From: Paul Jurcys
pjurcys@gmail.com

To: U.S. Copyright Office
101 Independence Ave SE
Washington, DC 20540-6000

Subject: Public Consultation Response on "Copyright and AI" [Docket No. 2023-06]

Dear Members of the U.S. Copyright Office,

We are writing to submit our response to the call for public consultation on the topic of "Copyright and AI." As professionals deeply engaged in the realms of data science, intellectual property law, creative arts and technology, we appreciate this opportunity to contribute our insights and perspectives on how copyright law should evolve in the face of rapid advancements in artificial intelligence (AI).

The intersection of AI and copyright law presents novel challenges and opportunities, necessitating thoughtful analysis and strategic adaptation of our existing legal frameworks. In our submission, attached below, we have endeavored to provide a balanced view, highlighting the need for clarity in legal definitions, the emphasis of human control in the creative process that involves AI tools, and the importance of incentivizing innovation while protecting creators' rights. Due to limited time we were not able to provide responses to all questions.

We are optimistic that this public consultation will pave the way for robust and informed discussions, leading to better insights and understanding how generative AI affects the existing principles of copyright law and creative processes. We are aware of the imperative to strike a delicate balance between fostering technological advancement and upholding the fundamental principles of copyright law.

We would like to applaud the U.S. Copyright Office for taking the lead in fostering the debate around the complex interplay between copyright law and AI. We woud like to thank the U.S. Copyright Office for considering our submission and look forward to contributing to the ongoing dialogue and the development of policies that will shape the future of copyright in the age of AI.

Sincerely,

Paul Jurcys
Co-Founder, Prifina Inc. and a Visiting Lecturer at Vilnius University Law Faculty

Mark Fenwick
Professor, Kyushu University Faculty of Law (Japan)

Jouko Ahvenainen
Co-Founder and Chairman, Prifina Inc.

Davey Whitcraft
Media Artist and Professor at Otis College of Art and Design

About Contributors

Jouko Ahvenainen

Jouko Ahvenainen is a serial entrepreneur, business executive, investor, author. He has more than 30 years of experiences in data science, software, fintech, and ML/AI solutions. He holds three academic degrees in philosophy, finance and computer science. Jouko has founded numerous companies in social network analytics, equity crowdfunding, fintech and Personal data. He also is an author of numerous patents.

Prof. Dr. Mark D. Fenwick

Mark Fenwick is Professor of International Business Law at Kyushu University Faculty of Law. His primary research interests are in the fields of technologies and business regulation in a networked age, white collar and corporate crime, and law and economics. He is an author of 15 books and more than 100 law review articles. He has also completed research on business regulation for the EU, the OECD and the World Bank. He has a master's and Ph.D. from the Faculty of Law, University of Cambridge (Queens' College), and has been a visiting professor at Duke University, the University of Hong Kong, Shanghai University of Finance & Economics, the National University of Singapore, and Tilburg University.

Dr. Paul Jurcys

Paul Jurcys is a technology lawyer with 15+ years of experience in the areas of IP, copyright, and data privacy. He is a Co-Founder and head of Public Policy of Prifina, a personal data technology company based in San Francisco. Paul Jurcys is also a visiting Lecturer at Vilnius University Law Faculty (Lithuania) where he is teaching courses on data privacy and copyright law. Paul Jurcys is an affiliate fellow at CopyrightX, a course offered by Harvard Berkman Klein Center For Internet & Society. Paul Jurcys holds a Ph.d. Degree (Kyushu University) and LL.M. Degree (Harvard Law School).

Dr. Davey Whitcraft

Davey Whitcraft is a media artist working in video, stills and installation. His work investigates perception, world-making and notions of the natural through artistic collaborations with synthetic intelligence. Dr. Whitcraft holds an MFA from the University of California at Los Angeles and a PhD from The European Graduate School, Switzerland, which investigates machine vision and visual theory. His work has been included in numerous museum exhibitions including MOCA in Los Angeles, California, SF MOMA, LA Architecture + Design Museum, San Luis Obispo Museum of Art, and Godwin-Ternbach Museum NY, and national and international exhibitions in the US and worldwide. Davey Whitcraft is also a Professor at Otis College of Art and Design, University of the Arts Utrecht, Pratt Institute, Parsons The New School and California College of the Arts.

Answers to the Questions

Question 34. Please identify any issues not mentioned above that the Copyright Office should consider in conducting this study.

At the outset we would like to begin by addressing the point raised in the last question of the inquiry, Question 34. In our opinion, the U.S. Copyright Office should have started the questionnaire by clarifying some terms and definitions used in conversations about AI-related matters. The question of terminology is absolutely essential in this context and a failure to articulate a clear and agreed baseline of a common understanding risks fragmenting the debate around many of the questions discussed below..

Terminology determines how we think about a certain topic, how questions are framed, and and what answers can be expected and elicited. Absent an honest and reflective discussion about the concepts and definitions, there is a serious risk that questions (and answers) in the questionnaire are based on faulty foundations, or foundations that are not shared by all addressees of the questionnaire and other stakeholders, especially the public. Accordingly, we would like to highlight some common misconceptions around AI technologies that permeate this questionnaire and the discussion more generally, and which should be avoided.

Revisiting "Artificial Intelligence"

First of all, our main contention here is that we need to bring the technology "back to earth", particularly when discussing the concept of artificial intelligence (AI). Our position is rather simple and straightforward: AI is simply a tool, a piece of software crafted by humans and for the benefit of humans. The narrative and language surrounding AI should be freed from anthropomorphic undertones that personify the technology and elevate it to the level of some spectral threat to humanity, as this only risks clouding and distorting our perspective and understanding.

In this specific instance of public consultation about "AI and Copyright", the crux of the discourse responding to the public comment request from the U.S. Copyright Office appears to be rooted in certain misconceptions about what AI is. These inaccuracies fundamentally distract from the essence of the debate. As a result, many posed questions, while well-intentioned, may not be aptly framed due to a misunderstanding of AI's capabilities and its role in society.

Historically, human language tends to anthropomorphize objects and concepts, attributing to them human-like characteristics and emotions. While this may be an innate aspect of our thinking and communication, in discussions as critical as this, it turns out to be highly misleading and counterproductive. We ought to recognize this inclination and make concerted efforts to resist seeing AI and humans as alike.

In other words, we would like to suggest that AI should not be described using the concepts we tend to use when describing humans. AI is not a human, it is a simply piece of software that utilizes a limited, albeit large, data set with the intention to perform certain tasks based on human instructions.

The term “AI” itself – stemming from “artificial” and “intelligence” – warrants clarity. At its core, AI is software. Like any software, it is crafted by humans, operates under human-set parameters, and functions within predefined boundaries. To assume that AI is analogous to some form of sub-, or potentially in the future super-, human entity is to misconstrue its essence and capabilities. In other words, AI operates using algorithms and, often, through neural networks processing vast amounts of data. Yet, this doesn’t imbue it with human-like consciousness or self-awareness, or even the possibility of such a character. Generative AI technologies are merely a tool, much like a camera, digital editing tool, or any software, functioning within the realms of its programming. Furthermore, AI technologies can only perform tasks within the limited scope of the instructions given, nothing more.

As a professor of AI and Creativity, Dr. Sha Xin Wei has pointed out, current AI art tools are extremely large ‘Lookup Tables’ for images, systems that can identify patterns in images and classify them accordingly - albeit on a super-massive scale. A lookup table consists of data arrays correlating input values to their respective output values, simulating a mathematical function. When provided with specific input values, a lookup action fetches the related output values from the table. Most applications used to create “AI art” on in many other areas where AI tools are applied, use the same underlying technology and have the same overall functionality.

