# AI and copyright – a licensor and publisher perspective

**Opening remarks**

We're in the Wild West phase of generative AI technology. Massive harm is being inflicted upon creative markets by fraudulent actors in the name of innovation and transformative fair use, and most organizations for creative professionals across the world have spoken up. While true innovation and lateral value creation does happen in the field, and is spectacular, IP law has its work cut out in urgently addressing the parts of professed ***value creation*** that is simply a direct ***value transfer*** from creators, their licensees, their heirs and other rights holders, to AI companies and their users.

The following is based on research into the workings, licenses and sources of **Stable Diffusion and Midjourney** over this past year, including the **Laion-5B** dataset and the **commercial exploitation ecosystem** around these systems, chiefly **Promptbase, CivitAI** and **Adobe Stock**, and its impact on creators. It is based on a professional understanding of international content licensing within print and digital publishing, and data-driven service and business development.

A key point for the purpose of the reasoning below is that AI models are being misconstrued as ***technology***, whereas by technical and commercial nature they are ***content repositories***. Key to their value proposition is **extraction of expressive content** from underlying works, in direct violation of the interests of authors and rights-holders, hugely detrimental to their marketplace value and the cultural and creative industries as a whole:

What were previously **display windows and marketplaces** for commissions and licensing have become scraping grounds for free raw material for replacements.

What was previously **cultural capital and a passive income source** for artists and authors has been weaponized into directly competing works.

In a cruel twist of irony, **online popularity**, having ones' works shared widely by fans, has been turned from free marketing or petty nuisance into replacement quality.

**Given this massive and ongoing theft, there is currently little to no marketplace incentive to create works using anything BUT generative AI, let alone to publish any works on the open internet nor any major platform. Informed opt-in collective licensing, as already proposed by Author's Guild and others before, and already extant for other media and for other uses, and indeed *standard procedure for AI training until just recently*, would restore balance.**

**Johan C. Brandstedt**
Digital business developer @ HiQ Göteborg
Former editor, Swedish editions of *Scooby-Doo Magazine*, *Tom&Jerry Magazine* (Warner Brothers), *Dragon Ball* manga, *Shonen Jump* magazine (Shueisha), *Manga Mania* magazine (Kodansha)
Former studio manager, *Jungle Peak Studios* (IP developer)
Illustrator and cartoonist (IP creator)

*1. As described above, generative AI systems have the ability to produce material that would be copyrightable if it were created by a human author. What are your views on the potential benefits and risks of this technology? How is the use of this technology currently affecting or likely to affect creators, copyright owners, technology developers, researchers, and the public?*

This past year has seen more images generated than photos taken in all of history. AI-music generated this year already makes up 14% of all music ever recorded.

Midjourney alone claims to have close to **16 million** customers and the research model Stable Diffusion has been downloaded **over 12 million** times, and spawned **hundreds** of spinoff models and projects.

As shall be seen below, the authors of the underlying original works — the source data — are being crowded out from their own marketplaces by derivatives of their own works.

Output *quantity* from generative AI evidently needs no market incentives, such as through legal protection.

Further, a number of **egregious violations of existing law and contracts** continue to power this boom, and ongoing market harm to artists and photographers can hardly be overstated.

The Stable Diffusion image model – a ***research*** model – has been included in the ***for-profit*** Midjourney service, as well as in Stability.ai's own for-profit image generator, Dream Studio, as well as a number of other products.

Adobe were quick to permit selling of this pirate model output on its **Adobe Stock** service, on equal terms and at equal price points to traditionally-produced designs – radically undercutting and crowding out its existing contributor base. They then co-opted their entire Contributor base to train their own directly competing image generator, **Adobe Firefly**, without offering prior information nor even an opt-out to this day.

This has been nothing short of devastating to traditional and commercial artists. Famous illustrator **Michael Whelan** called out July as the worst month ever in terms of infringement; fantasy artist **Greg Rutkowski** testified to the BBC of having been copied more times than Picasso, drowning out his real work from search results; artist **Kelly McKernan**, plaintiff in one of the class action lawsuits against Midjourney and Stability, testifies to a precipitous drop in commissions, as do many others. There have been numerous reports of layoffs at art departments, and traditional online marketplaces such as Art Station, Etsy, DeviantArt, Adobe Stock and other stock sites are flooded with AI images, drowning out traditional art.

We will here go through the steps of the generative AI value chain as it exists on the market today, outlining the legal loopholes and outright violations at each step. This will inform how copyright law can stem the worst abuse as experienced on the market currently, while stimulating the novel, lateral, non-competing value creation and innovation — the science and useful arts — that copyright is intended to incentivize.

**Data sourcing**

1. **Works published online without the creator's awareness**
2. **Publicly available taken as Public Domain**
3. **"Openly licensed content"**
4. **Blanket claim of transformative fair use**
5. **Violation of the EU Text- and Datamining (TDM) exceptions**
6. **Blatant disregard for machine-readable opt-outs**

**Exploitation**

7. **Commercial exploitation of research models**
8. **Targeted exploitation of quality works by living artists and registered trademarks**
9. **Examples of direct market harm**

**Protest**

10. **Creative professional organizations protest unanimously**
11. **150 legal procedures against generative AI**

**Criminal conduct**

12. **DMCA and GDPR non-compliance**
13. **Fraudulent copyrightability claims**
14. **Fraudulent attribution claims**
15. **Incitement to infringement**

### 1. Works published online without the creator's awareness

Works found online are most often published by fans without the knowledge or consent of the copyright holder in the first place; the definition of a popular artist, and a popular work, is that ***other people save and share it*** for their enjoyment on various platforms, as well as on the open internet and in pirated collections.

Historically, the onus on artists to police the internet through DMCA takedown has rarely been worth it, as publicity also drives sales.

The need for creators to be vigilant with regard to their works' online presence has been abruptly turned from non-issue to survival imperative by big-business scrapers, without prior warning – violating the moral right to consent to derivative uses.

### 2. Publicly available taken to be Public Domain

David Holz of Midjourney has publicly admitted to not licensing any of the copyright works that power his service – a blatant crime against the exclusive commercial exploitation rights of copyright holders and the right to informed opt-in consent for commercial use of their

work. This is admitted with a shrug in an interview, as being par for the course in the field:

> When asked: “Did you seek consent from living artists or work still under copyright?”
>
> Holz replies: “No. There isn’t really a way to get a hundred million images and know where they’re coming from.
>
> “It would be cool if images had metadata embedded in them about the copyright owner or something. But that’s not a thing; there’s not a registry.
>
> “There’s no way to find a picture on the internet, and then automatically trace it to an owner and then have any way of doing anything to authenticate it.”

This blatant falsehood sparked outrage, as relayed in the below article: https://petapixel.com/2022/12/21/midjourny-founder-admits-to-using-a-hundred-million-images-without-consent/

For DallE, OpenAI do the exact same thing; there have been claims of licensing, but they have yet to be verified by any licensor.

