is not one of compliance, but of seeking forgiveness after the fact. Major AI developers are currently embroiled in lawsuits alleging unauthorised use of copyright material to build the models they now seek to legitimise:

- **The New York Times vs. OpenAI & Microsoft (2023):** This landmark lawsuit alleges that millions of articles were illegally used to train ChatGPT, with the model sometimes reproducing content verbatim.
- **Getty Images vs. Stability AI (2023):** Getty alleges that the image generator Stable Diffusion was trained by scraping over 12 million of its images without permission, even citing examples where remnants of the Getty Images watermark are visible in AI-generated images.
- **Authors Guild vs. OpenAI:** Numerous authors, including Sarah Silverman and George R.R. Martin, have joined class-action lawsuits alleging their books were used to train large language models without consent, credit, or compensation.

This pattern of behaviour shows a disregard for the rights of creators, suggesting that any new "rules" would likely be treated as mere obstacles to be navigated or ignored.

Lessons from the Media Bargaining Code

We need only look to the recent past to see how global tech giants react to Australian laws that attempt to value local content. The introduction of the News Media Bargaining Code (NMBC) was met with hostility.

In 2021, in response to the NMBC, Google threatened to withdraw its search engine from Australia. In a more drastic move, Facebook (Meta) temporarily blocked all Australian news content from its platform, a clear act of political and economic pressure to undermine the legislation.

This experience demonstrates a clear playbook: platform companies leverage their market power to resist any framework that requires them to negotiate fairly or compensate Australian rights holders for the value they derive from their content.

A TDM exception would simply be the next battleground, where they would likely fight any attached licensing schemes or creator compensation models.

Opt-out is no easy solution

One proposal to protect licence holders' rights is to give them an "opt-out" from a TDM exception. Even if a TDM exception could overcome the transparency and compliance issues, opt-out proposals also face numerous practical challenges. These include:

- **Technological Nuances:** Watermarks or hashtags like "#NoAITraining" might not be legally effective or machine-readable. "Machine-readable means" (e.g., specific metadata, robots.txt files) are necessary, which are technologically complex for individual creators. Collecting societies may have an important role here.
- **Platform Control:** Web-crawler blockers are bot-specific, page-specific, and not work-specific, meaning they can be bypassed or require constant updates. Metadata can be easily altered, removed, or stripped by other platforms like social media sites.
- **Downstream Use:** Opt-out notices might be ineffective for "downstream" re-publications or derivative works, meaning creators could lose rights when their work is shared beyond their control.
- **Existing Works:** The opt-out model raises questions about whether existing online content without prior opt-out notices would become "fair game" for AI training. Once used for training, it is difficult to see how content could be opted-out retrospectively.

In summary, it is difficult to see how any TDM exception could be implemented in a way that does not significantly undermine licence holders rights, because there is no way to enforce any associated limitations.

The undesirability of a TDM exception

More fundamentally, we argue that a TDM exception, even if practicable, is not desirable.

International examples need careful interpretation

We note that the Interim Report lists several jurisdictions where TDM exceptions have been implemented. Many comparable jurisdictions, including parts of the EU, the US (through fair use), Japan, and Singapore, have some form of TDM exception or similar doctrine. The consequent claim is that failing to align with these international trends could make Australia less attractive for AI talent and investment, causing it to lag in AI take-up and potentially undermine its AI sector.

But we also note that these exceptions generally pre-date the rise of the current generation of generative AI tools, which require large amounts of data to train. The mere existence of these exceptions is not determinative, because the rise of generative AI has qualitatively changed the environment. The stakes involved, both in terms of AI benefit and the scale of copyright material required, are now much higher.

This means that the case for a TDM exception needs to be re-considered for this new environment.

The most relevant and recent example of the competing interests at stake is the UK's current process to review its existing limited TDM exception. The debate in the UK has canvassed several distinct models:

- "Do Nothing": This would maintain the current legal ambiguity, likely leading to an increase in litigation as AI developers and rightsholders test the boundaries of existing copyright law.
- Strengthening Copyright Law: This model would involve explicitly requiring licenses for most TDM activities, giving maximum control to rightsholders but potentially stifling AI innovation due to the complexity and cost of licensing at scale.
- A Broad TDM Exception (without opt-out): As initially proposed, this would have been highly favorable to AI developers but was deemed to undervalue the contribution of creators and potentially harm the UK's creative economy.
- The Opt-Out Model: The currently favored approach, which seeks to strike a balance by allowing TDM by default for lawfully accessed content while empowering rights holders to prohibit the use of their works.