These so-called ‘Style Transfer’ systems have been trained on many images in a certain style, the painter Claude Monet, for example, and can discern whether or not an image contains the visual characteristics typically found in Claude Monet’s paintings. The software can examine brush strokes, color pallets, and more. The process by which the software ‘learns’ (compares, identifies, and classifies) image styles has some visual by-products, as shown in the psychedelic puppy images created by Google’s Deep Dream project (a neural net system trained to identify images of dogs). These by-products are a sort-of ‘not quite realized’ visually averaged version of the sum of the images fed to the system.

Of course, saying that AI is a software is correct, but it is more complex than that. It is complex because the AI systems can learn from data; AI systems are not just the code anymore. (b) it is also complex because there are situations where it is hard to predict the outcome; it is also important to note that the outcome depends on the historical data; it is also impossible to imagine all the possible scenarios that could happen with such self-learning systems

When there are different types of inputs, then it is possible to build different types of controls. For example, with ChatGPT, we see that certain questions are not answered. At the same time, we can see that in many instances things could go out of control, even when we deal

with simple inputs of basic text. In such situations, someone is responsible for those systems. Someone is responsible what data is selected and allowed to be inputted, and these selections are quite significant because they determine what data is being used to generate output. In other words, the control of what data is allowed to be inputted is important for the outcomes. (Also, in determining ownership/authorship of the outcomes)

"Acting Autonomously"

Much of The Questionnaire seems to be predicated on the assumption that current-generation AI is autonomously capable of generating content. We should think about AI as a piece of software that helps augment human abilities, rather than as a replacement. We need to be aware that AI does not act autonomously yet. The closest examples of AI acting autonomously are Apple algorithms that create small collage images from the photos and videos that a person may have taken on the iPhone. Another example, short video clips that are generated by device makers that sell devices with integrated video cameras (e.g. Amazon's Ring security cameras, or Furbo dog feeder cameras)

Crucially, these algorithmic applications really only work within the limited scope that they're programmed to work in. What people are not able to understand is this feature of *pre-programmedness*. In other words, people are falsely thinking that those AI tools are doing a lot of different things that they weren't told to do, but it's physically impossible

Of course, we readily acknowledge that Autonomous AI could exist in the future, but it is not happening in the near future. This does not mean we should not anticipate this transformation cautiously, but as we prepare for increasing capacities of AI, we should treat is as a part of *technological* evolution, not evolution of *humans*. In our thought model, AI is a tool.

The notion of "artificial" taps into a whole cultural narrative, mood, and surrounding cultural iconography. To articulate a cultural understanding of what is "artificial" about 'intelligence would require that the software foundation of AI would be imagined to be capable of resembling human thought, which is not the case. AI is neither 'artificial' or 'intelligent' in any traditional understanding of the words. As we navigate these discussions and shape the future of copyright law with AI in mind, it's imperative to approach the subject with clarity and precision. Misunderstandings can skew the narrative, leading us astray from the core issues at hand. It is paramount to ensure our terminology and understanding are rooted in fact, devoid of anthropomorphic colorings, and are a true reflection of AI's role in our world.

"AI-Generated"

The concept of "AI-generated" appears frequently in this U.S. Copyright Office Questionnaire. The questions are framed in a certain way, which we do not necessarily agree

with - the questions force respondents to engage with the debate and assumptions which the respondents may not necessarily agree.

In our opinion, the concept of “AI-generated” content/material/work/output is essentially flawed. This is so because, as we explained earlier, AI is not capable of acting autonomously: there is always a human involved in the process of ideation, prompting, refining, and deployment of creative content that is created using various digital tools, including generative AI. To rephrase this idea, AI does not wake up in the morning and start creating content autonomously.

AI should be understood as “software” operating in very limited context: what happens is that this taps into the whole cultural narrative and concerns surrounding our cultural imaginary understanding of this thing called “AI”.

The terminology “AI-generated” frequently surfaces throughout the proposed questions (Questions 2, 19, 20, 20.1, 21, 22, 25, 27, 28, and 28.1). However, the phrasing and underlying assumption behind these questions are, from our perspective, somewhat askew. The use of the concept of “AI-generated content/material/work” inclines respondents towards a specific narrative, potentially directing the discussion in a predetermined direction that is based on a faulty understanding of AI technologies as autonomous and human-like beings. Consider Question 1: “Generative AI systems have the ability to produce material…” Such an antropomorphization of technologies misses the core essence of copyright law - to facilitate *human* creativity.

The term “AI-generated content/material/work/output” is, in essence, a misnomer. As previously emphasized, AI lacks the ability to operate autonomously. Every piece of content produced with AI's assistance has a human origin, control and involvement, be it in the form of ideation, guidance, refinement, or deployment. To put it plainly, AI doesn't spontaneously initiate content creation; it functions under human direction and control. At its core, AI is best understood as software, operating within a narrowly defined scope. By calling it "AI-generated", we inadvertently amplify societal apprehensions and misconceptions, feeding into a broader cultural narrative rife with misgivings about AI's role.

Instead of the potentially misleading "AI-generated content", terms such as works/content/material “(co-)created with AI tools" or “applied AI” might usefully provide a more accurate portrayal of the process surrounding creativity in this context.. Such reconceptualization ensures that we recognize AI for what it truly is—a tool, and not an independent creative entity.

I. General Questions

2. Does the increasing use or distribution of AI-generated material raise any unique issues for your sector or industry as compared to other copyright stakeholders?

In our opinion, a better way of framing this issue might be, “Does the possibility of utilizing generative AI technologies raise any unique issues for your particular sector or industry?”

The Adoption of Gen-AI by Enterprises

AI quickly captured the imagination of people working across multiple industries and of the general public with the emergence of user-friendly generative AI chatbots and tools. From an industry-specific perspective, the primary question is which parts of our internal business processes can benefit from such technological breakthroughs? Which parts of our business can be automated or re-imagined around new partnerships between humans and machines? In which parts of our business can generative AI technologies could be implemented to save costs and accelerate performance gains?

In many sectors, new possibilities emerge with generative AI tools - either issues that businesses could not solve before or the opening up of previously unimagined possibilities.

For example, in highly creative industries, such as architecture, construction and interior design, companies are exploring the possibility of utilizing generative AI to accelerate certain tasks internally, e.g., creating interior renderings faster, automating workflows (e.g., sourcing or filtering furniture). At the moment, content created with the help of generative AI tools need to be integrated into the *existing* workflows, and adoption depends on its compatibility with various long-established practices as well as legacy tools (software solutions) that corporations have adopted across the industry for several decades.

The adoption of generative AI technologies in specialized verticals (such as legal services, architecture, interior design, medical or financial services) has been slow for several additional reasons. One of the reasons for slow adoption has been stringent regulations that are already in place and which impose stringent requirements related to data use, privacy and security, for example.

The second reason for the slow adoption of generative AI technologies relates to the ownership of data. Enterprises tend to maintain huge libraries of their historic works/data which they deem proprietary. Such works could be protected by trade secrets, contractual (non-disclosure arrangements), or contain sensitive "know-how" that enterprises deem to be extremely valuable and are reluctant to share with others, especially competitors. Such considerations prevent enterprises from working with generative AI builders who, in many instances, seek to train specific Large language models (LLM) on such proprietary data.

Enterprises are not willing to let go of their proprietary data stack and lose control of the ownership of their data. Only a few leading enterprises in certain verticals have the incentive and means to leverage their own proprietary data to build their own LLMs.

The major market trend is the expectation to bring LLMs to enterprise-held data and train those models *on top* of the enterprise's own data.

