WORKING PAPER | ISSUE 09/2024 | 4 APRIL 2024

ECONOMIC ARGUMENTS IN FAVOUR OF REDUCING COPYRIGHT PROTECTION FOR GENERATIVE AI INPUTS AND OUTPUTS

BERTIN MARTENS

Generative artificial intelligence (GenAI) models have stirred considerable controversy about copyright protection for AI training inputs and model outputs. The European Union's AI Act will require model developers to be transparent about their use of training inputs such as text, images and music. The EU Copyright Directive allows free text and data mining of these media inputs unless copyright holders have opted out and want license payments.

The right to opt-out amounts to economically inefficient overprotection of copyright. Free use of media content for GenAI training does not affect media sales to consumers. Opt-outs only strengthen the bargaining position of copyright holders, who decide depending on their private interests. That generates windfall profits without any increase in consumer surplus or social welfare.

The licensing of training inputs reduces the quantity of data and the quality of GenAI models, creates transaction costs and reduces competition between GenAI firms. This slows down GenAI-induced innovation in media products and production processes, and productivity gains in all service sectors that apply GenAI. Ultimately, it slows down economic growth compared to what it could be with competitive and high-quality GenAI.

Bargaining over license pricing is arbitrary as there is no objective revenue benchmark to start from. Defenders of the moral right to remuneration argue that any arbitrary remuneration is better than no remuneration. But this private moral right comes at the expense of social welfare. The ongoing bargaining and court cases between media producers and GenAI developers risk entrenching this market failure in jurisprudence. Early regulatory intervention and elimination of opt-outs for GenAI training, or weakening them in AI Act implementation guidelines, would solve this.

There is no need for copyright on GenAI outputs either. GenAI reduces the marginal cost of machine-production of media outputs to close to zero, on par with the marginal cost of reproduction. That eliminates incentives for piracy. Moreover, composite human-machine outputs benefit from a *de-facto* extension of copyright on the human component.

Bertin Martens (bertin.martens@bruegel.org) is a Senior Fellow at Bruegel and a non-resident research fellow at the Tilburg Law and Economics Centre, Tilburg University

Recommended citation:
Martens, B. (2024) 'Economic arguments in favour of reducing copyright protection for generative AI inputs and outputs', *Working Paper* 09/2024, Bruegel

bruegel

1 Introduction

Copyright is a policy tool to stimulate innovation in society. Granting exclusive private intellectual property rights, in the form of copyright, to human authors is meant to be an incentive for investment in the production of creative content such as books, music and movies. It is a tool to prevent commercial free-riding by others on that content. In continental Europe remuneration for creative content, is not just an economic tool but a fundamental right, as reflected for example in Recitals 9-11 of the EU Copyright (Information Society) Directive (2001/29/EC). Reducing the scope of copyright protection is thus perceived as weakening fundamental rights.

However, this defensive stance makes it difficult to bring technological change and wider economic considerations into the copyright debate. By contrast, the economic approach to copyright that is more in line with legal thinking in English-speaking common-law countries, seeks to strike a balance between incentivising private innovators and capturing the benefits of innovation for society as a whole. Over-protection and under-protection of private rights both hamper innovation. But finding the right balance is difficult and involves many trade-offs and redistributional effects between producers and users of innovation. Changes in content (re-)production technology upset that balance.

The arrival of generative artificial intelligence (GenAI) [1] has upset these trade-offs in the creative and media industries. GenAI models, such as large language models that can process natural language, are pre-trained on very large text datasets taken from books, documents, Wikipedia and webpages. The earliest models used billions of word tokens as training inputs; current models use trillions. Language models are actually a subset of a wider category of foundation models that can process any type of structured data, including audio and images, mathematics and computer code (CMA, 2023). Their very large scale has led to unexpected higher-level meta-learning properties, including the ability to learn with a few natural language prompts that guide models to answers by describing the similarity with other tasks or examples, or even to respond to queries to which the models had no or limited exposure during training (Zhao *et al*, 2022).

Additional properties are likely to emerge with the ever-increasing scale of models. Model performance can also be enhanced post-training by 'grounding' using additional datasets and adding 'plugins' of application-specific prompt sets. Reinforcement learning from human feedback, including deliberate attempts to trick or misguide new models, can correct model responses, to avoid 'hallucinations' or erroneous responses and harmful output. These properties make GenAI a very powerful tool to increase production and productivity in a wide range of language, audiovisual media, mathematical, scientific and engineering tasks. GenAI is rapidly becoming a general purpose technology (Bresnahan and Trajtenberg, 1995; Brynjolfsson *et al*, 2023) that has spillover effects or externalities on product and process innovation and productivity in virtually all industries.

[1] The label 'Generative AI' is used here for AI algorithms that are capable of taking natural language prompts as inputs to generate text, images audio and visual media. This includes large language model (LLM) chatbots such as OpenAI's ChatGPT, Google Bard, and Meta's LLaMA, and text-to-image AI systems such as Stable Diffusion and DALL-E.

These properties also explain the ambiguous relationship between GenAI and the media industries. GenAI models are trained on human-produced work, including content over which artists hold exclusive copyright. At the same time, human artists use GenAI to leverage their media productivity and explore innovative new media outputs that build on and compete with existing human content. The impact of GenAI models spills over far beyond media industries to every sector in the economy.

The EU AI Act tries to settle these tensions by applying a generic formulation[2]. The draft AI Act (European Commission, 2021) and subsequent amendments by the European Parliament led to a new Article 52(c) that requires GenAI developers to respect existing copyright on training inputs, as defined in the text and data-mining exception to copyright in the Copyright in the Digital Single Market Directive (CDSM, Directive (EU) 2019/790). Art 4 of the CDSM Directive makes re-use conditional on copyright holders not exercising an opt-out. The AI Act also imposes transparency obligations on GenAI model developers who should provide a *"summary"* of all copyrighted materials used as *inputs* for training purposes. There is no mention in the AI Act about copyright issues in GenAI *outputs*.

Existing text and data mining (TDM) provisions in EU copyright law, and similar 'transformative use' exceptions in copyright law in the United States, have not eliminated tensions between copyright holders and GenAI developers, if only because GenAI has pushed TDM technology and applications far beyond what was envisaged at the time of writing CDSM Art 4. In the US, big media companies and individual artists are taking GenAI producers to court over alleged infringements of their copyright, both in training inputs and in model outputs. Major GenAI investors, such as OpenAI and Microsoft, are negotiating license agreements with media companies to ease the tensions and reduce legal uncertainty – and also because they want to avoid punitive statutory damage claims under US law. Smaller GenAI start-ups may not be able to afford licenses and court battles. In the EU, the AI Act holds out the promise of licensing revenue for copyright holders. So far, this prospect has limited the number of court cases.