**3. “Openly licensed content”**

When launching Firefly, Adobe’s unified message on data sourcing was “Public Domain, already licensed and openly licensed content”.

There is a rich data commons voluntarily shared under Creative Commons licensing, eg. on photo sharing platforms such as Flickr. CC relinquishes *some* of the rights otherwise reserved through copyright. None of those licenses except CC-0, equivalent to a Public Domain waiver, permit commercial use *without as much as attribution.*

This deliberate overreach is also demonstrated by OpenAI’s abuse of FOSS licenses in producing their GitHub CoPilot service. The class action lawsuit currently in progress since 2021 numbers simultaneous breach of 11 different free and open source software licenses.

**4. Blanket claim of transformative fair use**

As outlined in several lawsuits from this past year, Midjourney, OpenAI and others have not only ingested massive troves of ill-gotten copyrighted works, but also personal, private and proprietary content of various kinds, including trademarked characters and logotypes and names of living professionals, as well as the personal expressive literary and artistic style of the same. As we shall see below, these are **core to the value proposition and marketing** of these models and services.

The go-to legal appeal often repeated by proponents is that this somehow constitutes “transformative fair use”. This point can be traced back to at least the 2019 comment of OpenAI to the USPTO & DoC request for comment on AI and IP:

> “We submit that proper application of fair use factors requires a finding of fair use, especially considering the highly transformative nature of training AI systems. This conclusion is strengthened by reference to existing analogous case law holding that the reproduction of copyrighted works as one step in the process of computational data analysis is a fair use of those works.”
>
> https://www.uspto.gov/sites/default/files/documents/OpenAI_RFC-84-FR-58141.pdf

This may apply to a range of other usecases, but DallE is the exact opposite.

As mentioned above, professional names and registered trademarks are key product features, in unaltered form. Prompting these, as well as titles of several copyright works, ***extracts the expressive content of the underlying works*** from the model, to ***marketplace replacement*** quality.

Additionally, **entire works** can be extracted in cases of overfitting, famously so with Steve McCurrys ***Afghan Girl***, and according to this paper, at predictable and non-neglible rates.

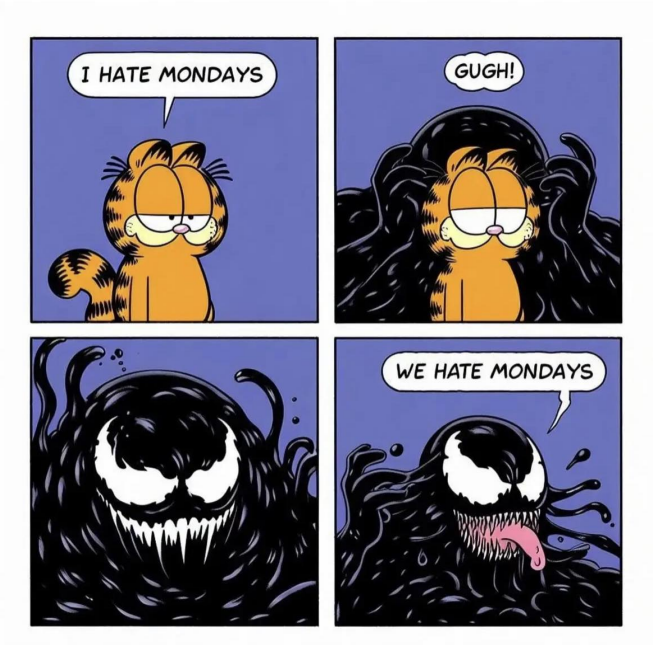
https://cfomax.com/paper-stable-diffusion-memorizes-some-images-sparking-privacy-concerns

(Overfitting is the well-known phenomenon of prevalent works in the training data producing near-matches in outputs.)

“Afghan Girl” prompted in Midjourney 5.2.

That specific work was banned from prompts in both Midjourney and DallE3 – but this can be easily circumvented by users of low technical skill.

The same phenomenon is prevalent in Dalle3:




Evidence like this defeats the line of reasoning that data extraction from the underlying works only extracts the *"facts and ideas"* as in fair use argumentation and the TDM exemption, rather than the *expressive content.*

**5. Violation of the EU Text- and Datamining exceptions**
Across the EU as of Jan 1st 2023 and in Germany as of 2021, the EU Digital Single Market directive permits research scraping of copyrighted works without prior informed consent. This comes with two important restrictions:

1) It does not extend to research institutes with close ties to for-profit institutions with preferential access to results.

2) Commercial use requires an opt-out possibility for copyright holders.

At the time when Stable Diffusion was trained, Stability.ai paid both the salary for the head of research at their academic research partner, later to become an employee, and the training compute for producing the model. In exchange, they got day one access to commercialize the results on their service Dream Studio. No prior notice was given, and **no opt-out system was even in place** when the services launched. As of May 2023, Stability partner Spawning.ai has processed **a billion opt-outs.** Still, models based on those works are available for download from Hugging Face, and incorporated in Midjourney.

In addition to direct monetization, the publishing of Stable Diffusion provided the nascent startup immense publicity and marketing benefit, quickly becoming one of the most popular downloads on Hugging Face.

**6. Blatant disregard for machine-readable opt-outs**

Several machine readable opt-outs are readily available as per the EU DSM TDM requirements, and have demonstrably been actively ignored by AI data scrapers:

- Platform Terms Of Use
- On-page copyright information
- CMI metadata embedded in image files
- Rights and attribution information in filenames and image descriptions
- Rights information and signature visible in the image itself
- Watermarks
- Common sense source URLs: stock sites and art marketplaces contain nothing BUT copyright works with all rights reserved, put on display for licensing and sales.
- Copyright notices embedded in source texts and code.

One recent study of sites Terms of Use showed that they largely explicitly forbid third party scraping.

The C4 dataset used by OpenAI and others contains 200 million instances of the copyright symbol according to a Washington Post investigation.

As for watermarks, many different ones are famously present in Stable Diffusion outputs; a smoking gun for large volume scraping and targeted training of proprietary material.

**7. Commercial exploitation of research models**

> "The model is intended for research purposes only."
> **– Stability.ai Stable Diffusion Terms of Use**
>
> https://huggingface.co/runwayml/stable-diffusion-v1-5
> https://huggingface.co/stabilityai/stable-diffusion-2
>
> "As between you and Stability, you own the Content that you generate using the Services to the extent permitted by applicable law."
> **– Stability.ai Dream Studio Terms of Service**
> https://dreamstudio.ai/terms-of-service

The same commercial entity here is ***publishing a research model*** with one hand, and ***renting out access to it and granting ownership of outputs from it*** with the other.