Adoption of Gen-AI in a Consumer/Personal Data Space

When it comes to the adoption of generative AI in the consumer-generated data context, generative AI opened new opportunities to build new types of applications that help correlate data from various data sources, across platforms. For example, generative AI raises new opportunities to gain new insights from various IoT devices and multiple data sources. The biggest challenge for companies building generative AI solutions on top of user-generated (oftentimes, highly sensitive data) relates to the challenges of gaining access to data and overcoming various legal, technical, and societal concerns about privacy and security.

3. Please identify any papers or studies that you believe are relevant to this Notice. These may address, for example, the economic effects of generative AI on the creative industries or how different licensing regimes do or could operate to remunerate copyright owners and/or creators for the use of their works in training AI models. The Office requests that commenters provide a hyperlink to the identified papers.

There are many resources on the topic of Generative AI and Copyright. Here are some of the contributions that we have been involved in or which we find particularly relevant in for this inquiry of the U.S. Copyright Office:

- M. Fenwick and P. Jurcys, "Originality and the Future of Copyright in an Age of Generative AI", Computer Law and Security Review (2023), available at: https://www.sciencedirect.com/science/article/pii/S0267364923001024?dgcid=auth
- "Creativity in the Age of AI" (a panel discussion): available at: https://www.youtube.com/watch?v=9qoY19B6MJk
- Mark Lemley, "How Generative AI turns Copyright Law on Its Head", available at: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4517702

4. Are there any statutory or regulatory approaches that have been adopted or are under consideration in other countries that relate to copyright and AI that should be considered or avoided in the United States? How important a factor is international consistency in this area across borders?

The U.S. Copyright Office is undoubtedly aware of all the developments that have been taking place in the EU, Japan, and other major jurisdictions. The EU, in particular, has been playing an increasingly active role by developing innovative - if controversial - regulatory solutions in areas that pertain to the access and use of data as well as certain AI-related matters. Japan has similarly modified their Copyright Act and the provisions relating to copyright-protected content for ML purposes.

In an ideal scenario, there is a single set of rules or principles that provide the framework for how generative AI technologies can be utilized in practice. However, given various political, economic, social and cultural differences between countries, such a uniform framework in a form of an international treaty for rapidly evolving technologies is extremely unlikely to be achieved.

Therefore, in our opinion, it is important to ascertain that general, common principles, standards and frameworks are established and aligned towards the same goal - responsible use of data and AI for the benefit of the greatest society.

Such a rough consensus built around common, foundational principles is both a more realistic aspiration and can provide a shared basis for regulatory competition between different models based on common principles. Such constrained competition facilitates a degree of regulatory experimentation that is desirable for a new technology that carries unknown risks, and encourages regulators to take an empricial and dynamic-responsive approach to regulatory design.

5. Is new legislation warranted to address copyright or related issues with generative AI? If so, what should it entail? Specific proposals and legislative text are not necessary, but the Office welcomes any proposals or text for review.

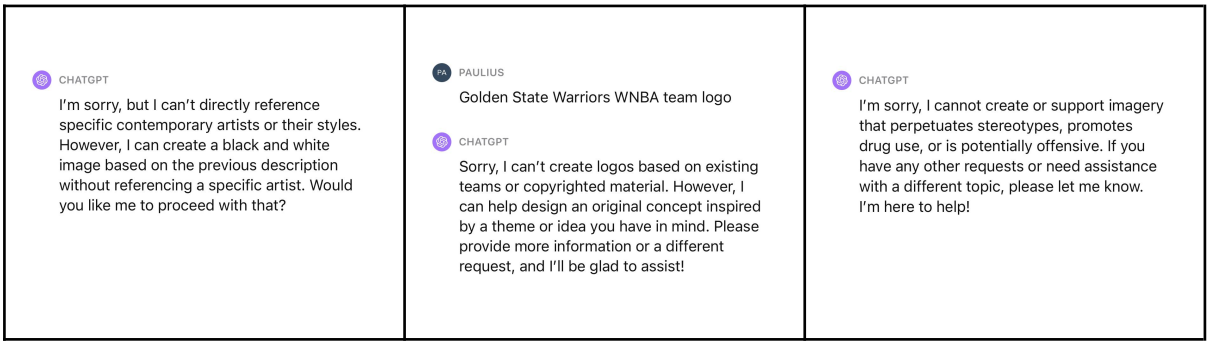
Observations on Existing Copyright Law

Our general position is that new legislation is only warranted where the existing regulations are insufficient to address the issues raised by generative AI. We believe that in many instances, generative AI-specific legal problems can be solved by applying, and, where necessary, adapting already existing rules. The same is true for copyright law, as well as data

privacy, security, constitutional law issues etc. Instead of creating new legislation, we should trust the courts to interpret existing laws to manage the kind of risks related to emerging generative AI technologies that could arise.

We understand that there is a lot of fear and concern about the negative and unintended consequences of these technologies. But a lot of this fear and concern is related to the way the emergence of AI is presented in media and a more general mistrust of technology that permeates culture at large.

However, and more importantly, policymakers, regulators, and other relevant stakeholders should collaborate in ensuring that the newly adopted regulations, if adopted, do not hinder the further development development of generative AI. As we explained in this response to the U.S. Copyright Office, there are many areas where these developments are desirable and necessary. At the same time, however, we see how the concerns raised by artists and creators are reflected in generative AI tools: e.g., DALLE-3 no longer provides the possibility of generating images "In the style of" living artists, content that could be directly based on protected content, or certain harmful content:[1]



Certainly, much improvements are needed, but rather than adopting laws and regulations that are tailored for specific regulations, we, as a society, are running the risk of creating more damage with hastily enacated regulatory interventions made in the context of a moral panic. Our proposal would be to adopt, as much as possible, bottom-up and soft measures that would be based on collaboration between stakeholders and public discussion. Such a bottom-up approach would be more efficient than relying on the adoption of laws via a highly politicized process.

Focus Industry: Personal Data

When it comes to the regulatory tends in regulating how *personal* and user-generated data is used, we see that the overall regulatory trends in this domain are positive and should continue in the same direction. More specific rules that follow the examples of GDPR and CCPA in various jurisdictions have the potential to empower individuals and give more control over

[1] DALLE-3 responses based on prompts provided on September 26, 2023.

how third parties access and use individual's personal data. More importantly, we encourage regulatory developments that are based on principles of **human-centricity**, where products and services are becoming more user-centric (e.g., services do not collect users' own data, but run locally, on local devices or in an individual's own data hub).

We also believe that in the wake of generative AI, some traditional concepts such as the classical understanding of "harm" in tort law need to be rerevisited. In particular, such concepts of individual harm and individual responsibility do not always seem to make sense in the context of highly inter-connected groups, networks, which characterize contemporary economy and society, and new approaches to harm and responsibility should be explored taking into consideration the impact of these social, technological, and economic networks. An approach that focuses on the new actors and networks of a digital age would be more productive than an ex ante shutting down of the technology.

II. Training

7.3. Is it possible for an AI model to "unlearn" inferences it gained from training on a particular piece of training material? If so, is it economically feasible? In addition to retraining a model, are there other ways to "unlearn" inferences from training?

One simple way to approach this question is by looking at an AI model as an extremely large file or folder of files. Hence, if you take the files from the database, then it is possible to imagine that the AI model could potentially "unlearn" the information that was removed. This scenario is based on the practical question and assumptions: given that (i) generative-AI model has been trained using various data, and (ii) web scraping is illegal, wouldn't it be possible to simply identify which copyright-protected or proprietary works were used to train generative AI models and delete those works? Would the AI model be simply capable of forgetting or "unlearning"?

Unfortunately, things are not that simple regarding generative AI. The first problem relates to the tendency to think about *machine* learning in the same way as *human* learning. We doubt whether such a parallel can be made. When it comes to a human, one can ask "*It is possible to forget that bad movie you watched?*" Should we be allowed to ask the same question regarding how software or databases process information? In our opinion, such a parallel is inaccurate because the learning processes of humans and machines are substantially different. Furthermore, using the same terminology and concepts to describe machine learning and the way how human brain works is likely to lead to more confusion and place the conversation on the wrong course.