This paper complements the legal debate around copyright and GenAI with an economic perspective that views copyright as a tool to overcome market failure in media products, and to promote innovation and increase welfare for society as a whole, not only for private interest groups. Establishing an appropriate scope of copyright protection would balance the negative economic impacts of granting exclusive copyright to content creators, against the benefits of a continuous stream of investment in new and innovative content. Both over-protection and under-protection reduce the societal benefits from copyright. A key test for the appropriateness of the scope of copyright protection is the impact of a change in protection on the supply of innovative content. Changes in content-production technology, such as the arrival of GenAI models, may affect the appropriateness of protection.

[2] At time of writing, the AI Act had been agreed by the EU lawmaking institutions and partly ratified, but not finalised with publication in the EU Official Journal. The provisionally agreed version of the text is available at https://data.consilium.europa.eu/doc/document/ST-5662-2024-INIT/en/pdf.

With GenAI, copyright issues have grown out of the confines of the creative media industries and become a wider economic issue. Finding a new GenAI-induced balance in copyright law should not only take into account the welfare of media industry producers and consumers, but of the entire economy.

On the GenAI training inputs side, we argue that granting an unconditional TDM exception for the use of copyright-protected media will not reduce the supply of creative media outputs. But it promotes GenAI-based innovation and productivity gains, not only in media industries but across all sectors. TDM opt-outs and licensing requirements are therefore superfluous. They strengthen the bargaining position and generate windfall monopoly profits for copyright holders, at the expense of other sectors. Such opt-outs and requirements fragment the knowledge base on which GenAI models are trained, increase prices and transaction costs for training inputs, reduce competition in favour of very large GenAI developers and represent a major comparative disadvantage for GenAI developers in the EU compared to developers in jurisdictions with more liberal TDM regimes. CDSM Art 4 strengthens the bargaining position of private copyright holders at the expense of welfare benefits for society as a whole. That private market failure can be overcome by widening the scope of the TDM exception, either by a modification of Art 4 CDSM and/or in the copyright implementation guidelines for the AI Act.

On the outputs side, GenAI increases human productivity, not only for language, cultural and audio-visual media, the traditional domain of copyright protection, but across a wide range of industries and human tasks where GenAI models are applied. The price effect of higher productivity and lower production costs will increase production and accelerate substitution between human- and machine-produced outputs, in media industries and elsewhere. While GenAI model development faces high fixed costs, and writing prompt sets to generate particular outputs can be costly, the marginal cost of producing GenAI machine outputs is very low and often close to zero. It does not require copyright protection to incentivise production. However, human authors may still claim copyright over the entire hybrid human-machine output, as long as the two are not separable. Any legal *de-minimis* human content threshold for copyright may slow down but not stop substitution.

The input and output sides cannot be considered entirely separately. More and better inputs will improve model performance and result in higher-quality outputs. Media industries may perceive the inability to raise licensing revenue on the inputs side as an opportunity cost, but that will almost certainly be compensated for by productivity gains from the application of GenAI models in media industries. For GenAI applications in other sectors, there are no opportunity costs from copyright licensing fees, only pure productivity gains. From a societal welfare perspective, the combined impact on all sectors should be taken into account.

This paper is structured as follows. Section 2 provides a short overview of the legal arguments in the GenAI and copyright debate, both on the input and on the output sides of media industries. Section 3 takes a closer look at various economic arguments, revolving around market failures. It concludes that there is no economic need to assign copyright to GenAI media outputs and argues in favour of widening the scope of the TDM exception in Art 4 CDSM to avoid generating new market failures in private bargaining between copyright holders and GenAI developers. Section 4 concludes.

2 The legal perspective on GenAI and copyright

This section briefly summarises the legal perspective on the two policy questions: copyright for GenAI inputs and outputs. Legal scholars tend towards a narrow interpretation of the copyright exception for GenAI training inputs, and towards a broad refusal to grant copyright on GenAI outputs.

2.1 The inputs side

The draft AI Act (European Commission, 2022) and subsequent amendments by the European Parliament led to a new Art 52c (1) that requires GenAI developers to

"(c) Put in place a policy to respect Union copyright law in particular to identify and respect, including through state of the art technologies, the reservations of rights expressed pursuant to Article 4(3) of Directive (EU) 2019/790

(d) Draw up and make publicly available a sufficiently detailed summary about the content used for training of the general-purpose AI model, according to a template provided by the AI Office"[3].

Paragraph (c) refers to Art 4(3) of the EU Copyright in the Digital Single Market (CDSM) (Directive (EU) 2019/790) that provides an exception to copyright for text and data mining (TDM) for lawfully accessible works. CDSM Art 4(3) restricts this exception to "*the use of works that has not been expressly reserved by their right holders in an appropriate manner, such as machine-readable means in the case of content made publicly available online*". Paragraph (d) enables rights holders and policymakers to verify compliance with Art 52(c). Quintais (2023) argued that the transparency disclosure obligations in the EU AI Act would be impossible to comply with at a meaningful level of fine-graining of the data.

TDM is legal jargon for web-scraping technology that is widely used and enables researchers to assemble large datasets by scraping webpages. CDSM Art 3 allows free-of-charge TDM for non-commercial scientific research. CDSM Art 4 allows free TDM for any purpose, subject to lawful access to the content and the possibility for the copyright holder to block web-scraping by inserting a machine-readable 'no scraping' or 'no robots' text string in the page, or in the terms and conditions for access to the page. Machine-readable opt-outs can also be digitally embedded in pictures and music. In such cases, TDM is conditional on negotiation and payment of a license fee to the rights holder. The EU AI Act simply confirms the applicability of CDSM Art 4 to training of AI models. Opt-outs and license requirements impose a significant constraint on TDM activities (Margoni and Kretschmer, 2022).

[3] Ibid.

Elkin-Koren and Weinstock (2020) argued that the EU TDM exception is too restrictive and puts EU GenAI research at a disadvantage compared to other jurisdictions that have more open and flexible provisions, including the US, Singapore, South Korea, Malaysia, Israel and Taiwan. Tyagi (2023) also recognized that the TDM exception in the EU CDSM is too narrowly defined to work for GenAI, and suggested that a broader general exception along the lines of Japan's copyright law would fit better. Article 30-4 in Japan's copyright law permits commercial TDM when not done for human enjoyment purposes (Ueno, 2021). Clearly, training of GenAI models falls in that category. However, the output of GenAI models can be used for human enjoyment. A similarly broad exception exists in the 2021 Singapore Copyright Act. It includes a mixture of a TDM exception for computational use by machine learning models (sections 243-244) and a US-style fair-use exception (sections 190-191) (Tan, 2024). However, the computational exception is subject to lawful access, including respecting paywalls and terms and conditions. The latter boils down to an opt-out possibility for content providers. The fair-use exception does not allow reproduction of the training inputs in the outputs of the AI model. Substitution with existing content remains a source of legal uncertainty. Care should be taken before jumping to conclusions about the EU's TDM disadvantage.