**8. Targeted exploitation of quality works by living artists and registered trademarks**

According to public information and the direct admission of Emad Mostaque, CEO of Stability.ai, Stable Diffusion filtered out half of the 5.6 billion images in the Laion-5B dataset before training. Furthermore, the subset **Laion-aesthetic**, comprising 600M images, was used for extra training to achieve better results.

According to publicly available information, the latter chiefly comprises material from **e-commerce, stock photo, art market, image board, image curation, professional magazines** and other sites that near-exclusively feature copyright works by living artists. **The product quality and marketplace value of these services is directly derived from unlicensed use of copyright works.**

Stable Diffusion comes distributed with a **prompting hit list** of named artists, many of which are active living professionals and recently deceased whose next of kin retain rights. This list was subsequently incorporated in a number of **commercial software products**, such as the ImagAIne app and Alpaca ML, a Photoshop plugin. It is also part of the basic prompt guide for Dream Studio, Stable Diffusion's official paid interface:

> "To make your style more specific, or the image more coherent, you can use artists' names in your prompt. For instance, if you want a very abstract image, you can add 'in the style of Pablo Picasso' or just simply, 'Picasso'."
>
> https://dreamstudio.ai/prompt-guide

Picasso's body of 45,000+ works will not enter the public domain until 2043, 70 years after his death in 1973. The rights for this marketing and commercial derivative use have most certainly not been cleared with his estate.

Further, these services actively encourage prompting for art marketplaces such as **ArtStation**, and living artist names, both in marketing such as the official Midjourney Magazine, curated user forums such as Midjourney community showcase, and the bustling Midjourney commercial exploitation ecosystem. A few examples:

- https://midlibrary.io/ features **3091 named artists** (painters, photographers, illustrators, etc many of which are active) and **245 titles** of popular properties at time of writing.
- https://www.artvy.ai/ features **600+ named artists**, as well as titles
- https://promptbase.com/ features several trademark properties as Midjourney prompts for sale: **1400+ Marvel, 450+ Disney Princess, 400+ Pixar, ~45 Barbie**, etc. These prompts also often state the intended use, which is invariably of direct marketplace replacement: **coloring pages, book covers, mascots, t-shirt prints, posters,** etc
- https://civitai.com/ **6000+ plugin models** for Stable Diffusion; many based on direct personal likeness and intended for deepfakes, others based on artist's entire catalogue of works to produce direct style likeness. Models are monetized off site on Patreon, Ko-Fi. Examples:

- **Sarah Andersen**, artist, plaintiff in class action lawsuit against Midjourney, Stable Diffusion; this is clearly an attempt at intimidation and harassment. https://civitai.com/models/85887/sarah-andersen
- **Greg Rutkowski**, famously prompted artist; this is unauthorized direct exploitation. https://civitai.com/models/117635/greg-rutkowski-inspired-style-lora-sdxl
- **Michael Whelan**, artist: same as above. https://civitai.com/models/120731/michael-whelan-sdxl-10-art-style-lora https://civitai.com/models/111028/michael-whelan-sd-15-art-style-lora
- Sam Yang AKA SamDoesArts spoke up against this abuse, only to find model makers announcing a competition on the official Stable Diffusion Discord to make the best replica model of his art. Evidence has since been erased, but several SamDoesArts models are available on CivitAI.

**9. Examples of direct market harm**

https://time.com/collection-post/6309445/kelly-mckernan/ Kelly McKernan: "freelance opportunities disappearing"

https://www.bbc.co.uk/news/uk-wales-66099850 Greg Rutkowksi: copied more times than Picasso; crowded out from search results.

https://twitter.com/whelanmichael/status/1688565174418964481 Michael Whelan: July 2023 "worst infringement month on record."

https://www.judiciary.senate.gov/imo/media/doc/2023-07-12_pm_-_testimony_-_ortiz.pdf Karla Ortiz, concept artist, gives several examples in her Senate testimony.

**10. Creative professional organizations protest unanimously**

> "for these platforms to be used ethically, there needs to be a **legal framework in place that does not freely exploit the intellectual property of creators** without permission or remuneration."
>
> **— Association of Illustrators (AOI), Feb 14th**
> supported by **The Association of Photographers (AOP), British Copyright Council (BCC), European Illustrator's Forum (EIF) and Creators Rights Alliance (CRA)**
> https://theaoi.com/2023/02/14/aoi-and-artificial-intelligence/

> “digital transformation must be shaped in full respect of **fundamental rights** (...) the right to **protection of intellectual property is a fundamental right** as it is defended in the Charter of Fundamental Rights of the European Union, the Universal Declaration of Human Rights, and the Universal Berne Convention.”
>
> **— European Visual Artists (EVA), May 11th**
> https://www.evartists.org/eva-statement-on-artificial-intelligence/

> “AI should not **disregard fundamental rights**, such as authors and performers rights, image, and personal rights, and should not be employed in ways that may deceive the general public. We urge the EU to promote original content and to respect human artistry and creativity.”
>
> **– 13 organizations, 100,000+ creative professionals, to the EU AI Act trilogue, Oct 2**
> https://www.ceatl.eu/ai-act-13-international-and-european-authors-and-performers-federations-call-for-a-human-centric-approach-to-generative-ai

> “the outputs of large language models (LLMs) rely heavily on authors’ ‘language, stories, style, and ideas’, without which they would be ‘banal and extremely limited’, however this is an **unlawful use of authors’ works.”**
>
> **– 10.000+ professional authors of the Author’s Guild, July**
> https://www2.societyofauthors.org/2023/07/25/join-thousands-of-authors-and-sign-the-authors-guilds-ai-open-letter/

> This is by no means an exhaustive list. **Japanese** creative professionals surveyed by one of Japan’s top newspapers expressed near unanimous concern. **American** associations of illustrators, painters, photographers, designers have issued a statement urging the EU to roll back the DSM exceptions, and so on.

**11. DMCA and GDPR non-compliance**

Due to the technical working of these models, **neither personal information nor copyright works can be removed** by the service provider upon request. This leaves rights holders with no means of having their professional names nor works opted out post fact upon discovery. This makes these services **inherently non-compliant** with DMCA Safe Harbor provisions and GDPR Right To Be Forgotten as extant in the EU and by extension in the US through Schrems-II, and with corresponding integrity law in other countries.

All of the for-profit actors – and they are many by now – **OpenAI, Adobe, Microsoft, Google, Meta, DeviantArt, Shutterstock, CivitAI, Promptbase** – choose the very narrowest definition of DMCA compliance by removing individual infringing works upon discovery, and/or offer indemnification to end-users.

DMCA takedown procedures were already a ceaseless whack-a-mole for artists; this new wave of rollouts further burdens individual professionals to assert their rights against some of the world’s most powerful companies.