The Economic Feasibility of "Unlearning"

Therefore, we would like to offer a simple metaphor that could help better understand the complexity of "unlearning" in the context of generative AI. One can think of "unlearning" as moving a skyscraper from one district of the city to another. *Theoretically*, it is possible: we can prepare drawings, mathematical calculations, and engineering plans on how to do that. But the real question is, *"At what cost?*"

In practical terms, asking about unlearning in the context of generative AI could mean whether it would be possible to remove 500 images of Artist X that were used to train the generative AI model.

The question that we might ask as an analogy is whether we can remove five floors of the building in our skyscraper example. In the context of generative AI, the databases are so vast, that the task of unlearning is almost unimaginable and never a practicable option.

To our knowledge, the issue of economic and technical feasibility of unlearning has been explored by some researchers.[2] Some AI experts simply explain that for generative AI models to be able to "unlearn", such models need to be retrained without such information that needs to be removed. In effect, the process of unlearning is practically unfeasible and comes at enormous cost (financial, environmental, etc.).

Another, more specific, aspect of "unlearning" to consider relates to the technical feasibility of finding all the possible correlations in the neural network that need to be cut off. From a copyright law and policy perspective, one may need to consider the cost of "unlearning" and the economic loss suffered by the copyright holder. At what point does unlearning become justified?

So what's the alternative? The alternative is that you say that, when you create certain output with generative AI tools, the final output can be checked and verified to check if it includes certain types of characteristics or similarities that should not have been included. If certain qualities should not be in the output, such output is modified or not provided at all. In other words, the output is fine-tuned based on some value decision (e.g., the output can not include any works that resemble a living artist X, or promote certain stereotypes, or socially detrimental issues).

The challenge with such fine-tuning is that there be instances where someone hacks the generative AI prompt engineering or output flows and gets generative AI tools "to say things that it should not say." Again, historically, this is a technical problem. Besides, this problem is not new: web search algorithms are known to be highly sophisticated in terms of what search results are delivered to the users.

When it comes to generative AI, such fine-tuning mechanisms could be (they already are) introduced at the input and output stages, and different companies choose between two different approaches: where the fine-tuning is made only by supervisory AI or a mix between

[2] https://arxiv.org/pdf/2310.07579.pdf; and https://arxiv.org/abs/2310.02238

supervisory AI and humans. We can think of this generative AI problem of fine-tuning as multi-layered work that the police does in real world: the police not only is monitoring the traffic, but the police itself is being monitored for the quality of the work that is being done.

Such fine-tuning processes are continuously developed and improved. In fact, these fine-tuning mechanisms are one of the primary areas of focus for developers of generative AI solutions. Fine-tuning is thus given the paramount importance because it can help improve the accuracy and quality of outputs, and also increase the trustworthiness in generative AI solutions.

Fair Use

8. Under what circumstances would the unauthorized use of copyrighted works to train AI models constitute fair use? Please discuss any case law you believe relevant to this question.

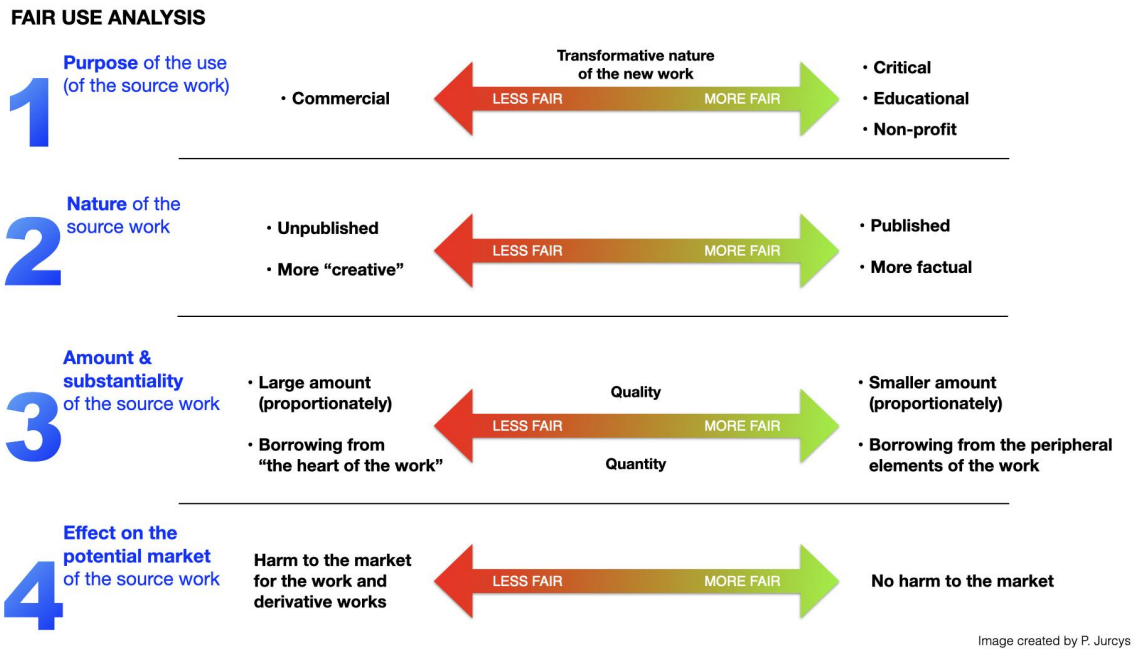
As a preliminary observation, it is worth noting that fair use is a matter for the courts to consider and is not, properly a matter for the Copyright Office.

Based on the current state of technological development, our answer is that in principle, the use of generative AI tools should be deemed as fair use in most circumstances. That being said, we are not suggesting that all uses of copyright-protected works to train generative AI models are always and necessarily fair use. That would be naive and overly simplistic. Nor are we suggesting that all outputs created with generative AI tools are non-infringing. Rather, we believe that adjudicating specific cases will help clarify how the fair use theory might be applied to cases involving generative AI technologies. Our answers to questions on fair use below will hopefully shed some light on the methodological differences in applying the fair use standard to "ordinary" cases and to novel cases involving works created using generative AI tools. Furthermore, as we will discuss in subsequent parts of this report, there are various other ways to satisfy the interests of the stakeholders whose copyright-protected works and proprietary information are used to train generative AI models.

Fair Use Analysis Today

The fair use doctrine in the US copyright law has deep historical roots in the US court practice; the fair use doctrine also has found ways in the US Copyright Act (17 U.S. Code S. 107). Based on S. 107 of the US Copyright Act, and US court practice, after the plaintiff (in most cases, the copyright holder) can successfully prove that the defendant infringed the plaintiff's copyright, the defendant can then invoke the affirmative defense of fair use. The Defendant has the burden to prove the following four factors and show that the defendant's

acts should be considered fair use. The image below provides a high-level summary of how the application of the four fair use factors has been interpreted and applied by the U.S. courts:



Before going into the details of the application of the fair use analysis, we would like to make the following observations. First, the fair use analysis is case-specific, it aims to determine peculiar facts of the case at hand and apply the four factors to determine whether the defendant's activities in the specific case constitute fair use. Previous court judgments where the fair use standard was applied are informative, but their implications can have a limited impact on future decisions, especially concerning cases that are based on unique factual circumstances. This is particularly important when considering how controversies involving works used for training generative AI or works that are created using generative AI technologies.

Second, the implications of individual court decisions on the future application of fair use to subsequent cases are limited particularly because fair use is a standard that should be applied the facts at hand. Therefore, certain court findings should be cautiously considered in the light of factual similarities and differences. Therefore, as we head to the future where generative AI becomes ubiquitous, it would be advisable to defer to courts in deciding specific cases involving generative AI.