Margoni and Kretschmer (2022) presented a different perspective. They offered legal arguments in support of the view that there is "*no need for a TDM exception as the extraction of factual information from protected content is external to the remit of copyright*". It is only a particular expression of factual content that is protected by copyright, not the content itself. In fact, news publishers themselves, who are in the front line of TDM and copyright infringement claims against GenAI developers, widely use TDM to search continuously for new news items on the webpages of their competitors, which they then rewrite for publication on their own sites to circumvent copyright claims (Cagé *et al*, 2020). Moreover, they increasingly use GenAI services to write news articles (WAN-IFRA, 2023). Media industries want to benefit from license revenue from AI model training, while also benefitting from productivity gains from the use of these AI models.

The United Kingdom tried but failed to introduce an unconditional TDM exception. The UK Intellectual Property Office launched a public consultation[4] on the TDM exception in UK copyright law that only allows TDM for non-commercial research only – similar to EU CDSM Art 3. This led to a policy proposal for a mandatory TDM exception for any purpose and without opt-out – going beyond EU CDSM Art 4. However, the proposal was withdrawn by the end of 2022 *"for review"*, under pressure from copyright holders in media industries[5]. They argued that GenAI outputs were already competing with human media production, and licensing income would boost media industry business models and provide an incentive to invest in curated databases for GenAI inputs. This is a recycling of the anti-TDM arguments

[4] See UK Intellectual Property Office, 'Consultation outcome on Artificial Intelligence and Intellectual Property and government response to the consultation', 28 June 2022, https://www.gov.uk/government/consultations/artificial-intelligence-and-ip-copyright-and-patents/outcome/artificial-intelligence-and-intellectual-property-copyright-and-patents-government-response-to-consultation.
[5] Robert Dickens, 'UK re-considers proposed exception for text and datamining', blog post on the Allen and Overy website, 2 March 2023, https://www.allenovery.com/en-gb/global/blogs/data-hub/uk-re-considers-proposed-exception-for-text-and-data-mining.

of scientific publishers at the time of the debate on the EU CDSM: publishers would have to incur additional costs to create machine-readable databases for TDM (Hargreaves *et al*, 2014). The AI industry has rejected that argument.

In the US, the Fair Use and Transformative Use copyright doctrines allow more freedom in the use of copyright-protected inputs for purposes other than the original purpose around which copyright holders build their business models. However, the interpretation of these doctrines is subject to uncertainty, in particular when GenAI models produce outputs that compete with the training inputs. Several copyright holders have opened court cases against GenAI firms that rely on this doctrine to access copyright-protected material for free as training inputs for their foundation models. This includes the New York Times against OpenAI[6], and Getty Images against UK-based StabilityAI[7]. The latter is particularly interesting because it revolves around alleged cross-border violations of copyright, which is essentially a territorial right. Getty Images claimed that UK-based StabilityAI used its pictures as training inputs in the UK, where a limited TDM exception would apply that does not allow this. StabilityAI claims training was done in the US, where the wider copyright exception of fair use would apply. Cross-border shopping for justice has already started. The UK High Court (2023) allowed the case to proceed for further investigation.

Investors do not like this legal uncertainty, in particular because US courts can impose punitive statutory damages for copyright infringement. To avoid this, several larger GenAI firms, including OpenAI, Google and Microsoft, have signed license agreements with major news publishers. OpenAI reached an agreement with the Springer Group, but was taken to court by the New York Times over alleged reproduction of newspaper articles in its GenAI model outputs. Big GenAI firms with deep pockets are able to pay for license agreements. This may push smaller GenAI start-ups out of the market and reduce competition in GenAI services. It may also lead to biased selection of inputs for training of GenAI models.

The calculation of a license fee is a problem in itself. GenAI algorithms use millions of training inputs that may all contribute to an output. Should they all get a *pro-rata* share of the value of GenAI outputs or revenues, even if outputs do not closely resemble any of the inputs? That value cannot be determined during the training phase; it will only become clear once the model is used in business applications. Moreover, this approach runs into insurmountable transaction-cost obstacles. Some authors suggest collective management of lump-sum license fees (Dermawan and Mezei, 2023; Senftleben, 2023). This runs into the same problem of absence of a benchmark for the calculation of fees. One may also argue that any non-zero fee rate is better than no fee at all. It creates a perception of fairness because it redistributes part of the GenAI developer's surplus to the input providers. That

[6] Michael M. Grynbaum and Ryan Mac, 'The Times Sues OpenAI and Microsoft Over A.I. Use of Copyrighted Work', *The New York Times*, 27 December 2023, https://www.nytimes.com/2023/12/27/business/media/new-york-times-open-ai-microsoft-lawsuit.html.

[7] Emilia David, 'Getty lawsuit against Stability AI to go to trial in the UK', *The Verge*, 4 December 2023, https://www.theverge.com/2023/12/4/23988403/getty-lawsuit-stability-ai-copyright-infringement.

argument goes back to the continental European view of an absolute and fundamental right to a remuneration for the authors of creative content, irrespective of the economic conditions or circumstances.

2.2 The outputs side

Novelli *et al* (2024) summarised the legal questions on the outputs side as reflecting two issues: the possibility of granting copyright to outputs, and the legal relationship between the inputs used for training and the outputs. In most countries, copyright law is on the side of humans: copyright only applies to human output, not machine-generated output. There are exceptions though. The UK, Ireland, India, South Africa and Hong Kong can grant copyright to human owners and operators of machine-generated works. Gervais (2023) saw no need for copyright for GenAI outputs, not only because there is no underinvestment and thus no incentive needed, but also because, from a legal perspective, only humans can produce creativity and machines cannot exercise moral rights to copyright. In contrast, Abbott and Rothman (2022) argued that GenAI outputs should be protected precisely to stimulate the production of innovative contents. Bulayenko *et al* (2022) proposed a four-step test for GenAI outputs to qualify for copyright: (1) a *"production in the literary, scientific or artistic domain"*; (2) the product of human intellectual effort; (3) the result of creative choices; and (4) the choices are *"expressed"* in the output. They concluded that many GenAI productions would meet these criteria, and illustrated this with examples from the music industry.

In February 2023, the US Copyright Registration Office[8] accepted the copyright registration of a human-authored text prompt set that generated a picture, but rejected copyright on the picture itself. It argued that the author's creative prompt choices were only weakly reflected in GenAI picture output. The same prompt set can generate many different pictures. Hugenholtz and Quintais (2023) argued that a *"prompt"* text string is insufficient for copyright purposes. Attributing copyright to prompt sets has led to an emerging commercial market in copyright -protected complex prompt sets[9]. However, in January 2024, a Beijing court for the first time granted copyright to the human maker of an AI-generated picture (Wang, 2024), claiming that the four criteria were satisfied.