**12. Fraudulent copyrightability claims**

The original Midjourney Terms of Service, since altered, claimed users, then Midjourney themselves, were **copyright holders of model outputs** despite their better knowing. Millions of $30/mo subscriptions were sold under this fraudulent pretense. This has contributed heavily to the proliferation of AI art in direct marketplace replacement use such as stock, editorial and publishing.

**Rights you give to Midjourney**
By using the Services, you grant to Midjourney, its successors, and assigns a perpetual, worldwide, non-exclusive, sublicensable no-charge, royalty-free, irrevocable copyright license to reproduce, prepare Derivative Works of, publicly display, publicly perform, sublicense, and distribute text, and image prompts you input into the Services, or Assets produced by the service at your direction. This license survives termination of this Agreement by any party, for any reason.

**Your Rights**
Subject to the above license, you own all Assets you create with the Services. This does not apply if you fall under the exceptions below.

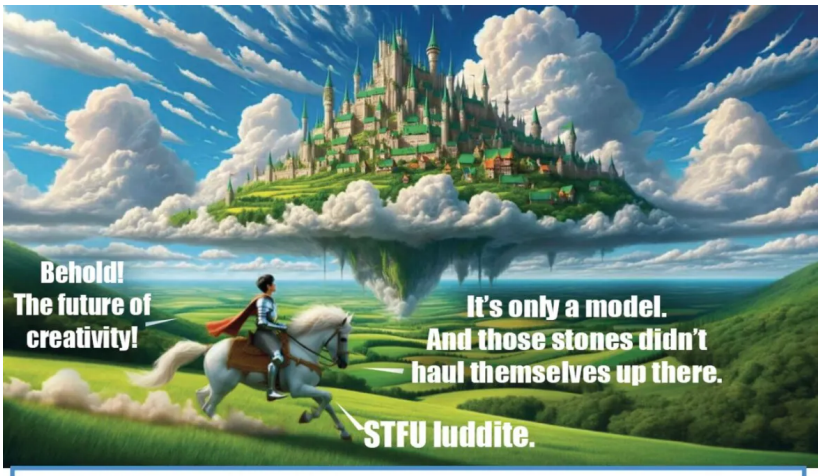
January version of Midjourney's Terms of Service. Prior versions have been erased from the Internet Archive.

http://web.archive.org/web/20230130031238/https://docs.midjourney.com/docs/terms-of-service

**13. Fraudulent attribution claims**

Core to AI product marketing is confounding ***idea* and *expression*** in the mind of users, selling the idea that commissioning an image from a model constitutes copyrightable **personal creative expression** despite lack of direct control over outputs.

One study of the Midjourney public gallery showed that prompts typically encompass a dozen or so directions on motif and visual style – a tiny portion of the myriad creative decisions large and small that go into the creation of a work, much more equivalent to **commissioning or art directing** an image than making it. The below illustration captures the balance here between creative effort and control:




"Painting of a fairytale knight in shining armor on a white horse riding by a castle in the sky on a cloudy summer day"

| CHOICES MADE BY YOU | CHOICES MADE FOR YOU | |
|---|---|---|
| Subject | Painting style | Landscape shape |
| Medium | Painting technique | Vegetation |
| | Layout | Horse breed |
| | Composition | Costume design |
| | Color palette | Ethnicity |
| | Aspect ratio | Pose |
| **SKILLS USED BY YOU** | **SKILLS USED TO MAKE A PAINTING** | |
| Typing | Imagination | Artistic style |
| | Observation | Landscape painting |
| | Artistic interpretation | Material studies |
| | Composition | Lighting and shading |
| | Perspective | Dynamic action |
| | Color theory | Horse anatomy |
| | Layout | Human anatomy |
| | Hand-eye coordination | Landscape design |
| | Penmanship | Costume design |
| | Color mixing | Architecture |
| | | Digital mastering |
| **INPUTS** | **INPUTS** | |
| ~100 characters | ~100.000 brush strokes | |

This "ghostwriter fallacy" has a strong psychological grip on amateurs – even in the face of direct targeted extraction of artist style, algorithmically derived from the training data, they will assert that they "created" it, whereas anyone who has ever made or commissioned image professionally know that this lands firmly in the second category.

Expect a deluge of testimonials based on this delusion, claiming Midjourney default aesthetics applied over pirated source material as personal creative expression.

**14. Incitement to infringement**

The blanket assertion of **transformative fair use**, along with the promise of **copyrightability** and **ease of extracting copyright works and "personal" expression** has spurred the user base to conduct **large scale, systematic infringement** with **disastrous market impact**, as exemplified above and below.

Active infringement has been encouraged every step of the way by service providers; from the Midjourney front page and Magazine parading prompts using artist names, to the first demo example shown to all Dream Studio users being to prompt for a painting by an artist, to distribution notes for the SD model and online manual for the Dream Studio product.

—

*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?*

It is not hyperbole to claim that there is a **full-blown extermination war** going on against professional artists. Entry-level opportunities and simple gigs are all but wiped out, and gleefully so, by grifters exploiting pirate models and targeting named artists:

"New Midjourney V5.2 Can Make You $25,570/month (Guide)"
https://www.youtube.com/watch?v=OujO9mYbiyA&ab_channel=Monice

"How I make 5000$ a Month using Midjourney"
https://www.youtube.com/watch?v=Q9feaVuqYIU&ab_channel=TheDorBrothers
This one refers explicitly to using living artist names as style prompts.

https://talkingbiznews.com/media-news/how-venturebeat-plans-to-use-ai-in-its-content/

Editorial decision to "embrace" pirate models to stay competitive.

Some platforms take a stand against AI-generated output, for legal or moral reasons:

> **Valve:**
> "We know it is a constantly evolving tech, and our goal is not to discourage the use of it on Steam; instead, we're working through how to integrate it into our already-existing review policies. Stated plainly, our review process is a reflection of current copyright law and policies, not an added layer of our opinion. As these laws and policies evolve over time, so will our process."
> https://www.eurogamer.net/valve-says-ai-generated-content-policy-goal-is-not-to-discourage-the-use-of-it-on-steam

**BackerKit:**
“This policy emphasizes that projects on BackerKit cannot include content solely generated by AI tools. **All content and assets must first be created by humans”**
https://www.backerkit.com/blog/backerkit-ai-policy/

However, **Adobe Stock** and **Amazon Kindle Direct Publishing** have become free-for-alls for publishing pirate model outputs – even with blatant infringement:

Etienne Habinger, artist, August:
“After seeing **Loish** and **Bastien Lecouffe Delharme's** post about their work copied and sold on **Adobe Stock,** I went there to see if I can give Adobe more money buying a crappy thefted AI generated stuff from my favourite artists in their store. The images isn't that great yet but at least they exist and I found most of my favourite artists!