The originally-favoured option of a broad TDM exception was withdrawn after significant pushback from the creative industries. The UK Intellectual Property Office's (IPO) initial 2022 economic impact assessment of a broad TDM exception was that a broad exception would provide a significant boost to AI development and innovation in the UK. While acknowledging negative impacts on creative industries, the initial assessment suggested that the overall economic benefit to the UK in terms of AI advancement would outweigh the costs, with IT-related benefits of up to £1 trillion by 2025.

The proposed benefits to the UK economy, though large, were speculative. In contrast, the creative industries were able to point to likely losses from a broad exception:

- The **Publishers Association** highlighted that it contributes **£6.7 billion** to the UK economy annually. They argued that a broad TDM exception would devalue their high-quality curated content, which is a crucial input for high-performing AI models. They commissioned their own independent reports suggesting that the loss of licensing revenue from TDM would significantly harm investment in new creative works.
- **The Associated Press (AP)** noted that it had already established licensing deals with AI developers and that a broad exception would destroy this emerging market. They stated that licensing for AI training is a vital new revenue stream that allows them to continue investing in high-quality journalism.
- **The British Photographic Council** argued that a broad exception would lead to the mass ingestion of copyrighted images without compensation, devaluing professional photography and making it harder for photographers to make a living. They pointed to the sector's significant contribution to the UK's creative economy and the risk of widespread job losses.

Apart from direct revenue losses, the creative industries were also able to highlight the emergence of new market arrangements between rights holders and the technology industry that would be disrupted by a TDM exception. The net effect of this disruption was a windfall transfer of value away from the creative industries to the technology industries, which they successfully argued was unjustifiable.

The UK is now considering a more restrictive TDM exception with an opt-out. But as we have discussed above, implementing an opt-out is not easy. The UK Government currently has no model for an opt-out scheme, and is in consultations with industry to find one.

The broad outlines of the Australian case are similar, and are developed below. No authoritative estimate of the value of the AI training data market exists in Australia. One clue is a report from Grand View Research, stating that the AI training dataset market in the healthcare sector in Australia was valued at US\$7.4 million (approximately AU\$11 million) in 2024. This niche market was projected to grow substantially, reaching an estimated US\$30.1million (approximately AU\$45 million) by 2030, demonstrating the rapid growth in datasets and their valuation. Across the economy then, we are looking at many billions of dollars.

Private gain is not the same as public interest

The strongest argument in favour of a TDM exception is that it would drive innovation and investment. Proponents of a TDM exception argue that current Australian copyright laws are too restrictive, preventing Australian companies from competing globally in AI development because they block access to the large amounts of data required for AI training. An exemption would reduce regulatory uncertainty, which can stifle innovation and investment if firms fear onerous or unclear regulations.

Separately, proponents argue that granting TDM exemptions could unlock billions of dollars of foreign investment into Australia.

But as we have already noted above, **there is little evidence to suggest that global AI companies have been at all restricted in their ability to (whether legally or illegally) access copyright material.** AI agents are already being trained on massive datasets. The issue, if it exists at all, is amongst smaller AI companies.

Looking at domestic AI development specifically, a TDM exception could be beneficial for smaller, low-compute models built and trained domestically by Australian research institutions and medical technology firms, fostering local innovation.

But we disagree that a TDM exception is the right way to achieve this access. Copyright exceptions in Australia, under the principle of fair dealing, have always been calibrated to serve a clear and direct public benefit. These exceptions allow for the use of copyrighted material for purposes like research, education, criticism, review, parody, and news reporting.

The common thread is that they facilitate public discourse, learning, and accountability - not direct commercial product development. The existing fair dealing provisions in the *Copyright Act 1968* (specifically sections 40-43 and 103A-103C) are narrowly defined for specific, non-commercial, or transformative public interest purposes.

The proposed TDM exception is inconsistent with the fundamental principle of fair dealing. Its primary beneficiaries would not be students, Australian IT researchers, or the public, but commercial

AI labs like OpenAI, Google, and Microsoft, entities with collective market capitalisations in the trillions.

