Copyright, AI, and the Harm We Are Not Measuring

The political and legal debate over artificial intelligence and copyright is already extensive: courts in multiple countries are adjudicating competing claims, legislatures are debating the relative merits of text and data mining exemptions and collective licensing proposals, and voluntary licensing markets are beginning to form. This breadth of activity is welcome. But one dimension of the problem has received considerably less attention than it deserves – that harm to rights holders is not caused by what goes into AI systems, but by what comes out.

Getting this right matters enormously. The creative industries and the AI sector are both large, both growing, and increasingly in conflict. The outcome of the copyright debate will shape incentives for human creativity for a generation, determine how the economic gains from AI are distributed, and set the terms on which AI development can proceed. A framework calibrated to the wrong problem will produce the wrong results.

Two markets, two problems

It helps to distinguish two markets in which AI and copyright interact, because they have different characteristics and call for different responses.

- The first is the market for training data. Building a capable AI model requires ingesting vast quantities of text, images, music and other content – much of it protected by copyright. The rights holders whose works were used often received nothing and were not asked. This is a genuine grievance. But the economic harm is less clear-cut than it might appear. Using a novel as training data does not obviously reduce sales of that novel; a well-trained model might, if anything, increase demand for original works through greater discoverability. The injury to rights holders in this market is primarily one of fairness and bargaining power: they are contributing to something of enormous value without any share of the proceeds.

- The second is the market for reproduced outputs – and this is where the more direct economic harm lies. AI models trained on large quantities of protected material can, in some circumstances, reproduce that material in their outputs: generating passages of text, reproducing lyrics, or producing images closely resembling specific artists' work. A model capable of reproducing copyrighted song lyrics on demand is a direct substitute for the licensed market for those lyrics. One that generates images in the style of a living artist can displace that artist's commissions. The harm here is concrete, measurable, and linked to specific capabilities of specific models.

The distinction between these two markets is not always clearly maintained in policy discussions. Most current proposals – text and data mining exemptions, opt-out registers, voluntary and collective licensing arrangements – address primarily the training data market. This is understandable: training-data access is the first point at which rights holders encounter AI, and collective licensing is an established model that can be adapted. But the reproduced outputs market, though increasingly central to litigation, has received comparatively less attention in the legislative debate.

What the courts have found

The courts have begun to make this distinction, albeit inconsistently. In November 2025, the Munich Regional Court found against OpenAI in a case brought by GEMA, the German music rights organisation, ruling that what matters is not whether a model stores copies of protected content in any technical sense, but whether it can reproduce that content when prompted. This is the right economic question: the harm to rights holders lies not in the act of training but in the capability the training creates.

By contrast, a California federal court found in favour of Meta in a case brought by authors whose books had been used to train the Llama model, on the grounds that the training was transformative and that the plaintiffs had presented no

meaningful evidence of market harm. The court was explicit that its finding turned on the evidence: a case demonstrating that a model regularly reproduced substantial portions of protected works, or materially reduced demand for the originals, might be decided differently.

Market evidence

This divergence reflects a genuine empirical uncertainty. The harm caused by AI-generated outputs varies across models, content types and markets. In stock photography, the evidence for substitution is already clear: AI-generated images are displacing human-created ones, particularly at the lower end of the market. In music, the evidence is more nuanced. On Deezer, AI-generated tracks now account for a substantial of newly uploaded content but for a still negligible share of actual streams, so aggregate displacement looks limited so far. But the underlying capability is real: in mid-2025, an undisclosed AI act, The Velvet Sundown, attracted over a million monthly Spotify listeners on the strength of music indistinguishable from human output – a reminder that, where AI-generated content is not labelled as such, substitution may simply be invisible in the data. For longer-form written work the picture is more mixed, though surveys of novelists and other writers report significant income effects.

The direction of travel is consistent across content types, even if the scale, timing and visibility of the effect vary considerably.

Charging for capability, not for inputs

If the harm lies in what a model can do rather than in what was used in its training, the remedy should be calibrated accordingly. This is the central insight behind a capability-based approach to AI copyright levies.