**Pascal Campion Art, Lorenzo Rendih Lanfranconi, John Howe, Ruan Jia, Alariko, Sylvain Sarrailh, Ian Mcque, Darek Zabrocki, Aaron Limonick, Jean-Baptiste Monge, Andreas Rocha, Magali Villeneuve, Paul Chadeisson, Simon Stalenhag, Rebecca Dautremerr, Vitaly Bulgarov, Tyler Edlin, Anton Fadeev, Maciej Kuciara, Nicolas Sparth Bouvier, Finian Finnian MacManus, Serguey Kolesov, Marc Simonetti, Dylan Cole, Alberto Mielgo, Guweiz etc.**

This is just the tip of the iceberg because most of the Ai generated content there don't mention the name of the plagiarized artist or the prompt in the title as it was the case here. Oh, and do you know how they build/train their "Ethical" AI software (Firefly) ? By using the images uploaded on Adobe Stock" 🤡“

https://www.linkedin.com/posts/johan-cedmar-brandstedt-a77b311_adobeartttheft-generativeai-createdontscrape-activity-7098940981860806656-7XDc/

–

Author **Jane Friedman** found reviews and entire books published in her name on Amazon Kindle Direct Publishing, also August:
https://janefriedman.com/i-would-rather-see-my-books-pirated/

–

Generative AI works have prompted Amazon to cap the bot flood of Kindle Direct Publishing submissions at three titles per day:
https://gizmodo.com/amazon-restricts-author-self-publishing-3-books-a-day-1850864471

*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.*

This study by KPMG frames AI adoption as a potential productivity boost, but also indicates 43% of writing jobs and 15% of graphic design jobs as impacted:

https://kpmg.com/uk/en/home/media/press-releases/2023/06/productivity-boost-from-generative-ai.html

An important side note here: AI PR tends to frame AI use as ***value creation***. TDM exceptions as currently interpreted effectively **erases cultural value and intellectual property** normally licensed and subject to informed consent, to the tune of ***trillions*** (as per the Getty lawsuit, 1.8trillion just for 12M images – not even a significant chunk of total training data volume). In the light of this, AI at present is ***value transfer*** – from creators and rights holders to AI companies and their users.

This recent study of Large Language Models shows the degree to which they produce verbatim quotations from underlying works using standard prompting. The conclusion is that quotation length rises linearly with model size, and that portions exceeding standard “fair use” quotation length is common:

https://arxiv.org/pdf/2310.13771.pdf

*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? [40] How important a factor is international consistency in this area across borders?*

International consistency is of utmost importance, as race to the bottom dynamics and copyright havens already drive this field.

**The law proposal currently on the table in France** gets it right by asserting the copyright holder's right to informed opt-in consent and compensation:
https://www.fieldfisher.com/en/locations/belgium/insights/french-national-assembly-proposes-new-law-to-secur

**The Chinese provisional law enacted in August** also puts forth important protections against unfair competition, as well as training data transparency and registration, informed consent by depicted subjects, mandatory watermarking:
https://www.chinalawtranslate.com/en/generative-ai-interim/

**The draft EU AI Act** has many similar requirements, data transparency being key:
https://crfm.stanford.edu/2023/06/15/eu-ai-act.html

**The EU TDM exceptions** work well insofar as they require opt-outs from commercial training and explicitly target the academic-private loophole exploited by Stability above, but should rather be seen as a **cautionary example** as the wording around what can be extracted has been abused to extend to the expressive content of the underlying works.
https://www.europarl.europa.eu/RegData/etudes/BRIE/2018/604942/IPOL_BRI(2018)604942_EN.pdf

**The Japanese TDM exception** is a full-blown disaster; the "non-enjoyment" clause is currently being used to put artists out of work at an alarming rate, and since training data is a black box, proving infringement is next to impossible:
https://ipwi.uj.edu.pl/documents/122195199/144296432/Ueno_2019_Cracow_Comment.pdf/adf57b6a-ad86-4bd6-94b0-4bbac61b049c

https://restofworld.org/2023/japans-new-ai-rules-favor-copycats-over-artists/

*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.*

**For *copyrightability of outputs***, current law suffices and present USCO deliberations draw the line in line with what ought to be incentivized in the marketplace. It does not unduly affect AI outputs negatively; models are stored expressive content by other humans, and does not merit special protections. Granting ANY protection to AI *generated* works – regardless of beneficiary – would be disastrous to the very idea of copyright, as they can be automated at insane rates, and would immediately skew the market – as has already been demonstrated in the market. If anything, taxation and tariffs should be levied to stem the flow of exploitative use.

As for AI ***assisted*** works, copyrighting the man-made portions, as presently, makes sense. The test of *initial medium* is a good first watershed; anything started in writing ought not to merit copyright claims over an image, as per the ***direct control*** and ***idea/expression*** tests (prompts express *ideas*, image and text generators provide stored *expression*)..

Compare also to the ***spontaneous double creation test*** established for generative art way before computers; if you and I could both have made the image, such that a third person can't tell them apart as being made by different people, they are evidently not an expression of the author's personality, and thus not sufficiently creative to merit copyright.

Thus, starting the work from an image or blank canvas would be the first qualifier for copyrightable images. From there on, a ***majority of the direct creative decisions*** such that the image ***expresses the personality of the author*** – rather than that of the tool or some third party trained on – ought to be the test.

**As for *protection of authors of the underlying works*** – personal expression being the basis of copyright in the first place, **extending that protection to style** makes sense. Prompting by name is a direct exploitation of professional reputation, and indirect exploitation of skills built, as expressed in works "trained" on.

AI "training" is statistical analysis; it has nothing to do with human learning. The value of models is only partly lateral to the expressive content of the underlying works. As shown above, most value as being exploited today is ***directly derivative*** – the kind copyright is set to protect.

The rights that need asserting here are the ones established in the Berne convention, lucidly expressed as ***consent, credit and compensation + transparency.***

Only with these in place can fair compensation be established.

As for outputs, **mandatory watermarking and embedded content provenance** will be key – both to maintain property rights, but also to safeguard our information commons.

## “Training”

*If your comment applies only to a specific subset of AI technologies, please make that clear.*

*6. What kinds of copyright-protected training materials are used to train AI models, and how are those materials collected and curated?*

First off: “training” is data processing. This is a technical process not different in nature to any other compression algorithm.

In addition to the examples above, **Common Crawl** feeds Laion. As for OpenAI and Midjourney, public statements and evidence points to use of **pirated training data** as well, such as the **Books3** dataset sourced from cracked ebooks shared via bittorrent trackers and collected at piracy hubs such as The-eye.eu. This is thoroughly documented here: https://aicopyright.substack.com/p/the-books-used-to-train-llms

As for *Midjourney Niji*, training data is likely focused on **“booru” anime/manga art boards** and possibly **manga book scans** – all copyright works.

Whereas the range, style and quality of cinematic imagery in *Midjourney 4* and *Midjourney 5.2* suggests collections of **high-quality photos and copyrighted movie stills.**

This will of course be up to the FTC to probe and prove.