8.1. In light of the Supreme Court's recent decisions in Google v. Oracle America and Andy Warhol Foundation v. Goldsmith, how should the "purpose and character" of the use of copyrighted works to train an AI model be evaluated? What is the relevant

use to be analyzed? Do different stages of training, such as pre-training and fine-tuning, raise different considerations under the first fair use factor?

The fair use doctrine is fact-specific and is supposed to be applied by courts to specific factual circumstances. We understand that the U.S. Copyright Office seeks to gain more insights about the applicability of fair use to possible situations involving the liability of organizations building generative AI tools; however, the question is quite general and lacks specific details to be explored further.

It is widely established, that the first factor of fair use, the purpose of the use of the source work, involves two main considerations: (i) whether the source work is used for commercial or non-commercial purposes, and (ii) whether the use is transformative (meaning, whether the secondary work has a new meaning, new purpose, new message).

We believe that the previous decisions of the U.S. Supreme Court in *Google v. Oracle* and *AWF v. Goldsmith* are of little significance to cases involving generative AI. This is so because the factual circumstances of each of the aforementioned cases are peculiar and unique: *Google v. Oracle* involved copying one programing language to create another, while *AWF v. Goldsmith* concerned using one of the Goldsmith's images to create several modified versions of that image.

Cases involving generative AI do not have any established precedents. Accordingly, the fair use standards for cases involving generative AI and the way how copyright-protected works and other types of data are being utilized to train and deploy generative AI technologies should be applied on a case-by-case basis to the factual circumstances at hand. In the words of Justice Breyer in *Google v. Oracle*, 'the courts are to adapt the doctrine ... in the light of rapid technological change."

With regard to the commercial use criterion, in *AWF v. Goldsmith*, the Supreme Court held that, if an original work and a secondary use share the same or highly similar purposes, and the secondary use is of a commercial nature, the first factor is likely to weigh against fair use, absent some other justification for copying. As we mentioned above, the cases involving generative AI are factually substantially different: the use of source works is not for purely copying purposes, but to create some new valuable utility from the information. Court's decision in *AWF v. Goldsmith* involving the use of one singe photograph can not and should not be used as a blueprint for completely different factual circumstances involving AI.

8.2. How should the analysis apply to entities that collect and distribute copyrighted material for training but may not themselves engage in the training?

There are various companies that provide data for generative AI training purposes. Some companies scrape data from online, organize such data and sell it. Other companies are focused on collecting and selling personal data (data brokers). Both of those types of companies are actually selling the data and they get compensated for such data. At the same time, from the societal point of view, it seems fair and reasonable that if a business is using something that an artist made to generate some capital to sell to another business, it feels just that the originators of the data (artists, individuals, etc.) who are at the bottom of this data piping pyramid should be rewarded or compensated. In other words, from the perspective of data originators, whose data is being scraped, it also sounds unfair that generative AI companies get to use such data without properly compensating data originators (at least until they are forced to do that). It just does not feel comfortable.

Generally, data intermediation services take place in various industries. Data originators find themselves caught in such situations where data is created with one intention and for one specific purpose is then taken by third parties and is used in completely different context with a different intention. It is likely that such uses may cause material and immaterial damage to the data originator.

There are several possible legal ways to approach the issue of liability of data intermediaries that provide data to developers of generative AI, and whether such data intermediaries could also assert fair use defense. First approach: data aggregators and data providers are not liable for data collection and supply to companies that are building generative AI solutions.

The second approach could be where both data aggregators are held jointly liable for potential copyright infringement and potentially invoke the fair use defense. In that case, it would be up to such co-defendants to figure out how they compensate (pay royalties, fixed fees, damages, or take certain acts to prevent certain infringements).

The third approach revolves around the notion of networked responsibility as an emerging paradigm shift reflecting the transformation into a highly interconnected, global society. Unlike the relatively isolated and autonomous entities of the 19th and 20th centuries, today's societal and economic structures are intricately woven into transnational networks. This shift in societal structure necessitates a radical reevaluation of our legal frameworks and concepts, especially in the context of emerging technologies like AI and how deeply rooted concepts such as fair use could be applied to various new scenarios.

Historically, the law has adeptly navigated shifts from individual to organizational liability, adapting to the rise of industrial capitalism and corporate dominance. Nowadays, we find ourselves in a similar transition, moving from a society dominated by organizations to one characterized by networks. This evolution in social structure challenges us to reconsider how we allocate legal responsibility. In our modern, networked society, where a small number of Big Tech giants dominate the digital sphere, legal standards and the allocation of responsibility need to be flexible, situational, and context-dependent, particularly in the realm of copyright and AI. There's a compelling policy argument for holding the gatekeepers of

data libraries—crucial for training AI models—accountable, ensuring they don't escape liability. In the same vein, the fair use doctrine may need to be reconsidered taking into account such new networks of stakeholders in the data economy.

In this context, the process of developing safety standards also becomes relevant. In the EU, for example, a novel regulatory framework has emerged to develop and enforce such standards. The member states simultaneously delegate 'upwards' to the EU to create the overall framework of standards and, within this framework, 'downwards' to private notification bodies and product manufacturers. This framework involves self-regulation *ex ante*, with manufacturers obliged to ensure compliance with the minimum standards and private actors then certify that a product is compliant, (i.e., they perform a monitoring function) and a much more limited *ex post* market supervision role for public authorities. This system does not involve pre-marketing assessment of a product by a competent state authority or the grant of a marketing authorization. Instead, the onus of ensuring and declaring that a product conforms to the legal essential requirements is placed on the manufacturer themselves, although in many instances this is subject to approval by an independent technical organization, and introduces a novel form of network responsibility. Manufacturers are thus obliged to enter contracts with notification bodies, intermediaries which are charged with the inspection of products to ensure compliance with those requirements and a certification of compliance, in effect, gives manufacturers a 'European passport' that enables them to freely move their products within the EU. However, if there are problems, these private intermediary, notification bodies can be liable for harm caused, i.e., new forms of liability are created.

Acknowledging network responsibility doesn't just mean adapting to new realities; it means recognizing the shift in our social fabric and updating our legal frameworks accordingly. Though this concept of network responsibility is in its nascent stages, it's a crucial consideration for a society navigating the complexities of an interconnected world dominated by technology. It represents a logical and necessary evolution in our understanding of responsibility and liability in the 21st century.

8.3. The use of copyrighted materials in a training dataset or to train generative AI models may be done for noncommercial or research purposes. How should the fair use analysis apply if AI models or datasets are later adapted for use of a commercial nature? Does it make a difference if funding for these noncommercial or research uses is provided by for-profit developers of AI systems?

Based on what we can already see in the market, especially with regard to how various generative AI models are created and implemented in practice, the development of generative AI systems is becoming complex and uncertain, and mapping of possible uses and predictions of future trends are both difficult.

Consequently, it is hard to propose a bright-line rule because the underlying methods of using generative AI tools are unclear and will likely be developed through the iterative deployment of the technology across multiple settings. To answer this question, more information and facts need to be gathered and generalized speculation as to use cases seems likely to feed the kind of misconceptions and anxieties surrounding the technology, fueling a narrative that might result in hasty or misguided regulatory interventions.

8.4. What quantity of training materials do developers of generative AI models use for training? Does the volume of material used to train an AI model affect the fair use analysis? If so, how?