Dreier (2023) sided with the US Copyright Office in his examination of the attribution problems that emerge if copyright were to be granted to GenAI outputs. First, attributing copyright to the owner or developer of the GenAI algorithm boils down to an extension of copyright from the algorithmic code to the output of the algorithm. It would be equivalent to granting the producer of a musical instrument copyright on the music produced by that instrument. Second, attributing copyright to copyright holders in GenAI training data would make sense only if the output closely resembles one or several training inputs. It is not easy however to determine what constitutes an infringing output (Nordemann, 2024). In theory, infringement occurs when an output is identical to, or recognisable in, an existing

[8] US Copyright Office, Zarya of the Dawn, registration # VAu001480196; see https://www.copyright.gov/docs/zarya-of-the-dawn.pdf.
[9] See for example https://promptbase.com/.

original work. In practice, there are no standardised differentiation rules and courts have to judge this on a case-by-case basis. Under some conditions, GenAI models can fully replicate a copyright-protected training input (Emanuelov and Margoni, 2024), and that would clearly constitute an infringement. This claim underpins the New York Times court case against OpenAI.

Third, attributing copyright to the author of the prompt set confuses the content's composer and performer. In music or theatre for example, composers can have copyright over the musical notes or text that they compose, but not over a specific performance of this composition. In the case of GenAI, the machine is the performer. As a machine, it cannot claim copyright on a specific performance of the composition. In the next section, we argue that there is no economic need to grant copyright to a machine performance.

The refusal to grant copyright to GenAI outputs can be circumvented with hybrid outputs that combine human and machine performances. Hybrids are widely use, for example in music (Bulayenko *et al*, 2022). German copyright law now explicitly recognises these hybrid practices, including remixes, memes, image files, mashups, fan art, fan fiction and sampling (Bischoff, 2023). This raises the question of what the minimal human creative input would be to qualify for copyright protection. Fixing a *de-minimis* human input requirement in copyright-protected hybrid GenAI outputs is conceptually difficult. As a result, any human contribution to a mixed GenAI output may be sufficient to claim copyright over the entire output.

A specific case occurs when a GenAI algorithm produces a near-perfect imitation of a human voice or face. This may be classified as a 'deep fake' when it is done without the consent of the human source of the voice or face. However, humans may also deliberately insert their own voices or faces into GenAI model to generate lookalike or soundalike output. One could argue that these human attributes constitute personal data, subject to the EU general data protection regulation (GDPR), not copyright rules. Some artists have already agreed to the use of their voice or face in GenAI applications. The GDPR attributes some data rights to natural persons as data subjects. While a data subject cannot 'sell' her voice or face, she may grant a usage right in return for remuneration.

3 An economic perspective

The continental European legal view on copyright revolves around the moral rights of content creators to receive remuneration for the use of their works. By contrast, the economic perspective on copyright does not consider license fees as an absolute right for content creators. The scope of copyright protection and remuneration depends on content production technology and on the economic impact that a change in scope would have on production of innovative content and on downstream use of that content. The scope of copyright protection, including applicable conditions and exceptions, should be adapted to changes in content (re-)production technology.

The emergence of GenAI technology marks a decisive change in content-production technology. For the first time in human history, a technology competes with mankind's natural monopoly on creativity.

GenAI algorithms can produce creative content ranging from truly original to perfect copyright-infringing substitutes for existing human-produced content. Human creators will want to collaborate with GenAI machines to benefit from productivity gains and cost savings. This offers the prospect of major innovations. Consumers can become producers and co-producers of music, films and games. This opens the door to interactive music concerts and movie experiences, co-produced with AI avatars of well-known artists. It could blur the borderline between movies and games.

These changes in production technology affect the scope of copyright protection, both on the input and the output sides. This section explains why the economic view tends to reject the opt-out condition on copyright on the GenAI input side, but supports not granting copyright on the GenAI output side.

3.1 The inputs side: a TDM exception for GenAI training inputs

It is useful to start by recalling the economic reasoning behind the TDM exception for non-commercial scientific research in Art 3 EU CDSM. The main economic argument in favour of that limited TDM exception was that it would not have a negative supply-side effect: it will not diminish research output or the number of published research papers (Hargreaves *et al*, 2014). On the contrary, it will increase research productivity and thereby increase research output, which is beneficial for society. Moreover, positive spillovers from research output to productivity and growth in all industries will be far more important than any impact on the scientific publishing industry (Hargreaves *et al*, 2014). In other words, reducing the scope of copyright protection and granting a TDM exception for scientific research is a proper pro-innovation policy change. It uses copyright as an economic tool to stimulate innovation. Not granting the exception would have given excessive protection to the scientific publishing industry as the ultimate copyright holders on published scientific research. It would have reduced research and slowed down innovation and economic growth.

The same economic reasoning can be applied to a TDM exception for any commercial purpose in EU CDSM Art 4, including for using media inputs for training GenAI models. The standard business model and incentive for investment in media products, including books, newspapers, music and movies, is sales for human consumption, for *"human enjoyment"* to use the wording from Japan's copyright law (Ueno, 2020). Media industry revenue from sales for human enjoyment will not be affected when media products are used for GenAI training. It will not reduce investment or creative content output, and therefore will not result in under-protection of the creative content industries. Quite to the contrary, granting a TDM exception will stimulate GenAI model development and improve the quality of models by giving them access to larger sets of training data. That quality increase, in turn, will increase productivity of GenAI models, not only for producing creative content in media industries, but far beyond in many other industries in which GenAI models are applied for services production.

Not granting a TDM exception would restrict the supply of copyrighted training inputs to LLMs and will significantly increase data-acquisition costs, potentially hindering the development of GenAI. Yang and Zhang (2024) found that, when abundant model training data is available, unconstrained use of copyright-protected training inputs, without opt-out clauses and licensing conditions, increases GenAI model quality, resulting in higher profits for the AI company, higher aggregated incomes for content creators and greater consumer surplus. In short, TDM opt-outs on GenAI training data inputs give excessive protection to copyright holders, at the expense of the wider economy, and also at the expense of media industries as users of GenAI models.

Like scientific publishers at the time of the TDM debate in the context of the EU CDSM, media producers would like to benefit from an additional windfall profit by rejecting TDM and relicensing the same media products, for which they already have a sustainable business model, to GenAI firms, at no additional production cost. This would be a pure monopolistic windfall profit because it is not associated with the original purposes and main business models for which the media were produced, and it is not associated with an increase in consumer welfare or in overall welfare for society. It would cause an inefficient welfare shift in the copyright equilibrium towards over-protection of media industries at the expense of the wider economy and innovation in society at large. It would increase static monopoly rents for copyright holders at the expense of dynamic re-use and GenAI innovation welfare gains.

Some media producers claim that additional licensing revenue would give them an incentive to invest in the curation of GenAI training inputs. GenAI developers spend on the cleaning and curation of unstructured input datasets collected on the internet. However, there is no need to fix this in copyright law. GenAI developers can always offer media input providers payment for the supply of curated data, if that would be a more-efficient solution than in-house curation by GenAI developers. Such payments can be settled through the market.