***Curation*** is very much a community effort – the open source and Midjourney user communities have been actively sourcing and selecting.

*6.1. How or where do developers of AI models acquire the materials or datasets that their models are trained on? To what extent is training material first collected by third-party entities (such as academic researchers or private companies)?*

See above for Stability and OpenAI.

*6.2. To what extent are copyrighted works licensed from copyright owners for use as training materials? To your knowledge, what licensing models are currently being offered and used?*

To my knowledge, OpenAI claims to have licensed training data, but neither them nor anyone else has named sources beyond the research datasets mentioned above. Nor have any sellers of training data come forth about selling training data to any AI companies.

As for licensing models – bulk licensing of royalty-free and rights cleared image-text pairs are offered by data sellers such as Vaisual: https://www.datasetshop.com/

The **Getty v Stability lawsuit** also mentions bulk licensing of images as training data as par for the course.

When it comes to more specific datasets, such as artist styles based exclusively on their body of work, royalty is the normal deal structure from other media, publishing, movie adaptations, merchandise, etc. Royalties range from 15% to 50% based on the portion of works used and negotiation. It is unclear at this time if any of the major players engage with this sort of licensing arrangement.

*6.3. To what extent is non-copyrighted material (such as public domain works) used for AI training? Alternatively, to what extent is training material created or commissioned by developers of AI models?*

Public Domain works make up some portion of training data. Adobe alleges to use “openly licensed content” which is not a legal term, and likely a sign of using Creative Commons content – which is questionable, as all CC licenses except CC-0 comes with some rights reserved.

*6.4. Are some or all training materials retained by developers of AI models after training is complete, and for what purpose(s)? Please describe any relevant storage and retention practices.*

To the best of my knowledge – training data needs to be retained for the duration of AI training, which can last weeks. Original training data is also needed when re-training with new data to avoid model collapse from AI inbreeding, so is typically retained between runs.

*7. To the extent that it informs your views, please briefly describe your personal knowledge of the process by which AI models are trained. The Office is particularly interested in:*

*7.1. How are training materials used and/or reproduced when training an AI model? Please include your understanding of the nature and duration of any reproduction of works that occur during the training process, as well as your views on the extent to which these activities implicate the exclusive rights of copyright owners.*

Asking how the sausage is made is not a productive line of inquiry as to whether the butcher owes money to the rancher. Even so —

AI PR as iterated by Stability and Adobe in their Senate Judiciary testimonies are fairytale versions for the gullible. Key of which is to anthropomorphize what is simply advanced statistical analysis, to obfuscate liability.

If we instead listen to the techies, “Intelligence is compression” in the words of Hardmaru, former Stability CTO. Image generator models are radically efficient, lossy, interleaved stores of expressive content mapped to descriptive text tokens.

The dataset of image-text pairs is first downloaded without the informed consent of rights holders, which is the first issue. The image files in question are publicly available viewing copies intended for human consumption; not public domain waivers of commercial exploitation rights.

Once downloaded, quality, NSFW and celebrity filters are applied, and automatic categorization is performed. The works are then retained for analysis for the duration of several training “epochs”, where quality works – works with a high aesthetic rating – are processed for additional epochs. This demonstrates that higher quality input data means better models with better output, meaning higher market value.

Copyright works used for models intended to produce ***market replacement derivatives*** (the core use case of generative AI is to generate new instances of the training data) should of course have been subject to the artist’s informed consent.

*7.2. How are inferences gained from the training process stored or represented within an AI model?*

Pass. Irrelevant to the issues at hand.

*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?*

No. A recent Microsoft paper clearly shows the intent of unlearning research: not to actually remove the ill-gotten data, but rather the most glaring traces of it:

https://winbuzzer.com/2023/10/09/microsoft-researchers-develop-unlearning-technique-for-large-language-models-xcxwbn/

*7.4. Absent access to the underlying dataset, is it possible to identify whether an AI model was trained on a particular piece of training material?*

Attribution is next to impossible as per current state-of-the-art research. Which is why **training and prompting records must be retained separately** for legal compliance and licensing reasons – see Training section below.

*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.*

Insofar as the use of the models can be restricted to wholly non-infringing and non-marketplace replacement use. ***Author's Guild v Google Books*, *LinkedIn v HiQ Labs* and *Perfect 10 v Google*** all point to **marketplace replacement** as the bright red line. This is clearly the exact opposite from Midjourney, Firefly, Dream Studio, Alpaca, ImagAIne, WomboDream, etc which all aim to directly substitute for the creative labor behind the source material in the market.

*8.1. In light of the Supreme Court's recent decisions in Google v. Oracle America [41] and Andy Warhol Foundation v. Goldsmith,[42] 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,[43] raise different considerations under the first fair use factor?*

As mentioned above, the core use case of generative AI is to substitute directly for instances of the training data. The purpose of an image generator model is to generate images that compete directly against the underlying works. The character of the use is producing a source for production of ***second-order derivatives*** of ***marketplace substitution grade.***

*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?*

See EU TDM exceptions above. There are legitimate uses for storing links, such as Laion, for research without ties to for-profit entities. But informed opt-in consent has to be the standard PRIOR to training.

Furthermore, sources such as http://The-Eye.eu which are clearly piracy-oriented and only nominally comply with the DMCA ought to be taken offline as they wreak havoc on the livelihoods of rights-owners.

*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.[44] How should the fair use analysis apply if AI models or datasets are later adapted for use of a commercial nature? [45] Does it make a difference if funding for these noncommercial or research uses is provided by for-profit developers of AI systems?*

See Stability example above. The EU TDM exceptions got this one right.

*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?*

Total volume does not say much, as both scraping and training is sequentially focused and weighted towards quality data through filtering, categorization, tuning, etc. – see Stability example above.

The substantiality factor for fair use ought to be considered from the share of total works in terms of training (concept artist Karla Ortiz testified to the FTC having ALL of her works taken) and in terms of stylistic likeness when looking at model outputs. 1:1 likeness of individual underlying works to individual outputs would be a terrible measure.

*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? [46] 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?*

**Generative AI is class warfare.** It targets entire classes of professionals by taking the entirety of their works and turning it against them in their own markets. Only in rare cases does that impact the market value of individual works. But as amply demonstrated above, it directly cannibalizes *individual artists, and also celebrities and models* by targeting their entire portfolios to extract expressive content and personal likeness, only to then produce replacement works for that particular artist or model.

Prompting professional names without that person's permission in order to produce marketplace substitutions of their works or likeness can never be fair use.

As for targeting specific media by living professionals – collective licensing should be compulsory, as per the French law proposal and as per the Author's Guild commentary on the 2019 USPTO RFC on AI & IP:
https://www.uspto.gov/sites/default/files/documents/The%20Authors%20Guild_RFC-84-FR-58141.pdf

*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)?*

As outlined above, ***entire bodies of works*** end up online continually, without the copyright holders knowledge or consent, so opt-out is untenable. Opt-in is the only option.