The argument that their commercial success will eventually trickle down as a "public benefit" is speculative and unsupported. Even if it were true, it cannot by itself justify the expropriation of private property rights. Weakening property rights always benefits someone - the issue is whether that weakening can be justified. In this case, the direct benefit is private profit, making this a departure from the principles underpinning Australian copyright law. A TDM exception for building a commercial product does not align with the established purpose of fair dealing exceptions.

Arguably, the reasonable needs of local not-for-profit AI researchers could be met with modest and narrow re-interpretation of the current fair dealing provisions relating to research. And since these researchers are subject to local law, there is a reasonable expectation they would respect these limits.

The needs of commercial AI developers are a separate case. However, **the answer is not a race to the bottom where local AI developers are given a free hand to ape the excesses of their global counterparts.**

“Non-expressive use” claim rests on an irrelevant distinction

Another argument for TDM exceptions is that the use of copyrighted material for AI training is "non-expressive." Copyright typically protects the expression of ideas, not the underlying information or data itself. From this perspective, using content to identify patterns for AI training should not constitute infringement.

The distinction between expressive and non-expressive use has gained importance with the rise of AI technologies and big data. There is a recognised distinction between expressive and non-expressive use of copyright material. Key characteristics of expressive use include:

- **Focus on Creativity:** The use is concerned with the artistic, literary, or musical qualities of the work.
- **Communicating the Author's Expression:** The user is often communicating the author's original message, style, or aesthetic to an audience.
- **Substitution:** The use could serve as a market substitute for the original work, as it offers the same creative experience.

Examples of expressive use would include:

- Copying and pasting a chapter from a novel into another book.
- Including a scene from a movie in a compilation video without permission.
- Creating a t-shirt that features a direct copy of a popular image.
- Playing a protected song in its entirety in a commercial or a film.

In contrast, non-expressive use is a use of a copyrighted work that does not engage with the creative expression intended for human consumption. Instead, it utilises the work for purposes that are purely functional, analytical, or informational, where the creative aspect is incidental. Key characteristics of non-expressive use include:

- **Focus on Information:** The use is aimed at extracting underlying data, facts, or patterns from the work, rather than appreciating its creative expression.
- **No Communication of Expression:** The user is not communicating the author's original creative message to an audience. The final output is often a new set of data or a computational analysis.
- **Not a Market Substitute:** The use does not replace the original work in the marketplace. No one would analyze a dataset of novels to gain the same experience as reading one of them.

Examples of non-expressive use include:

- A researcher using software to scan thousands of copyrighted books to analyse linguistic patterns, the frequency of certain words, or the evolution of sentence structure over time. The software "reads" the books, but no human consumes the creative expression of any single book. The output is a statistical analysis, not a story.
- A search engine creating copies of web pages to index them. This copying is necessary to allow users to find information, but its purpose is functional (to build an index), not to present the creative content of the websites to users as its own.
- Software like Turnitin making copies of student essays and scholarly articles to compare them for textual similarities. The purpose is to check for originality, not to enjoy the literary quality of the works being copied.

The difference between expressive and non-expressive use is important in the context of fair use (in the United States) and fair dealing (in jurisdictions like the UK, Canada, and Australia).

A use that is non-expressive is more likely to be considered fair dealing. This is because it does not harm the commercial market for the original work and is often "transformative" - that is, it uses the original work for a completely new and different purpose with different form.

The problem with TDM proponents using this argument is that it rests on a distinction between AI training and AI inference that is irrelevant in practice. It is true that AI training "extracts" information from copyright material, and does not retain a copy of that material. In that sense, the use is non-expressive.

However, once that model is trained, it is used for AI inference to produce text, audio and/or visual outputs. These outputs are of the same kind as the original inputs, and they do compete directly in the same market as the original creators. That is the entire purpose of training a generative model in the first place.

So the practical result of the training is indeed something that looks very much like expressive use. This becomes obvious when one considers that it is even possible to ask some AI applications to produce art in the style of a particular artist, something that is plainly communicating the author's expression, and is also a market substitute for the original inputs.

The claim that AI training is non-expressive use is therefore on shaky ground. It might hold for a research model that is never used to produce commercial outputs. But if the whole purpose of AI training is to create a model that will generate new material that can substitute for original creations, then the argument plainly fails.

As a side-note, this shows that traditional copyright concepts need to be applied carefully in the new environment generative AI has created.