The opt-out clause in Art 4 CDSM strengthens the bargaining position of media copyright holders. They can to decide on the basis of their private interest in matters that have spillover effects on society, far beyond their private interests. Private interests will take centre stage, at the expense of society's interest in innovation. Spillovers and externalities are not taken on board in private decisions. That results in over-protection of copyrighted content and a new market failure: positive externalities of the exception are not taken into account. We cannot label it yet as a regulatory failure because regulators could not have foreseen the emergence of GenAI at the time when the EU CDSM was negotiated. Until 2022, before the arrival of GenAI models, it could be argued that Art 4 CDSM represented an appropriate balance between the interests of copyright holders and society at large. With the arrival of a new GenAI technology, that balance is clearly broken and requires correction. Art 4§3 CDSM will become a regulatory failure if policymakers do not intervene to correct this over-protection and cancel the opt-out clause for the use of media products as training inputs for GenAI models. The opt-out clause in Art 4 CDSM will lead to monopolistic pricing of licenses for training inputs and reduce the amount of training inputs available for GenAI. GenAI developers will be more selective in training data. It may result in biases in the training set. It may reduce competition between GenAI developers, because smaller firms may not be able to afford the same amounts of training data. That reduces the quality of GenAI models for society at large.

Steps in this direction are already being taken, for example by the French music copyright management organisation SACEM. SACEM imposed a collective opt-out from GenAI model training for all French music that falls under its authority, based on Art 4 CDSM (Spitz, 2024). As a result, French music is likely to be marginalised in GenAI models. That may have negative spillover effects to French artists who want to use GenAI for the production of new audiovisual materials, and to other industries that rely on French language audio and musical inputs for the production of GenAI-driven services. Whether French GenAI start-ups are able and willing to pay for licenses is not clear. GenAI producers may turn to other sources of musical inputs and marginalise typical French music in GenAI-based synthetic media. SACEM, as an intermediary in the music supply chain, might increase its own revenue from this stance, but French cultural production and society at large may lose out from this.

3.2 Pricing of copyright licenses for training inputs

Apart from quantity rationing, Art 4 CDMS also introduces monopolistic pricing for copyright licenses giving access to GenAI training inputs. Pricing of TDM licenses runs into substantial theoretical and practical problems.

From a theoretical perspective, the standard economic model of production considers that, to be fair and efficient, remuneration of inputs should be in accordance with their marginal contribution to the value of outputs. This marginal remuneration rule is known in economics as the Euler Theorem[10]. It runs into problems when applied to GenAI models, for at least two reasons. First, unlike physical goods, data inputs and media outputs are non-rival products. They can be re-used without limits. That is the reason why copyright on these products is needed in the first place. However, non-rivalry undermines the Euler Theorem and remuneration according to marginal productivity. Doubling the volume of data inputs will not double GenAI model outputs or productivity (Romer, 1994; Buchanan and Yoon, 1999). Larger input datasets improve model performance but the marginal contribution of a single data input can vary from zero to very high, depending on the use of the model. This makes it impossible to determine an economically meaningful and efficient licensing price for media inputs for GenAI training purposes. Pricing becomes a pure bargaining issue.

License pricing also runs into practical problems. Pricing could be an arbitrary one-off lump sum, irrespective of the value of the GenAI output. Pricing could also be calculated as a proportional share of the output value of GenAI models. That opens up a new set of insurmountable computational problems. At what point in the GenAI value chain should output value be used as a benchmark – the revenues of GenAI developers, or the revenues of redeployers who add additional features and datasets to models, or prices charged to final users? What about open GenAI models that are available for free? For example, OpenAI's close collaboration with Microsoft results in several implicit and explicit financial transfers between the two firms. Microsoft, in turn, builds OpenAI's models into a wide range of consumer and business services and charges for this. Whose output value should be the benchmark

[10] For a more detailed explanation, see for example *Wikipedia*, 'Euler's theorem', https://en.wikipedia.org/wiki/Euler%27s_theorem.

for computing a share for the providers of media inputs to the training of OpenAI models? GenAI foundation models are extremely versatile and can be used for an infinite range of purposes and applications. If GenAI redeployers and users 'ground' their models with additional data, or add prompt-based applications on top of the model, what should be the benchmark for calculating license prices? Lump-sum pricing is the easiest way out, but is also completely disconnected from the value of applications and economic reasoning.

On top of license prices, negotiating TDM licenses with many rights holders generates high transaction cost for GenAI model developers. Developers will try to reduce these costs by restricting the number of licenses and will want to negotiate only with a few big global media copyright holders, such as major newspaper, book, music, picture and movie producers. That saves costs compared to negotiating with large numbers of smaller media producers. This results in further data fragmentation. Dermawan and Mezei (2023), Senftleben (2023) and Quintais (2023) acknowledged the problems of individual bargaining and suggested that a collective bargaining approach between national copyright management offices and copyright holders would reduce these costs. Geiger and Iaia (2024) proposed statutory licensing as a compromise solution for GenAI model training that preserves the fundamental right of authors to remuneration. Statutory licenses allow use of copyrighted material without the permission of the rights holder, but subject to the payment of 'equitable' remuneration, whatever this might be.

Imposing copyright licensing for inputs fragments the available data pool for GenAI training and weakens the benefits from data aggregation, in several ways. Only the largest firms with deep pockets will be able to pay for licenses that are often beyond the means of smaller AI start-ups. Smaller firms may only have access to unlicensed data that covers only part of the available data universe. That may result in a biased selection of fragmented training data, and therefore biased and less-performant models. It will reduce competition between GenAI developers.

GenAI represents a structural shift in learning processes. Until recently, symbolic learning was an individual human process. With digital algorithms, symbolic learning became possible outside human brains. Algorithms can collect and analyse information in digital symbolic format, and learn from it to guide human behaviour and produce all kinds of natural language, sound and image outputs. GenAI foundation models (Bommasani *et al*, 2020) learn by extracting insights from a huge pool of human knowledge and skills that resides in books, documents and texts, webpages, art and audio-visual products, etc. The sheer scale of that pool and algorithmic learning, using trillions of word tokens as inputs, results in unexpected emerging learning properties, such as the ability to learn in-context and apply insights to categories of data that were not part of the pool. GenAI models are still on an exponential growth path in terms of data pooling and computing capacity (Sevilla *et al*, 2022). Performance still improves with the amount of data they can access and process. There is no sign yet that GenAI models have run into diminishing returns on the size of the training datasets. On the contrary, some research claims that the largest models will soon be running short of data (Villalobos *et al*, 2022). This suggests that GenAI foundation models are still benefitting from positive marginal returns to economies of scale and scope in data aggregation (Bajari *et al*, 2018; Calzolari *et al*, 2023; Caballa *et al*, 2023): more insights can be extracted from larger and more aggregated datasets. Only large data pools can reap these data-driven externalities. Smaller models exist and can be sufficiently powerful for many applications, but they are not at the performance frontier. Any reduction in the volume of data that GenAI developers can access would have a negative impact on model performance.

The European continental law tradition considers the right of copyright holders to remuneration as a fundamental moral right of authors. In that legal context, one could argue that, even when fee calculations are difficult, any fee above zero is fairer than no fee at all because it redistributes some of the profits from GenAI developers to input providers. But there are no good economic reasons for redistributing that surplus: it will reduce the quality of GenAI models and productivity gains in media industries and far beyond, and thereby reduce the social-welfare gains from GenAI technology. Fairness will only benefit copyright holders, at the expense of the rest of society. It may even reduce benefits for copyright holders because they get worse GenAI models with which to produce their own content outputs.