The model draws on a well-established precedent. When magnetic tape recording became widely available in the 1970s, several European countries introduced levies on blank tapes. The legal and economic logic was that the tape manufacturer, while not itself infringing copyright, was producing a product whose value derived substantially from the ability to copy protected content at scale. A levy on each unit sold compensated rights holders without requiring anyone to identify which specific recordings had been copied.

Though much maligned for the arbitrariness in its implementation, this concept provides a helpful analogy for AI. Rather than taxing the inputs to AI development – the training compute, the training data – a capability-based approach would levy AI models based on their demonstrated capacity to reproduce protected expression. Technical researchers have developed methods for measuring what they call the memorisation rate: the proportion of a model's training material that can be extracted through systematic prompting. A model with a low memorisation rate, whose outputs reflect genuine generalisation rather than reproduction, would attract little or no levy. One with a high memorisation rate – capable of reproducing substantial quantities of protected text, images or music on demand – would attract a proportionally higher charge.

This approach has several advantages over the alternatives. It aligns the payment obligation with the actual harm created rather than with the act of training, which may cause little harm at all. It creates direct incentives for AI developers to adopt techniques that reduce memorisation – desirable independently of the copyright question, since memorisation also poses risks for the privacy of individuals whose personal data appears in training sets. And it is more resistant to territorial arbitrage: unlike measures that target the training process, which can be relocated to more permissive jurisdictions, a deployment-based levy applies where the users are, regardless of where the model was built.

Attribution: recognising the creative source

A levy, however well-designed, addresses only part of the problem. Research consistently finds that many creators are not primarily motivated by royalty income. The absence of credit – the use of their work without acknowledgement – is experienced as a significant harm in its own right, and one that demonstrably affects willingness to create. Studies have found that creators who discover their works have been used for AI training, even before any measurable financial loss, reduce their subsequent creative output available for training.

This matters for practical economic reasons beyond sentiment. Most professional artists and writers derive most of their income not from royalties but from commissions, performances, teaching and other activities that depend on reputation. An AI system that draws on a creator's work without acknowledgement gradually erodes the reputational foundation on which that income depends. A policy that compensates financially while ignoring attribution addresses only part of the incentive structure that sustains creative production.

Mandatory source attribution in AI outputs would address this directly. If users were informed that a response drew substantially on particular sources – a specific legal text, a particular artist's catalogue, a given author's body of work – the creator would receive recognition even where no direct infringement has occurred.

Crucially, an attribution mechanism that addresses this non-pecuniary harm could also solve a practical problem inherent in any levy system: how to distribute proceeds equitably among rights holders when the link between specific training works and specific outputs cannot easily be traced.

Building attribution capability into AI systems from the outset, rather than attempting to reconstruct it retrospectively, is substantially cheaper and more accurate. Some AI architectures already do this as a matter of course. Extending the requirement across deployed systems, backed by regulatory mandate, should be feasible in principle. AI developers have commercial reasons to resist it – detailed attribution reveals sensitive information about training choices and creates evidence useful in infringement claims – which is why market forces cannot be relied upon to deliver it.

The international dimension

One structural difficulty with any copyright framework for AI is that copyright law is territorial while AI development is global. Measures targeting the training process are readily circumvented by relocating training operations. Both the capability-based levy and the attribution requirement attach to deployment rather than training and are therefore considerably harder to avoid. But a fully effective framework ultimately requires coordination across jurisdictions – at minimum, mutual recognition of collective licensing arrangements and agreed standards for attribution reporting.

Conclusion

The AI copyright debate has already generated a rich body of analysis and a range of serious proposals. Text and data mining exemptions, collective licensing schemes, opt-out and opt-in regimes, and voluntary market arrangements each have genuine merits, and the trade-offs between them have been carefully examined. The approach proposed here is intended to complement rather than displace that existing work. A capability-based levy and mandatory attribution address the reproduced outputs market – a dimension of the problem that current frameworks largely leave unresolved – and can in principle sit alongside whatever arrangements govern training data access. The technical tools required already exist. The architectural choices that would make attribution cheap to implement are still being made. The question is whether policymakers will engage with the output side of the AI copyright problem before that window closes.

A longer paper discussing the issues raised in this note in more detail is available at https://dx.doi.org/10.2139/ssrn.6814498