Risks of market disruption

Beyond undermining the fair dealing principle, **the erosion of intellectual property rights injects risk into efforts to develop well-functioning markets for copyright material.** We have already mentioned the examples raised in the content of the UK's review of its TDM exception. This will undermine business models and result in a lower level of production that benefits no-one.

A historical example is the emergence of illegal online music sites in the late 1990s like Napster. The most immediate and profound impact was the widespread, and illegal, free exchange of digital music files. This unprecedented access to an unlicensed library of songs without cost led to a sharp downturn in physical music sales, particularly CDs, which had been the industry's primary revenue source.

At its peak in 1999, the global recorded music industry had revenues of approximately US\$28.6 billion (equivalent to over US\$45 billion in 2023 dollars). By 2014, the industry's global revenue had cratered to a low of US\$14.3 billion, less than half of its 1999 peak.

This had a profound effect on the industry's ability to invest in new talent, distribution and marketing, and industry skills:

- **Artist & Repertoire (A&R)** budgets were slashed. This was often the first area to see cuts, leading to fewer new artists being signed. Labels could no longer afford to gamble on developing acts, and focussed on artists who already had a proven track record, a significant online following, or a sound that fit a commercially successful formula. This led to a decline in innovation.
- **Reduced marketing budgets:** Music video productions, magazine spreads, and promotional tours were significantly curtailed for all but the biggest superstars. As physical distribution declined, record stores closed and physical distribution infrastructure became redundant. This led to significant job losses.
- **Skilled workforce decline:** Major labels merged and laid off thousands of employees worldwide, from A&R scouts and marketers to administrative staff. This resulted in a significant "brain drain" and the loss of institutional knowledge and mentorship opportunities. With smaller recording budgets, many large, iconic recording studios closed. The industry shifted towards smaller project studios and home recording, reducing opportunities for sound engineers, producers, and technicians to learn their trade in a traditional apprenticeship model. Many roles that were once in-house, such as producers, mixers, and even marketing specialists, became freelance positions. This created a more precarious "gig economy" within the industry.

Unchecked, this would have led to a permanently lower level of music production globally. This did not happen for two reasons:

- First, the industry struck back with **vigorous copyright enforcement actions**, asserting their intellectual property rights against those who argued that "copyright is dead". Napster was shut down in 2001, and its successors subsequently.

- Second, they **worked with new partners** on the creation of legal digital download services like Apple's iTunes. In the longer term, Napster's legacy is most evident in the rise of streaming services such as Spotify, Apple Music, and others.

As a result, the music industry was able (slowly) to recover and sustain a reasonable level of investment.

This example is instructive because it shows that building a new ecosystem to meet unmet demand delivered better long-term outcomes than the short-term "sugar hit" of free content. If control of intellectual assets had not been recovered, the result would have been a permanently lower level of output and activity.

In the context of AI, the risk of copyright infringement is that incentives to invest in content production will be undermined. This will especially be the case if copyright inputs are used to train AI models that can produce new content in competition with human artists.

But this production would be based on a stagnating stock of inputs, as human artists reduced output in response to declining returns on their effort. The result would be diminished creative industries and a low-quality market of "slop" that would undermine the benefits of AI technology.

A TDM exception would not reduce this risk; it would increase it because it would weaken creative investment incentives.

Supporting market development is the answer

Rather than the "sugar hit" of a TDM exception, **long-term development of generative AI is best served by the development of a value chain and ecosystem that preserves investment incentives for all participants.**

This will require adaptation by all participants in that ecosystem as well. This adaptation is another example of "microeconomic reform" designed to improve market efficiencies and promote economic growth. Creating a sustainable value chain incorporating both the creative content industry and the generative AI industry will almost certainly require new technical and institutional arrangements, and will require both to do business differently.

These solutions are still emergent, and their development could easily be disrupted if property rights in these markets are not maintained. Any policy solution that is both practicable and principled needs to build on the existing licensing structures and the markets they currently sustain, not undermine them.

Generative AI companies need to step up

The first challenge for the generative AI providers is to recognise that they are indeed part of an ecosystem, not a whole one, and that the long-term health of creative content industries is in their interest.

The first initiative by generative AI providers that would serve this end would be transparency about which training inputs they have historically used.

This is not difficult for them to ascertain. There is no need to “look inside” the models to check, because these companies know exactly which inputs they have obtained and used. The only question is whether they should be compelled to reveal what they already know.