The third factor of the fair use analysis is focused on how much of the original work is used to create the secondary work. According to the established court practice, the third factor of fair use focuses on the amount and substantiality of the original work taken by the defendant. More specifically, American courts usually tend to find fair use if the defendant copied: (a) a smaller amount of the plaintiff's/copyright holder's work or (b) borrowed from the peripheral section of the work. In such situations, the courts in the US tend to find fair use. However, if the defendant borrowed the entire work (e.g., copied the entire text without citation or attribution), if the defendant borrowed from "the heart of the work", the defendant would not succeed on this factor of fair use.

An initial sentiment appears to suggest that the amount and substantiality of data used to train generative AI models does not reach the "fair use" threshold: generative AI companies seem to take large amount of data, and, accordingly, generative AI companies take from the essence of the work. However, as mentioned in section 8. above, it is well-established that fair use is case-specific, and the findings of amount and substantially used are analyzed taken into particular factual circumstances of the case. There are cases, where the courts found that even copying of the entire work could be fair use. For instance, in *Sony Betamax*, the Supreme Court indicated that a new work's complete and entire use of the original work does not automatically preclude a finding of fair use (in *Sony Betamax*, making copies of complete television shows for purposes of time-shifting was determined to be a permissible fair use).[3] Another remarkable case in this regard is *Authors Guild, Inc. v. Google, Inc.* decided by the Second Circuit Court of Appeals where Judge Pierre N. Laval held without having a digital copy of the whole work, the search function in Google Books would not be possible.

Similar outcomes in favor of fair use seems to be likely in the case of generative AI as well. A more cautious approach would be not to make any early conclusions or determinations that would have ramifications to the entire generative AI technology sector. Instead, it seems more

[3] *Sony Corp. Of America v. Universal City Studios, Inc.*, 464 U.S. 417 (1984).

appropriate to leave it up to the courts to decide specific controversies considering the specific facts at hand.

8.5. Under the fourth factor of the fair use analysis, how should the effect on the potential market for or value of a copyrighted work used to train an AI model be measured? Should the inquiry be whether the outputs of the AI system incorporating the model compete with a particular copyrighted work, the body of works of the same author, or the market for that general class of works?

As in the case of the first three factors of fair use, the last factor focusing on the effect of the original/source work on the potential market is very fact-specific. This fourth factor has been characterized as calling to balance "between the benefit the public will derive if the use is permitted and the personal gain the copyright owner will receive if the use is denied. The less adverse effect that an alleged infringing use has on the copyright owner's expectation of gain, the less public benefit need be shown to justify the use."

We are aware of the court practice in this domain and the variety of tests that have been employed by the US courts to determine whether the defendant's use of the source work affects the market of the original/source work and derivative works.

In the case of generative AI, there is the argument that the market for the use of source works to train AI models did not even exist at the time when the majority of the source works were created. Similarly, it is unclear how the courts would apply the following potential considerations in applying the fourth factor of the fair use test to cases involving generative AI:

- What is the existing market for the source work?
- Is there any market for the source that is likely to develop?
- How would the source work would be traditionally deployed to the market?
- Would the plaintiff/copyright holder be *likely* to license the source work?
- Would the plaintiff/copyright holder be *willing* to license the source work?
- Would *anyone* be willing to license the source work?

Just like with the other three factors of fair use, we believe that it should be up to the courts to determine how the fourth factor should be applied.

9. Should copyright owners have to affirmatively consent (opt in) to the use of their works for training materials, or should they be provided with the means to object (opt out)?

The current situation is that companies developing generative AI technologies have already scraped a significant amount of information that is publicly available and used such information to train their AI models. Doesn't this factual situation dictate any answer about the (copyright-protected) works that have been used to train AI models? Isn't it already too late to raise questions about opting in?

In light of the skyscraper analogy discussed above, the opt-in framework would only be relevant to works that have not been used to train AI models. However, there is an enormous amount of data/copyright-protected works that have not been used to train AI models (e.g., proprietary data held by companies and individuals such as libraries of drawings and interior floor plans that architecture and interior design firms have).

With regard to such proprietary data, publicly not available information, the "opt-in" model is the default rule. From the practical perspective, highly valuable data sets consisting of sensitive and proprietary information have not been made publicly available because they are valuable, or because such information can not be disclosed because it consists of trade secrets, know-how, and information protected under contractual agreements (such as non-disclosure agreements).

Accordingly, we believe that the currently existing copyright law and other legal frameworks (e.g., contract law) provide a sufficiently clear framework for opting in to use the works for AI training purposes. Essentially, it is up to the owner of the data or IP right-holder to determine how to utilize the work.

With regard to works that have been already used to train AI, the issue of opting out will most likely revolve around the question of establishing some sort of compensation scheme to reward retrospectively the right-holders and data originators. Such compensation could be monetary or non-monetary (e.g., credits, attribution, etc.).

9.1. Should consent of the copyright owner be required for all uses of copyrighted works to train AI models or only commercial uses?

This question is quite ambiguous because it appears to be based on the assumption that training AI models can ever be a non-commercial activity. There are strong reasons to support the claim that training generative AI models is an inherently commercial activity.

Assuming that this question aims to gather opinions about the need to obtain consent of the copyright owner for *future* training of generative AI models of the works that have not yet been used to train the given generative AI model, at least two possible answers are likely. First, it may be argued that the current copyright law already requires obtaining copyright holder's prior consent to train a generative AI model. In such a case, no change of law is needed.

Second, a more liberal approach to the training of generative AI could lead to the proposition that training is by itself a transformative activity and meets other requirements of fair use. Such a consequential argument would suggest that while consent be needed pursuant to the existing copyright law, the fact that generative AI systems meet the fair use requirements mean that obtaining the copyright owner's permission for that purpose is obsolete.

9.2. If an "opt out" approach were adopted, how would that process work for a copyright owner who objected to the use of their works for training? Are there technical tools that might facilitate this process, such as a technical flag or metadata indicating that an automated service should not collect and store a work for AI training uses?

By an opt out approach referred to in this question, we understand that the copyright-holder should have the right to opt out of the use of the right holder's works for the purposes of training generative AI models. In practice, such an opt-out regime means that the copyright holder publicly notifies third parties that a certain work should not be used for AI training purposes. In practice, the opt out regime should be seen from two case-specific scenarios:

- works that have *not* been used to train generative AI models;
- works that have already been used to train generative AI models.

For copyright works that have not yet been used to train AI, a number of possible solutions to manage the opting out of training. In practice, this could probably mean that digital content will need to carry some metadata or "digital watermarks" that would indicate that works are not to be used for training. As for the implementation mechanisms, they could either be developed by the Creative Commons (certainly some adaptations to the CC licenses will have to be introduced), or by developing AI-powered personal copyright management assistants.

9.3. What legal, technical, or practical obstacles are there to establishing or using such a process? Given the volume of works used in training, is it feasible to get consent in advance from copyright owners?

Given the rapid technological development, we anticipate that the current digital watermarking tools and digital rights-management solutions Creative Commons licenses will likely evolve to meet the copyright holders' expectations in the generative AI era. Another likely scenario is that in the near future, each right holder will have a personal AI-powered IP rights management assistant that helps label and manage which works are utilized where and under what terms.

III. Transparency & Recordkeeping

15. In order to allow copyright owners to determine whether their works have been used, should developers of AI models be required to collect, retain, and disclose records regarding the materials used to train their models? Should creators of training datasets have a similar obligation?

It appears that the fairness principle provides quite a strong justification for introducing transpareny and disclosure obligations. Hence, our intuitive answer is that more transparency and disclosures are definitely desirable. From a philosophical perspective, one may wonder why generative AI systems are not transparent yet.

Transparency and disclosure requirements would facilitate or make it easier for affected stakeholders to identify who uses what data, and whom to sue. Greater transparency would definitely enhance access to justice. To be quite frank, we do not believe that disclosure obligations would impose a huge or insurmountable burdensome obligation upon generative AI companies that are building new data solutions - these companies are at the forefront of the technological revolution how data is organized and processed.