In conclusion, the TDM opt-out n Art 4(3) CDSM only increases the private bargaining power of copyright license holders in their negotiations with GenAI model developers (Quintais, 2023). It does not increase economic efficiency for society at large. On the contrary, it reduces innovation, competition and economic growth.

3.3 Who owns the outputs from GenAI models?

GenAI is particularly suitable for generating new text, audio and images, based on human input through natural language prompts. Just to pick a few random examples: an award-winning Japanese novel was co-written by a human author and ChatGPT[11]. GenAI models can help novice artists produce music, often with an overwhelming number of varieties that require additional steering to guide them to the most suitable outputs (Louie *et al*, 2020). GenAI can also produce images and video, with very little human effort. For example, the Promptbase.com platform sells thousands of cheap application tools to produce endless varieties of image outputs. OpenAI rolled out an app store with a rapidly growing number of application-specific versions of ChatGPT designed for a very wide spectrum of human tasks.

GenAI models are applied far beyond the media industries, in virtually all sectors of the economy. Brynjolfsson *et al* (2023) presented empirical evidence that GenAI boosts human productivity in customer support services by 14 percent on average and up to 34 percent for novice and low-skilled workers. It facilitates product innovation in domains that were until recently beyond the reach of human invention. For example, Google-DeepMind AI models mapped the folding of over 200 million

[11] Sonja Anderson, 'ChatGPT Helped Write This Award-Winning Japanese Novel', *Smithsonian Magazine*, 24 January 2024, https://www.smithsonianmag.com/smart-news/this-award-winning-japanese-novel-was-written-partly-by-chatgpt-180983641/.

proteins that may have useful applications in new pharmaceutical products, a task that would have taken human researchers centuries at the pre-AI rate of progress[12]. The importance of GenAI beyond media industries is obvious when looking at GDP data. In 2019, the cultural and creative media industries were estimated to represent 4.4 percent of EU GDP (Ernst and Young, 2021). Inputs from these industries in GenAI model training affect innovation, productivity and growth across the entire economy, as recognised by some legal scholars (Keller, 2023).

Companies that apply GenAI models to increase productivity can easily appropriate the revenues and profits from these services applications. The situation is more complicated for media industries that use GenAI to produce media content outputs and sell them to wider audiences. Copyright law takes an inherently anthropocentric stance because, until very recently, all creative content was human work. Legal scholars reject copyright on GenAI machine outputs and agree that only human authors can claim copyright on creative content. However, as long as there is a human contribution to a composite human-machine output, humans can claim copyright over the entire media output, unless the human and machine contribution can be separated in the output. Bulayenko *et al* (2022) illustrated how this hybridisation of outputs is already happening in music production. Human artists can increase their marginal productivity substantially by combining near-zero marginal cost machine outputs with their own positive marginal cost human contributions. That constitutes a strong incentive to substitute human outputs with machine outputs.

Claiming an exclusive copyright on the composite human-machine output *de facto* increases the price of composite GenAI outputs to a monopolistic price, above marginal cost, by hedging it to the monopolistic price that human authors can set for their contributions. Human artists appropriate the combined profit margins. That boosts human monopoly rents on content. The attribution problems identified in legal studies (Dreier, 2023) are unlikely to be very relevant in that economic substitution process. Setting a *de minimis* human contribution threshold may slow down but not eliminate that substitution. Media products with less human content are likely to crowd out products with higher human content because of price, quantity and productivity effects, irrespective of the substantially regime that applies to these outputs.

Economic reasoning supports the rejection of copyright on machine outputs, though for different reasons: there is no reason to grant an exclusive copyright to machine outputs because the marginal cost of the original production equals the marginal cost of reproduction. Copyright protection is necessary to avoid free-riding when marginal reproduction costs are lower for consumers than for content producers (Belleflamme and Peitz, 2014; Landes and Posner, 1989). For most of the history of mankind, there was no need for copyright protection because the marginal cost of producing a creative work was equal to the marginal cost of reproducing or copying it. Re-performance was the only technology available for reproduction. Artists who produced a painting, a song, a theatre play or a text could only reproduce it by painting, singing, playing or writing it again. That changed when more cost-

[12] See https://deepmind.google/technologies/alphafold/.

efficient reproduction technologies were invented, starting with the Gutenberg press in 1436. Books became an industry with high fixed but low marginal reproduction costs, compared to the cost of manually rewriting books. The phonograph had the same effect on music: the marginal cost of reproduction dropped dramatically compared to original music performances. Cameras did the same for theatre and spawned an entirely new derived industry: movies. For a while, media producers had exclusive control over reproduction devices. Copying a phonogram was too costly for piracy, until magnetic sound recording devices for consumers arrived. Movie producers controlled cinema entry tickets and video rental, until video recorders circumvented these technical means of exclusion and gave consumers the advantage with lower reproduction costs. With the arrival of digital information technology, media reproduction costs for consumers fell to the lowest possible level, near-zero, unleashing a wave of piracy. That lasted until music and video streaming services levelled the marginal cost playing field again between producers and users. With fixed monthly subscriptions, consumers and producers equally benefitted from zero marginal reproduction costs for music and movies, reducing piracy and stimulating legal consumption.

The arrival of GenAI marks a new turn in this technological arms race between content producers and consumers. Previous technologies reduced the cost of *reproduction* of original content. GenAI reduces the cost of original *production* of content. Content may still require an intellectual effort by the author to construct a prompt set. But the actual cost of production, the transformation of the intellectual effort into a digital or analogue 'performance', is dramatically reduced. It reduces the marginal cost of original production or performance to near-zero. Artists can write a new story, produce a new song or even a movie by feeding a GenAI algorithm with a few prompts. The fixed cost for GenAI model developers to set up and train algorithms can run into many millions of dollars. The costs for model deployers and users to design complex prompt sets, and feed an existing model with specific data, can also be substantial. But once the model is trained, additional data is uploaded and the prompt set designed, the marginal cost of actually producing a new media product, or performing the prompt set with GenAI, is very low.

This is where economic considerations converge with the legal view that there should be no copyright on GenAI outputs, though for very different reasons. With near-zero marginal production costs, there is no need for exclusive copyright protection to incentivise the production of creative media output. GenAI algorithm owners and users may implement technical means to control distribution of output to users and charge a positive price through these channels, similar to media streaming service providers. Algorithm owners may sell access to the algorithm with fixed cost and zero marginal cost subscriptions, without claiming rights over media output. GenAI output markets can still be profitable without copyright as long as producers can control distribution through tickets and subscriptions. Human artists can still be remunerated for GenAI content provided they can organise excludable 'live' performances or mix human and machine contributions in mixed media output that benefits from copyright protection for the human component. It may spawn entirely new interactive media industries in which artists and consumers co-generate content in real-time.