In the interests of developing a sustainable market for copyright material, they should be so compelled, because this knowledge is essential for copyright holders to evaluate their inputs to AI training, and develop appropriate pricing.

The second initiative by generative AI providers is that they should enter into the market for copyright material on the same terms as other licence seekers.

There is no need for any specific exception to cater for their needs, which are in principle no different to any other commercial player. If they are reluctant, then vigorous copyright enforcement, backed by governments, may be necessary, just as it was against pirate music providers.

While the prospect of paying for training data may increase costs, it also offers a solution to the legal and ethical ambiguities that have plagued the AI industry. A clearer licensing framework could provide AI companies with more legal certainty and a sustainable way to access high-quality training data.

This approach would benefit local AI companies, who cannot sidestep local law, in two ways:

- It would create a clear path for content acquisition that did not involve parsing the legalities of a fair dealing style TDM exception.
- It would level the playing field between local and global AI companies, who would both have access to content on the same terms.

Role of collecting societies will grow

The collecting societies have a crucial role to play in streamlining the licensing process. By representing multiple copyright holders, collecting societies can negotiate and issue licenses on their behalf, making it easier and more efficient for AI developers to obtain the necessary permissions without having to approach individual creators.

The Copyright Agency in Australia exemplifies this, stating they can assist with licensing third-party content for AI and are exploring “collective licensing solutions”. The Copyright Agency, representing authors, publishers, and visual artists, has been at the forefront of both policy advocacy and practical market development. It has been a key member of the coalition opposing any weakening of copyright law, arguing that a TDM exception would unfairly “prefer the interests of multinational technology companies”.

Initial steps

The Copyright Agency is the only Australian collecting society to have launched a specific licensing product related to AI. It has introduced an extension to its Annual Business Licence, which represents the first concrete “deal” in this space. This initiative is a highly strategic but carefully circumscribed first step:

- The extension allows staff in licensed businesses to use news media content in the prompts of generative AI tools for internal purposes.

- The license explicitly forbids the use of content for the core purpose of AI development. This includes training, fine-tuning, augmenting, or validating AI models, as well as text and data mining or compiling datasets. Furthermore, it requires businesses to ensure the content is not captured by an open AI system and used externally.

This limited licence is a pioneering move. It establishes the legal and commercial principle that using copyrighted content in conjunction with AI requires a paid license. It carves out a specific, low-risk use case, generates a new revenue stream for its members, and allows the agency to test compliance and administration mechanisms while holding a firm line against unlicensed model training. The agency is now in consultation to potentially expand this model to other content types, such as books and journals.

Next steps

Other potential benefits of collecting society management of AI content licensing include integration of safeguards in licensing agreements that address concerns about the outputs of AI models. For instance, negotiations could include provisions to prevent infringing AI-generated outputs or to manage issues like "in the style of [artist]" prompts that violate artists' moral rights, ensuring responsible AI development. The collecting agencies are well-placed to identify and manage such issues.

The approach of selling structured data sets on an open market for data will characterise the AI ecosystem in the future. This will require the collecting societies to develop these structured data sets and the associated digital marketplaces - a significant challenge. However, this is also a task for policy-makers, who will need to create regulatory frameworks and to help enforce rights to support the emergence of these markets.

The next phase for Australian collecting societies will involve the proactive construction of licensing solutions. Assuming that their members' work has value and requires a license for use in AI, they must now build the mechanisms to facilitate that licensing at scale. The path forward will likely involve:

- **Development of Collective Frameworks:** Australian societies are expected to accelerate the development of their own collective licensing schemes, looking closely to international models like the UK's CLA/ALCS framework as a template. This will require extensive consultation with members to establish opt-in mechanisms, moral rights protection, data management protocols, and equitable distribution models.
- **Incremental Market Entry:** The Copyright Agency's limited licence for AI prompts serves as a crucial pilot program. It allows the society to test the market, refine compliance procedures, and establish commercial relationships in a controlled, low-risk environment. The success and learnings from this initiative will likely inform the development of more comprehensive licenses that cover the core issue of model training.

In summary, we argue that focusing on and strengthening licensing frameworks, underpinned by transparency and the collective bargaining power of collecting societies, presents a more equitable and sustainable path forward for AI development in Australia. This approach aims to foster innovation while ensuring the continued vitality and compensation of the creative industries. In contrast, a TDM exemption which would undermine intellectual property markets and the associated industries.