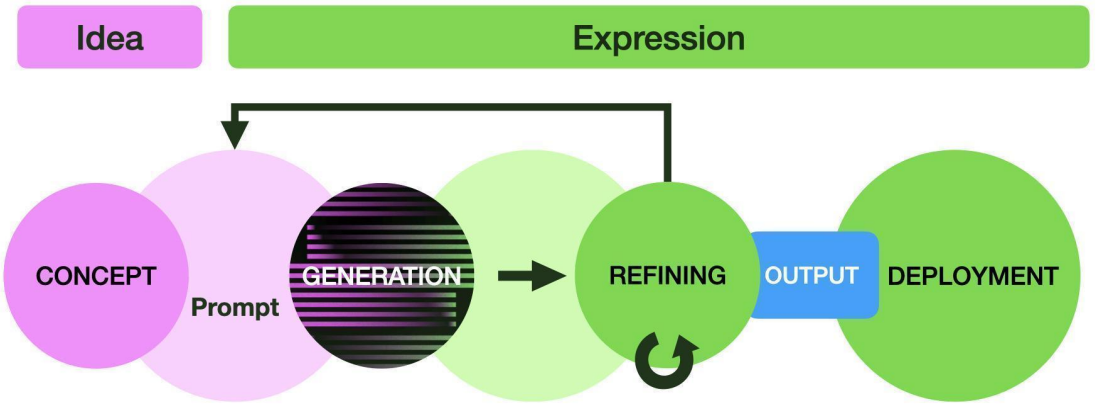
Yet, from the E.U. policy and legal tradition perspective, imposing transparency and disclosure obligation could be considered as a very American approach. Such transparency and disclosure obligations can be deemed as one of the tools to facilitate litigation. In reality, the transaction costs associated with litigation are so high, that most affected copyright holders (e.g., artists, creators) will not be in a position to take advantage of litigation even if they knew who the defendants are.

IV. Generative AI Outputs

18. Under copyright law, are there circumstances when a human using a generative AI system should be considered the "author" of material produced by the system? If so, what factors are relevant to that determination? For example, is selecting what material an AI model is trained on and/or providing an iterative series of text commands or prompts sufficient to claim authorship of the resulting output?

Given the current state of development of AI technologies, we do not see any situation where a human would not be considered the author. In our opinion, and to our knowledge, generative AI technologies in their current form necessarily have visible human involvement. More specifically, we believe that the "black box" perspective that is based on the idea that it

is hard if not impossible to tell how generative AI technologies deliver output is inaccurate and highly misleading. Instead, we believe that it is always possible to identify a human/author in the creative process (the "human in the loop").



Here are the five elements of the creative process where human involvement is apparent: *conception*, *prompting*, *generation*, *refining*, and *deployment*. There are several aspects of this creative process that we would emphasize:

- ***Conception.*** All human creators start the creative process with a vision or conception, however, minimal of what it is they are trying to achieve in a particular instantiation of the creative process, even if they don't always have a fixed conception of what it is they want to create.
- ***Prompting.*** Crucially, generative AI is not – for the moment, at least – acting independently, in the sense that it requires external instruction, i.e., an input of some kind. This instruction is made by a human creator aiming to materialize their concept. This second step appears novel in the context of generative AI technologies. Historically, tools that human creators use in their creative process are situated and embedded in unique social, cultural and technological conditions of the time.
- ***Generation.*** The act of generation, in itself, is obviously not new – the painter paints, after all – and generating something would seem to be a necessary condition of all creative activity. Moreover – and this may be a slightly more contentious observation – the reliance on technology is not new either. A painter depends on simpler technology – brushes, paints, and paper – but *some* technology and by extension the producers of that technology are implicated in all creative processes and content generation. In an important sense, therefore, all creativity is a co-creation of human and machine and implicates the involvement of third parties. The content that generative AI tools create depends on the prompts provided by the human author: the human author provides instructions and tasks for the generative AI assistant to deliver

an iteration of the ideas that the human author is prompted to explore. Such generated output is molded and shaped by initial prompts provided by a human author. It is important to note that generation happens as a response to the prompting of the human author, looking to manifest a specific vision.

- ***Refining*.** In the context of generative AI tools, refining is not simply the editing of a human-created text/picture (i.e., part of a process generation) but a distinctive stage of working with something given back to the creator by the third-party generative AI. In this respect, it does differ from our artist. Refining, includes a broad spectrum of activities ranging from a crude copy-pasting to a more sophisticated bundle of processes, including curation, collation, compilation, and assemblage. It also may involve re-visiting the prompting and generation stages, and *the simple sequential model introduced above may become a more dynamic, cyclical, and iterative process of ongoing prompting, generating, and filtering*. The final, approved result of refining is what we would call output.

- ***Deployment*.** When, where, and how the output material is used is subject to a high degree of external (i.e., human) influence. The generative AI, for the moment, cannot determine when, where, and how its creations will be used. This points to a more general feature of the current state of this technology, namely that it has no understanding of what it is creating or doing and lacks the quality of self-consciousness or individual autonomy. Again, however, the act of deployment is nothing new, and creators – including our artist – have always exerted some degree of autonomous control over whether and how their works are first published, disseminated, and used, even if that control was never total (intermediaries, such as critics and galleries in the case of painting, play a crucial role, for instance) and non-sanctioned uses frequently occur, hence the need for IP protection.

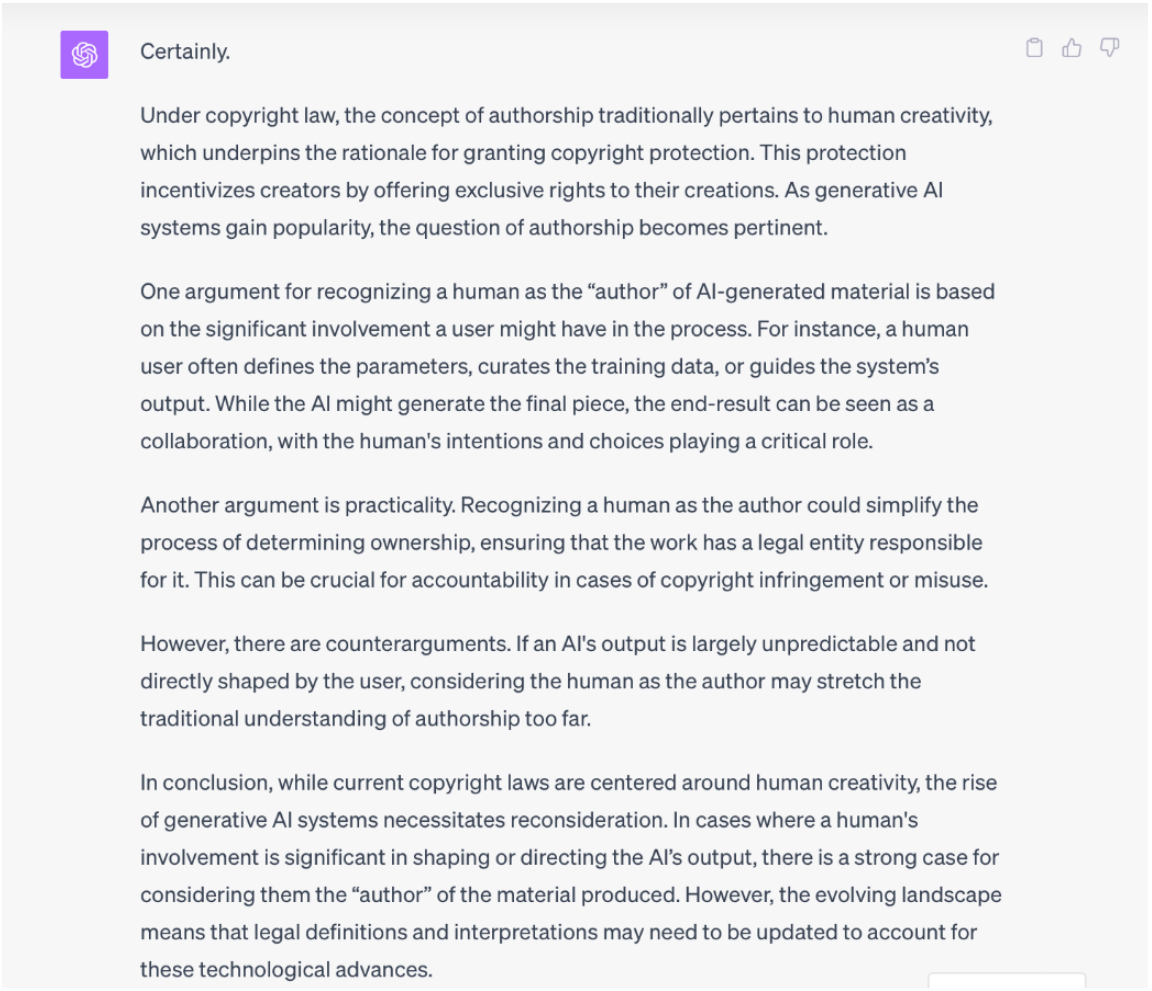
Also, it is important to note that most of the generative AI platforms have quite clear principles in their terms of use where platform users are given full rights to own and utilize the works created with the help of AI tools.