*9.1. Should consent of the copyright owner be required for all uses of copyrighted works to train AI models or only commercial uses?* [47]

Academic research use is fair game. There are also many non-profit uses the creator does not necessarily agree with, and it is be their moral right to abstain from those. A good example is the reckless release of Stability 1.4, which forever enables any number of vile uses in the style and name of the artists hijacked into that model.

*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?* [48]

As mentioned under question 1, answer 5, there are at least seven machine readable opt-out methods available today which are currently ignored by AI companies. They don't bother reading any of them. They even remove watermarks, for Christ's sake! The onus should not be on every artist, model and citizen to police the entire internet in the first place.

*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?*

There is an opt-out registry covering **ONE BILLION** works already, provided by Spawning.ai. OpenAI **IGNORED** it in building DallE3. This question is an insult to copyright holders everywhere.

> "In March, we delivered 80 million image opt-outs to StabilityAI ahead of their training Stable Diffusion v3. Less than two months later, we've exceeded 1 billion opt-outs." https://spawning.substack.com/p/our-seed-round-of-funding-to-build

*9.4. If an objection is not honored, what remedies should be available? Are existing remedies for infringement appropriate or should there be a separate cause of action?*

Any company that deliberately launches a commercial service, or continues to provide a service, based on **ONE BILLION** opted-out works, ought to be banned outright from operating. But as it stands, all of Big Tech – Meta, Microsoft, Alphabet, Amazon – has adopted or endorses this predatory rights-abuse.

*9.5. In cases where the human creator does not own the copyright—for example, because they have assigned it or because the work was made for hire—should they have a right to object to an AI model being trained on their work? If so, how would such a system work?*

Moral rights apply universally as per the Berne convention.

*10. If copyright owners' consent is required to train generative AI models, how can or should licenses be obtained?*

Opt-in, from artists themselves, or their appointed data brokers, agents and other licensees.

*10.1. Is direct voluntary licensing feasible in some or all creative sectors?*

Pass. Can only speak for licensing within illustration, comics, animation, publishing and syndication.

*10.2. Is a voluntary collective licensing scheme a feasible or desirable approach? [49] Are there existing collective management organizations that are well-suited to provide those licenses, and are there legal or other impediments that would prevent those organizations from performing this role? Should Congress consider statutory or other changes, such as an antitrust exception, to facilitate negotiation of collective licenses?*

Yes, voluntary collective licensing is a feasible approach. There are organized guilds and rights management organizations in most countries well suited to handle remuneration schemes. Models based on unlicensed property are illegal and must be destroyed.

*10.3. Should Congress consider establishing a compulsory licensing regime?* [50] *If so, what should such a regime look like? What activities should the license cover, what works would be subject to the license, and would copyright owners have the ability to opt out? How should royalty rates and terms be set, allocated, reported and distributed?*

Yes, there is a vast resource of amateur works and works by unorganized professionals and next of kin being exploited by AI companies. AI services should be taxed for exploiting such value, and the proceeds should be distributed towards the promotion of science and the useful arts internationally by the WIPO, nationally by the Department of Arts and Culture, via the corresponding entity or industry body on the state level.

Yes, personal and proprietary information should always be able to be opted out.

*10.4. Is an extended collective licensing scheme* [51] *a feasible or desirable approach?*

Yes.

*10.5. Should licensing regimes vary based on the type of work at issue?*

Yes, as today, clearances for biometry need to be handled separately from clearances for works, as they ground in different legal domains. A voice or face is personal and subject to integrity and privacy law, impersonation, defamation, libel, etc. Yet it may be part of a copyrighted audio recording or photo capture.

*11. What legal, technical or practical issues might there be with respect to obtaining appropriate licenses for training? Who, if anyone, should be responsible for securing them (for example when the curator of a training dataset, the developer who trains an AI model, and the company employing that model in an AI system are different entities and may have different commercial or noncommercial roles)?*

Curators are liable for data collection, model makers are liable for rights clearances and licensing.

*12. Is it possible or feasible to identify the degree to which a particular work contributes to a particular output from a generative AI system? Please explain.*

Yes, as long as the prompt is supplied one can correlate artist name, title, media descriptors, publishing source and other keywords along with visual similarity of the output in both specific details, overall composition, and particular features present in the training image and output but NOT in prompt. This can never provide 100% certainty, but a reasonable indication. Error sources include copycat works, naming overlap and model tuning.

*13. What would be the economic impacts of a licensing requirement on the development and adoption of generative AI systems?*

A balancing of interest between authors of the crucial underlying works and their exploiters. There is presently ZERO market incentive to create artistic works in any other medium than generative AI, let alone publishing any works online, as everything ever published is fair game for unfettered market exploitation by anyone. This needs correcting.

*14. Please describe any other factors you believe are relevant with respect to potential copyright liability for training AI models.*

Music licensing models show the way: base licensing for all training data, multipliers for specific titles and author names, cutoffs with remainder pooled towards general stimulus of the arts.

## 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?*

Yes and yes.

*15.1. What level of specificity should be required?*

Public declaration of aggregate information on sources and licenses. Full disclosure of rights-holder property on request.

*15.2. To whom should disclosures be made?*

Regulators, license collectors, and legal authorities in case of suspicion of fraud.

*15.3. What obligations, if any, should be placed on developers of AI systems that incorporate models from third parties?*

Full disclosure of models used.

*15.4. What would be the cost or other impact of such a recordkeeping system for developers of AI models or systems, creators, consumers, or other relevant parties?*

It is an absolute, non-negotiable prerequisite for a white market to exist.

*16. What obligations, if any, should there be to notify copyright owners that their works have been used to train an AI model?*

That's an absurd question. This would be like a magazine publisher notifying a writer, illustrator or photographer, or model, that their work or likeness is already available in newsstands across the nation, and abroad: the reputational and market damage is already done.

*17. Outside of copyright law, are there existing U.S. laws that could require developers of AI models or systems to retain or disclose records about the materials they used for training?*

Privacy and integrity law, in states where such exists.

## Generative AI Outputs

*If your comment applies only to a particular subset of generative AI technologies, please make that clear.*

**Copyrightability**

*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?*

Finding a pretty stick in the woods does not make one an artist, no matter how long one roams the woods. Models are derivatives of Other People's Works. Prompting is *retrieval*. And it can be – and is – automated. Bot flooding is already destroying markets with cheap generic derivatives. Quantity does not need incentives.

On the other hand, AI *models* or *finetunes* of licensed models *themselves*, if based entirely on the artist's own works, or duly licensed works, are derivative works prepared by the author under their direct control, and ought to be copyrightable.