Other players may have a role

There may also be roles for other players in the value chain, and we should not assume that all of these will be Australian. A recent example is Cloudflare's "Pay-Per-Crawl" initiative. Cloudflare is a major global player in internet infrastructure, so this initiative is significant.

The company is now blocking AI crawlers by default across its network and has introduced a "Pay-Per-Crawl" system, aiming to create a new economic model for the use of online content in AI training. Key features of the system are:

- **Default Blocking of AI Crawlers:** For all new domains on the Cloudflare network, and as a new default for existing customers, most known AI crawlers will be blocked from accessing content. This proactive stance aims to prevent unauthorised scraping of data for AI model training.
- **Granular Controls for Website Owners:** Publishers and website administrators now have access to more sophisticated tools to manage bot traffic. They can choose to:
 - **Allow:** Permit specific AI crawlers to access their content freely.
 - **Block:** Continue to deny access to certain or all AI bots.
 - **Rate-limit:** Control the frequency of requests from individual crawlers.
- **The "Pay-Per-Crawl" Initiative:** A new mechanism for AI companies to license and pay for the content they use. This system is designed to facilitate a direct financial relationship between content creators and AI developers. While the exact pricing and payment structures are still in their early stages and part of a private beta, the initiative aims to establish a marketplace for training data.

We do not mention this example to endorse it - content holders and AI companies need to determine themselves whether this system meets their needs. And there are other questions about whether this positions Cloudflare as a gatekeeper in the emerging marketplace, which might generate its own issues.

The point of the example is that **solutions that do not rely on any copyright exception are emerging.** At this critical stage, tinkering with intellectual property rights threatens to disrupt these market-oriented developments. What is required at this time is not short-term thinking, but a steady policy framework and a commitment to positive, incremental change that is respectful of content holders' rights.

About Venture Insights

Venture Insights is an independent company providing research services to companies across the media, telco and tech sectors in Australia and New Zealand.

For more information go to ventureinsights.com.au or contact us at contact@ventureinsights.com.au.

Important notice: By accepting this research note, the recipient agrees to be bound by the following terms of use. This research note has been prepared by Venture Insights Pty Ltd and published solely for guidance and general informational purposes to authorised users under the terms of a licence agreement between Venture Insights Pty Ltd and its subscriber. You need to be expressly authorised to use it, and it may only be used for your internal business purposes and no part of this note may be reproduced or distributed in any manner including, but not limited to, via the internet, without the prior permission of Venture Insights Pty Ltd. If you have not received this note directly from Venture Insights Pty Ltd, your receipt is unauthorised. If so, or you have any doubt as to your authority to use it, please return this note to Venture Insights immediately.

This research note may contain the personal opinions of research analysts based on research undertaken. This note has no regard to any specific recipient, including but not limited to any specific investment objectives, and should not be relied on by any recipient for investment or any other purposes. Venture Insights Pty Ltd gives no undertaking to provide the recipient with access to any additional information or to update or keep current any information or opinions contained herein. The information and any opinions contained herein are based on sources believed to be reliable, but the information relied on has not been independently verified. Neither Venture Insights Pty Ltd nor its officers, employees and agents make any warranties or representations, express or implied, as to the accuracy or completeness of information and opinions contained herein and exclude all liability to the fullest extent permitted by law for any direct or indirect loss or damage or any other costs or expenses of any kind which may arise directly or indirectly out of the use of this note, including but not limited to anything caused by any viruses or any failures in computer transmission.

Any trade marks, copyright works, logos or devices used in this report are the property of their respective owners and are used for illustrative purposes only. Unless otherwise disclosed, Venture Insights has no affiliation or connection with any organisations mentioned in this report. However, the information contained in this report has been obtained from a variety of sources, including in some cases the organisations themselves. In addition, organisations mentioned in this report may be clients of Venture Insights.

The recipient hereby indemnifies Venture Insights Pty Ltd and its officers, employees and agents and their related entities against any direct or indirect loss or damage or any other costs or expenses of any kind which they may incur directly or indirectly as a result of the recipient's use of this note.

© 2025 Venture Insights Pty Ltd. All rights reserved.

David Kennedy, Managing Director
david.kennedy@ventureinsights.com.au

Venture Insights
Level 3, 461 Bourke Street, Melbourne, VIC 3000