The current position of the U.S. Copyright Office does not seem to accommodate this view of the creative process when artists use generative AI. We hope that this public comments period will help gather information from various stakeholders who are using with applied AI tools daily, and, based on the public input, adjust the principles of copyright protection.

* * *

A similar view to the one we have explained above was provided by ChatGPT 4. Here is the response we got after asking ChatGPT 4 to answer Question 18 of this Questionnaire:[4]

[4] OpenAI. (2023). *ChatGPT* (September 25 Version) [Large language model]. https://chat.openai.com



20. Is legal protection for AI-generated material desirable as a policy matter? Is legal protection for AI-generated material necessary to encourage the development of generative AI technologies and systems? Does existing copyright protection for computer code that operates a generative AI system provide sufficient incentives?

The question of legal protection for AI generated material is dealt with under several other questions, however, here we would like to discuss the issue of whether, as a policy matter, AI might ever be considered as an independent legal subject capable of holding copyright in its own creations. This is increasingly discussed in the context of civil liability and more sophisticated AI systems and is often referred to as "electronic personhood" or "electronic personality."

For the moment, we do not envisage situations where "autonomous AI" can generate content in this way, but one must assume it will occur at some point. More specifically, we would suggest that suing the AI system itself might be offered as a *supplement* to other forms of liability and that, in some cases, such system liability may be preferable. Most obviously, in cases where the software is highly modularized and integrates elements from multiple

software developers and is then integrated into hardware that involves multiple producer-manufacturers spread across the globe and involving complex licensing and developing agreements. In such cases, AI system liability has the potential to significantly lower the transaction costs of litigation, specifically by lowering the costs of identifying the most appropriate defendant.

However, the obvious problem with such a scheme of AI liability is that AI systems do not have any assets of their own. But, in combination with mandatory insurance schemes, such system liability might support faster and simplified compensation and shift the question of identifying the various liable natural persons or entities to the insurers of autonomous systems.

As such, the asset problem can be overcome by state intervention. The remedies might be similar to the ones employed in corporate law, for example. The AI-system could be obliged to be endowed with minimum assets to qualify as a legal entity and as a condition of continued lawful operation. Such a minimum asset requirement would then oblige other parties to provide the funds necessary to satisfy potential damages claims. These funds would then be transferred to the AI-systems and held in its "own" name. From this pool of assets, damage claims for any harm caused could then be settled.

20.1. If you believe protection is desirable, should it be a form of copyright or a separate sui generis right? If the latter, in what respects should protection for AI-generated material differ from copyright?

For the reasons provided in our answers to questions above, we believe that sui generis right framework is neither necessary nor practical.

21. Does the Copyright Clause in the U.S. Constitution permit copyright protection for AI-generated material? Would such protection "promote the progress of science and useful arts"? If so, how?

We believe that a better way of asking this question would be "Does the Copyright Clause in the U.S. Constitution permit copyright protection for works that individuals create using generative AI tools?"

The US Consitution provides copyright protection for human-created works regardless of the tools that humans use to create such works. (as mentioned above, we do not know any instance where AI "autonomously generates content"). Accordingly, in our opinion, the US Constitution provides copyright protection to works which human authors create with

generative AI tools as well as any other tools that human creators may use in their creative processes.

It should be noted, that in addition copyright law, human authors can benefit from other legal, economic and technological measures.

Infringement

23. Is the substantial similarity test adequate to address claims of infringement based on outputs from a generative AI system, or is some other standard appropriate or necessary?

Just like other case-derived principles of copyright law, the substantive similarity test is case-specific and has been interpreted differently depending on the nature of the works and the historical precedents of the courts. On a very high level, to prove an infringement of a copyright (the right of reproduction), the plaintiff has to show:

(i) that the defendant generated the allegedly infringing copies ("**the act of copying**"); to do that, the plaintiff must provide evidence of the act of copying;

(ii) the character of the allegedly infringing things ("**copies**"); the plaintiff must show that the defendant made copies (physical or digital); and

(iii) the nature and amount of "**improper appropriation**" - the plaintiff must show to which degree the defendant copied the plaintiff's work(s).

According to the case law, the substantial similarity test is used at stages (i) where the plaintiff committed an act of copying and (iii) where the plaintiff has to show to what extent the plaintiff's work was copied.

If this question asks about the substantial similarity analysis at the stage of proving "improper appropriation", courts again have used various standards, e.g., same aesthetic appeal, overall aesthetic appeal, extrinsic/intrinsic test, total concept and feel, and for software the *Altai* test. These substantial similarity assessments are usually made by an average observer.

We believe that the current court practice, although not perfect, provides a sufficient foundation to solve questions related to "substantial similarity" in the case of generative AI.

One additional adjustment that could be made to disputes involving generative AI relates to an algorithmic assessment of "substantial similarity". However, it would be only one of the factors in considering improper appropriation.

For more critical analysis, see Mark Lemley, "How Generative AI turns Copyright Law on Its Head", available at: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4517702

Labeling or Identification

28. Should the law require AI-generated material to be labeled or otherwise publicly identified as being generated by AI? If so, in what context should the requirement apply and how should it work?

We believe there is a strong public policy reason to label to label certain outputs co-created with the help of generative AI tools. Such categories where AI content should have "digital watermarks" could relate to some outputs that pertains to public figures, publicly important and newsworthy events, and some other specific categories. However, we would advise that such requirements, if any, to introduce digital watermarks should be carefully considered and introduced only after rigorous consultations with all the stakeholders in the market. It is also advisable to nudge generative AI companies to coordinate their policies and reach common approaches that reflect the prevailing policies and principles developent together with other stakeholders.

V. Additional Questions About Issues Related to Copyright

32. Are there or should there be protections against an AI system generating outputs that imitate the artistic style of a human creator (such as an AI system producing visual works "in the style of" a specific artist)?

It should be noted that style has never been protected by copyright for obvious constitutional and public policy reason. We do not believe that any additional protection is needed. Besides, as we also discussed above, some generative AI platforms already prevent generation of outputs that are based on the style of, or are inspired by, certain living artists. More generally, based on our conversations with many creative artists and professionals working in applied arts (e.g. interior designers), the transfer of style does not seem to be the core issue related to generative AI.