Although prompts are considered legally insufficient as human input to qualify for copyright on GenAI output (Hugenholtz and Quintais, 2021), complex and well-designed prompts have value in marketplaces, as the examples of Promptbase.com and the OpenAI apps store show. A complex set of prompts amounts to human-written computer code. When assembled into applications on top of a foundation model, they become 'apps' in their own right and may be subject to copyright protection. The judgement of the US Copyright Registration Office that granted copyright on a unique and humanly designed prompt set, as reported above, confirms this view. Foundation model platforms with app stores on top, such as the OpenAI apps store, are increasingly becoming a standard business model for GenAI model developers. The production of prompts is a fixed cost for creative content. However, the marginal cost of producing an original expression of prompts in the form of an content output is close to zero. Copyright for prompt sets and the existence of prompt marketplaces suggest that human designers can still charge a price above marginal cost to make a living out of that trade, even in the absence of copyright protection for the performance output of their prompt sets. This suggests that there is no market failure in prompt applications and no economic need for copyright protection.

4 Conclusions and a policy recommendation

Copyright is an economic policy tool to promote investment in innovative content, in the presence of market failures in non-rival and hard-to-exclude media products. Granting exclusive copyrights to authors overcomes free-riding that would discourage investment in creative content. It is important, however, to get the scope of copyright protection right. Both over- and under-protection may reduce innovation. The arrival of GenAI technology that can be used to produce creative media content upsets the existing trade-off. GenAI reduces media production costs and triggers price, quantity and substitution effects. Substitution between human and machine outputs triggers concerns about human employment in all product and services markets. In media markets, it raises questions about copyright for machine-generated outputs and the payment of copyright license fees on media inputs used for training of GenAI models. Decisions on these issues will have spillover effects reaching far beyond media markets to all sectors that use GenAI models.

This paper argues that the conditional TDM exception in Art 4 CDSM, which provides copyright holders with the possibility of an opt-out, amounts to over-protection of copyright, for several reasons. First, there is no need for an opt-out clause because TDM will not have a negative effect on the supply of creative media contents – the contrary, in fact. Copyright is an economic policy instrument to promote innovation. Any extension of the scope of copyright protection beyond what is needed for the production of innovation only strengthens the bargaining position of copyright holders and generates windfall profits, without any increase in consumer surplus or welfare for society as a whole. Second, copyright holders in media industries take decisions on the basis of their interests and ignore spillover effects to other industries. TDM licensing would reduce the quality of GenAI models across the entire economy, not only in the media sector. This results in a market failure that requires regulatory intervention. Third, bargaining over the price of TDM licenses under opt-out conditions in inherently inefficient and arbitrary. There is no meaningful economic way of setting an objective revenue benchmark. Defenders of the moral right to remuneration may argue that any arbitrary remuneration is better than no remuneration at all. But economists would argue that this private moral right would come at the expense of welfare for society as a whole.

There is no need for copyright on GenAI outputs. There is no risk of free-riding because the marginal cost of GenAI machine production is very low, close to zero, and most likely on par with the marginal cost of reproduction. That creates an economic level playing field between content producers and copiers, both operating at social welfare-maximising marginal cost. Composite human-machine outputs can already benefit from a *de-facto* extension of copyright from the human part to the entire output, unless the two are separable. However, GenAI models risk producing outputs that closely resemble training inputs and may therefore be classified as copyright-infringing. In the absence of general standards for infringement, this remains a case-by-case decision for courts. Granting copyright to machine outputs or licensing of training inputs will not affect that risk.

This leads to the policy conclusion that the conditional TDM exception in Art 4 CDSM results in market failures in private bargaining on TDM licenses and creates excessive copyright protection for GenAI training inputs. This can only be corrected by widening the scope of Art 4 CDSM and dropping the opt-out from the TDM exception for copyright holders in case of GenAI training applications. A second-best solution could be to introduce a very flexible interpretation of TDM conditions for GenAI training in the implementation guidelines for Art 52(c) of the AI Act.

The ongoing wave of private bargaining between major media companies and GenAI developers, including court cases and out-of-court settlements, in the EU and the US, is risky because it may create precedents and further entrench this market failure in jurisprudence. Early regulatory intervention would prevent inefficient outcomes that benefit the private interests of copyright holders at the expense of innovation and growth in the wider economy, and thus at the expense of society at large.

References

Abbott R. and E. Rothman (2022) 'Disrupting creativity: Copyright law in the age of generative artificial intelligence', *Florida Law Review*, August

Bajari, P., V. Chernozhukov, A. Hortaçu and J. Suzuki (2018) 'The impact of big data on firm performance, an empirical investigation', NBER Working Paper 24334, National Bureau of Economic Research

Belleflamme, P and M Peitz (2014) 'Digital piracy: an update', *Discussion Paper* 2014/19, UC Louvain Center for Operations Research and Econometrics, available at https://dx.doi.org/10.2139/ssrn.2537269

Bischoff, S. (2023) 'The dawn of pastiche: First decision on new German copyright exception', *Kluwer Copyright Blog*, 7 June, available at https://copyrightblog.kluweriplaw.com/2023/06/07/the-dawn-of-pastiche-first-decision-on-new-german-copyright-exception/

Bommasani, R., D.A. Hudson, E. Adeli, R. Altman, S. Arora, S. von Arx ... K. Zhou (2022) 'On the Opportunities and Risks of Foundation Models', Center for Research on Foundation Models, Stanford Institute for Human-Centered Artificial Intelligence, Stanford University, available at https://arxiv.org/abs/2108.07258

Bresnahan, T. and M. Trajtenberg (1995) 'General purpose technologies "Engines of growth"?' *Journal of Econometrics* 65(1): 83-108

Brynjolfsson, E., D. Li and L. Raymond (2023) 'Generative AI at Work', *NBER Working Paper* 31161, National Bureau of Economic Research

Buchanan J. and Y. Yoon (1999) 'Generalized Increasing Returns, Euler's Theorem, and Competitive Equilibrium', *History of Political Economy* 31(3): 511-523

Bulayenko, O., J.P. Quintais, J. Poort and D. Gervais (2022) 'Challenges to CR in AI music outputs', *Kluwer Copyright Blog*, 22 April, available at https://copyrightblog.kluweriplaw.com/2022/04/22/ai-music-outputs-challenges-to-the-copyright-legal-framework-part-i/

Cagé, J., N. Hervé and M.-L. Viaud (2020) 'The Production of Information in an Online World', *Review of Economic Studies* 87(5): 2126-2164

Calzolari, G., A. Cheysson and R. Rovatti (2023) 'Cooperative data-analytics: a market for machine-data', *Discussion Paper* 17842, Centre for Economic Policy Research

Carballa, B., N. Duch-Brown, S. Hacuk, P. Kumar, B. Martens, J. Mulder and P. Prüfer (2023) 'Economies of scope in data aggregation: the case of health data', *TILEC Discussion Paper* 2023-3, Tilburg University

CMA (2023) *AI Foundation Models: Initial review*, UK Competition and Markets Authority