*19. Are any revisions to the Copyright Act necessary to clarify the human authorship requirement or to provide additional standards to determine when content including AI-generated material is subject to copyright protection?*

No. Perfectly clear as-is.

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

Again, **MODELS ARE CONTENT**. The software/content distinction is key here. Content is passively stored information, like a database or an mp3 file. Models themselves should be copyrightable, given proper licensing of underlying works.

*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?*

No. Works either qualify for copyright, or they don't.

*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"? [52] If so, how?*

Not hardly, as that would undermine the entire idea of copyright and incentivize quantity, not quality. If anything, **AI generated works should be taxed**, by the same logic of other mass production.

AI-*assisted*, yes, if sufficiently directed by human hand.

**Infringement**

*22. Can AI-generated outputs implicate the exclusive rights of preexisting copyrighted works, such as the right of reproduction or the derivative work right? If so, in what circumstances?*

Yes, most certainly. Most blatantly so with prompting for overfitted works and when using copyrighted works as *image-to-image* and *remix* model inputs, but also in the case of style prompting as described under point 31 below. It is also possible to circumvent prompt blocks of artists and works through prompt engineering as long as the works are retained in the model.

*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?*

Data transparency is necessary: embedded prompt and model metadata in images, easily searchable training data records for models.

*24. How can copyright owners prove the element of copying (such as by demonstrating access to a copyrighted work) if the developer of the AI model does not maintain or make available records of what training material it used? Are existing civil discovery rules sufficient to address this situation?*

Not by a long shot. Ban blackbox sourcing.

*25. If AI-generated material is found to infringe a copyrighted work, who should be directly or secondarily liable—the developer of a generative AI model, the developer of the system incorporating that model, end users of the system, or other parties?*

Model makers for sourcing, service providers for prompt filters and DMCA compliance, end users for circumvention of such filters.

*25.1. Do "open-source" AI models raise unique considerations with respect to infringement based on their outputs?* [53]

Yes. AI models are CONTENT, not software. "Open-source" therefore does not apply in the first place! This is deliberate misinformation on the part of model makers. Publishing derivatives of copyright works whose purpose it is to produce second-order derivatives that serve as marketplace replacements of the underlying works with just a notice saying "research use only" is CLEARLY not enough, as entire direct exploitation ecosystems spring up in their wake, as demonstrated above. In practice, this is an assault on the property rights and livelihoods of entire classes of professionals.

*26. If a generative AI system is trained on copyrighted works containing copyright management information, how does 17 U.S.C. 1202(b) apply to the treatment of that information in outputs of the system?*

The inability of AI systems to process CMI is by design, as is the stripping of CMI by platforms such as Instagram, Threads, Twitter and Facebook. This is deliberate rights laundering and needs to stop. All training data must be kept on record.

*27. Please describe any other issues that you believe policymakers should consider with respect to potential copyright liability based on AI-generated output.*

–

**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?*

Source needs to be included as visible watermark, prompt and model needs to be in the metadata, and written declaration needs to accompany publication. C2PA shows the way, but approaches the whole issue as a voluntary afterthought, whereas it needs to be a by-default, non-negotiable, built-in feature.

*28.1. Who should be responsible for identifying a work as AI-generated?*

Model makers and service providers should be responsible for including watermarks and metadata. Model users should be held responsible for tampering. Platforms for securing correct metadata and retaining CMI information (this is deliberately being stripped by Meta et al today, as mentioned above). Publishers for accompanying source declaration at publication.

*28.2. Are there technical or practical barriers to labeling or identification requirements?*

Any supplementary information can be easily removed, just as any content protection or watermark. The tools to do this at scale are readily available on legal platforms despite violating DMCA. GitHub publishes watermark removal tools and Stability tout watermark removal as a featured use case of one of their products despite it being unlawful.

*28.3. If a notification or labeling requirement is adopted, what should be the consequences of the failure to label a particular work or the removal of a label?*

Fraud charge and fines for commercial infringement. Other charges for deliberate disinformation, impersonation, etc.

*29. What tools exist or are in development to identify AI-generated material, including by standard-setting bodies? How accurate are these tools? What are their limitations?*

Several are available and in development and they all produce a significant number of false positives and can be easily gamed.

**Additional Questions About Issues Related to Copyright**

*30. What legal rights, if any, currently apply to AI-generated material that features the name or likeness, including vocal likeness, of a particular person?*

A number of laws on harassment, impersonation, defamation, libel, etc. But foremost, data privacy and integrity.

*31. Should Congress establish a new federal right, similar to state law rights of publicity, that would apply to AI-generated material? If so, should it preempt state laws or set a ceiling or floor for state law protections? What should be the contours of such a right?*

Again, harmonizing such rights needs to be a multilateral international endeavor.

*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)? Who should be eligible for such protection? What form should it take?*

AI users are quick to point out that "one can't copyright style". Well, this is because one until very recently couldn't *copy* style without significant effort and overcoming barriers of workmanship honor and balancing the fine line between hommage and plagiarism.

As the ample examples under question 1, point 5 above show, copying style is a key feature and core value proposition of AI models, and carries a lot of marketplace value that is *being directly exploited from the artist of the underlying work*. Just as with dancing, photography, voice acting and other forms of artistic expression, developing a recognizable art style takes years and is a fusion of personal traits and trained skills – a result of lots of hard work and something not easily altered. Every novel and distinct art style is an enrichment of human artistic expression of the kind copyright law is set out to incentivize. **Every unlicensed exploitation thereof is a removal of such incentives.**

Furthermore, statistical analysis is fundamentally different in nature from human learning in that it is an *extraction* of personal expression from underlying works which, again, weren't licensed in the first place. *Any style LoRa, or successful "in the style of" prompt, is evidence of the model maker copying and processing copyrighted works – ie. preparing derivative works, transgressing the author's commercial monopoly to the fruits of her labor.*

Any named living artist or photographer should be protected from this exploitation. Service providers who wish to provide a named artist as a product feature must obtain a license in accordance with the author's moral rights, and negotiate compensation. When such license is withdrawn or expires, the bare minimum obligation by service providers is to block that name and titles of works from prompts. But this is far from enough. Research efforts into machine unlearning must be pursued, and present generation models must be disgorged as

per the mandate of the FTC, as the value of the stylistic content still resides in the model regardless of prompt blocking, and new ways of circumventing blocks are continually discovered.

This isn't about "copyrighting style" as such – humans will copycat, and that's its own set of issues – but rather about reasserting already present monopoly to commercial exploitation of works produced *through machine means*.

*33. With respect to sound recordings, how does section 114(b) of the Copyright Act relate to state law, such as state right of publicity laws?* [54] *Does this issue require legislative attention in the context of generative AI?*

Voice is biometric data and is foremost subject to laws of privacy and integrity. These need to be strengthened and clarified. Right to publicity may apply to character voices and singing style in addition, by the same logic as artist style above

.