Dermawan, A. and P. Mezei (2023) 'The Jakarta Effect: Could Southeast Asia's Consensus-Based Remuneration Approach Be the Blueprint for Human-Centric AI?' *Kluwer Copyright Blog*, 14 December, available at https://copyrightblog.kluweriplaw.com/2023/12/14/the-jakarta-effect-could-southeast-asias-consensus-based-remuneration-approach-be-the-blueprint-for-human-centric-ai/

Dreier, T. (2023) 'The challenge to attribute AI output to inputers', presentation to the European Copyright Society conference, 26 May, available at https://www.law.kuleuven.be/citip/en/citip-conferences/ecs/ecs-2023/presentations/ecs-2023-thomas-dreier.pdf

Emanuilov, I. and T. Margoni (2024) 'Memorisation in generative models and EU copyright law: an interdisciplinary view', *Kluwer Copyright Blog*, 26 March, available at https://copyrightblog.kluweriplaw.com/2024/03/26/memorisation-in-generative-models-and-eu-copyright-law-an-interdisciplinary-view/

Elkin-Koren, N. and N. Weinstock (2020) 'Transplanting Fair Use across the Globe: A Case Study Testing the Credibility of U.S. Opposition', 72 *Hastings Law Journal* 1121

Ernst and Young (2021) *Rebuilding Europe: the cultural and creative economy before and after the Covid crisis*, study commissioned by the European Grouping of Societies of Authors and Composers

European Commission (2021) 'Proposal laying down harmonized rule on artificial intelligence (the Artificial Intelligence Act)', COM/2021/206

Geiger, C. and V. Iaia (2024) 'Towards a statutory remuneration right for machine learning of GenAI', *Computer Law & Security Review* 52

Gervais, D. (2023) 'AI and CR: a reply to Matt Hervey', *Kluwer Copyright Blog*, 16 February, available at https://copyrightblog.kluweriplaw.com/2023/02/16/ai-and-copyright-a-reply-to-matt-hervey/

Hargreaves, I., L. Guibault, C. Handke, P. Valcke, B. Martens and R. Lynch (2014) *Standardisation in the area of innovation and technological development, notably in the field of Text and Data Mining*, Report from the Expert Group, European Commission, Publications Office of the EU

Hugenholtz, B. and J. Quintais (2021) 'Copyright and Artificial Creation: Does EU Copyright Law Protect AI-Assisted Output?' *International Review of Intellectual Property and Competition Law* 52: 1190-1216

Keller, P. (2023) 'Generative AI and copyright: Convergence of opt-outs?' *Kluwer Copyright Blog*, 23 November, available at https://copyrightblog.kluweriplaw.com/2023/11/23/generative-ai-and-copyright-convergence-of-opt-outs/

Landes, W. and R. Posner (1989) 'An Economic Analysis of Copyright Law', *Journal of Legal Studies* 18(2): 325-363

Louie R., A. Coenen, C. Zhi Huang, M. Terry and C. Cai (2020) 'Novice-AI Music Co-Creation via AI-Steering Tools for Deep Generative Models', *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*: 1–13

Margoni, T. and M. Kretchmer (2022) 'A Deeper Look into the EU Text and Data Mining Exceptions: Harmonisation, Data Ownership, and the Future of Technology', *GRUR International* 71(8): 685-701

Nordemann J.B. (2024) 'EU law: Generative AI, copyright infringements and liability', *Kluwer Copyright Blog*, 23 January, available at https://copyrightblog.kluweriplaw.com/2024/01/23/eu-law-generative-ai-copyright-infringements-and-liability-my-guess-for-a-hot-topic-in-2024/

Novelli, C., F. Casolari, P. Hacker, G. Spedicato and L. Floridi (2024) 'Generative AI in EU Law: Liability, Privacy, Intellectual Property, and Cybersecurity', mimeo, available at https://arxiv.org/abs/2401.07348

Romer, P.M. (1994) 'The Origins of Endogenous Growth', *Journal of Economic Perspectives* 8(1): 3-22

Quintais, P. (2023) 'Generative AI, Copyright and the AI Act', *Kluwer Copyright Blog*, 9 May, available at https://copyrightblog.kluweriplaw.com/2023/05/09/generative-ai-copyright-and-the-ai-act/

Senftleben M. (2023) 'Generative AI and Author Remuneration', *International Review of Intellectual Property and Competition Law* 54: 1535-1560

Sevilla, J., L. Heim, A. Ho, T. Besiroglu, M. Hobbhahn and P. Villalobos (2022) 'Compute Trends Across Three Eras of Machine Learning', mimeo, available at https://arxiv.org/abs/2202.05924

Spitz, B. (2024) 'AI data mining: French music collecting society SACEM opts out', *Kluwer Copyright Blog*, 25 January, available at https://copyrightblog.kluweriplaw.com/2024/01/25/ai-data-mining-french-music-collecting-society-sacem-opts-out-with-what-consequences/

Tyagi, K. (2023) 'Generative AI: Remunerating the human author & the limits of a narrow TDM Exception', *Kluwer Copyright Blog*, 13 December, available at https://copyrightblog.kluweriplaw.com/2023/12/13/generative-ai-remunerating-the-human-author-the-limits-of-a-narrow-tdm-exception/

Tan, D. (2024) 'Generative AI and Copyright Fair Use', Centre for Technology, Robotics, Artificial Intelligence & the Law, National University of Singapore, March, available at https://law.nus.edu.sg/trail/generative-ai-copyright-fair-use/

Ueno, T. (2021) 'The Flexible Copyright Exception for "Non-Enjoyment" Purposes – Recent Amendment in Japan and Its Implications', *GRUR International* 145

Villalobos P., J. Sevilla, L. Heim, T. Besiroglu, M. Hobbhahn and A. Ho (2022) 'Will we run out of data? An analysis of the limits of scaling datasets in Machine Learning', mimeo, available at https://arxiv.org/abs/2211.04325

Wang, Y. (2024) 'Beijing Internet Court Grants Copyright to AI-Generated Image for the First Time', *Kluwer Copyright Blog*, 2 February, available at https://copyrightblog.kluweriplaw.com/2024/02/02/beijing-internet-court-grants-copyright-to-ai-generated-image-for-the-first-time/

WAN-IFRA (2023) Gauging Generative AI's impact on newsrooms, World Association of News Publishers and Schickler, available at https://wan-ifra.org/mp-files/gauging-generative-ai-impact-on-newsrooms.pdf/

Yang S. and A. Zhang (2024) 'Generative AI and Copyright: A Dynamic Perspective', mimeo, available at https://dx.doi.org/10.2139/ssrn.4716233

Zhao, X., S. Ouyang, Z. Yu, M. Wu and L. Li (2022) 'Pretrained language models can be fully zero-shot learners', mimeo, available at https://arxiv.org/abs/2212.06950

bruegel

Bruegel, Rue de la Charité 33, B-1210 Brussels
(+32) 2 227 4210
info@bruegel.org
www.bruegel.org