General Questions

The Office has several general questions about generative AI in addition to the specific topics listed below. Commenters are encouraged to raise any positions or views that are not elicited by the more detailed questions further below.

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?*

There are **no benefits** of the generative AI systems to the original author, whose work is repurposed and plagiarized by the artificial intelligence training model for the intent of later re-selling the reconstituted image, video or written text without the creator's consent, authorization and/or compensation.

How is the use of this technology currently affecting or likely to affect creators, copyright owners, technology developers, researchers, and the public?

The creators, copyright owners, technology developers, researchers and the public are likely to lose significant sources of income from the unlawful copyright violations based on reconstituted generative AI systems.

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?*

Yes, the images, video and songs that are plagiarized by the generative AI systems are eerily similar to the original source material.

Napster and other systems that preceded OpenAI and other generative AI distributed processes placed direct facsimiles of music online.

The current technologies attempt to circumvent copyright, trademark and patent control, by arguing that taking a plagiarized *hybrid* version (i.e. a version that is undetectably different) of a written text by John Steinbeck (and/or any other author, including works that our family has generated and/or contributed to) from a pirate site (i.e. one that is operating outside of the parameters of the law (illegally)) is fair use.

Every movie currently pressed as a DVD, Blu-ray and/or streamed online has a warning from the FBI in the beginning of the film that defines the penalties for stealing copyrighted movies.

Books document the copyright registration at the beginning pages of the text or digital copy.

The Terms and Services of most technology startups prohibit reverse engineering and/or duplication of a company's code or intellectual property.

Microsoft, Facebook, Apple, Adobe, Oracle, Sun, HP, Twitter, Dell, IBM, Canon, Pandora, Spotify, Xerox, Sony and Google (at the larger end of the spectrum) all have specific provisions and legal teams in place to protect their intellectual property and/or software tied to their hardware.

Do these businesses consider repurposing their code by members of the public and/or their competition fair use? **No they do not**. They aggressively pursue anyone who misuses their material in any way. Consequently, the copyright, trademark and/or patent holder has a legal right to also to have protections and compensation for their intellectual property. To deny individual content creators their right to the same legal umbrella *is to deny their right to due process under U.S. law*.

Generative AI mechanisms are appropriating important works of literature, movies, music and works of art without written consent *of any* of the generators is not fair use, in the same way that repurposing a photograph for record album cover without authorization and payment is not fair use.

For example, our family provided the rooster image for the *Ambergris?* album by the band of the same name. It was a complicated endeavor requiring a professional animal trainer, full production studio with lighting, backdrops and film. The film had to be delivered to a photo lab (at that time), with processing fees and travel expenses to and from the facility to pick up and deliver the negatives.

The image was later re-used *without consent* and *authorization* by the band Pavement on their EP *Watery Domestic* without written permission, agreement and/or any compensation (of any kind) to our family.

Finally, there was also significant time and energy used in generating these responses, including printing and stationary costs to prepare physical mailings to the U.S. Copyright Office and U.S. elected officials (Congressional – House and Senate), postage and driving to and from the U.S. Post Office. For electronic submissions there are costs associated with domain and email registrations, annual hosting and programming (when and where applicable).

The personnel at the copyright office and the members of U.S. Senate and House of representatives involved in overseeing this process are being compensated for their time, effort, travel and expenses (per diem). The original copyright holders (the respondents) are not.

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.*

The European Union, and U.S. Academic Institutions (Harvard and Stanford) have begun to address the costs of artificial intelligence. The following paper(s), research and databases can be found at the subject hyperlinks listed below:

I. https://www.europarl.europa.eu/RegData/etudes/BRIE/2019/637967/EPRS_BRI(2019)637967_EN.pdf
II. https://hbr.org/2023/04/generative-ai-has-an-intellectual-property-problem
III. https://hai.stanford.edu/news/2023-state-ai-14-charts

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?*

The Brunel University London and the European Union have begun to address the artificial intelligence. The following paper(s), research and databases can be found at the subject hyperlinks listed below:

I. https://www.europarl.europa.eu/RegData/etudes/BRIE/2019/637967/EPRS_BRI(2019)637967_EN.pdf

II. https://www.wipo.int/export/sites/www/about-ip/en/artificial_intelligence/call_for_comments/pdf/org_brunel.pdf

How important a factor is international consistency in this area across borders?

International consistency is very important. For example on a DVD, Blu-ray Interpol's regulations (are also displayed during the course of the film).

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.*

Yes, new legislation is warranted to ensure that every author of creative material is notified in writing, similar to a mailing for opting in and/or out of a class-action lawsuit, and given the opportunity to remove their plagiarized videos from both the: (1)AI training model; and (2) the commercial system.

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?*

Artificial Intelligence (AI) systems scan and utilize nearly every line of programming code (unless protected by Robot.txt code, a firewall, pay wall and/or other electronic means), photographs or images (including those with watermarks), songs, writings, text messages and/or any other material currently being hosted on the web.

6.1. *How or where do developers of AI models acquire the materials or datasets that their models are trained on?*

Artificial Intelligence (AI) systems scan and utilize nearly every line of programming code (unless protected by Robot.txt code, a firewall, pay wall and/or other electronic means), photographs or images (including those with watermarks), songs, writings, text messages and/or any other material currently being hosted on

the web.

To what extent is training material first collected by third-party entities (such as academic researchers or private companies)?

It is uploaded directly into AI corporations' technology architecture, unless specified in the startups Terms and Services that they are providing it to academic researchers or private companies first.

6.2. *To what extent are copyrighted works licensed from copyright owners for use as training materials?*

There is currently no licensing of artificial intelligence materials and no compensation to the original generators of the work.

To your knowledge, what licensing models are currently being offered and used?

There are no known licensing models for any of the information being used.

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?*

All material on the web (i.e. online) whether it is public domain and/or copyrighted goes into the AI training systems.

6.4. *Are some or all training materials retained by developers of AI models after training is complete, and for what purpose(s)?*

Yes, the information is permanently stored in their databases, unless they are issued a legal request to remove the material from the training system and database(s) by a process server as a part of a lawsuit and/or specified in the company's Terms & Services.

Please describe any relevant storage and retention practices.

Everything that is hosted and visible and/or not visible on the web is stored permanently in AI systems.

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?*

As previously noted in 6.1, Artificial Intelligence (AI) systems scan and utilize nearly every line of programming code (unless protected by Robot.txt code, a firewall, pay wall and/or other electronic means), photographs or images (including those with watermarks), songs, writings, text messages and/or any other material currently being hosted on the web.

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.

As previously noted in 6.4, yes, the information is permanently stored in their databases, unless they are issued a legal request to remove the material from the training system and database(s) by a process server as a part of a lawsuit and/or specified in the company's Terms & Services.

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

AI extrapolates all of the useful data-points, shapes and text for any representation of an object that it interprets (i.e. a raven). For example, the training model takes multiple images of ravens in various forms of rest and flight and develops a disconcertingly analogous image.

7.3. *Is it possible for an AI model to "unlearn" inferences it gained from training on a particular piece of training material?*

AI can "unlearn" inferences from the training model. In order to prevent copyright, trademark and patent infringement, the model will need to be modified to remove the unlawful encroachments.

If so, is it economically feasible?

Yes, it is economically feasible. If the AI can train on billions of images. It can also in turn be programmed to "unlearn" information.

In addition to retraining a model, are there other ways to "unlearn" inferences from training?

Yes, the AI model can search for the words: Copyrighted, Copy Control, Trademark, Trademarked, Patent, Patented, etc.; scan for watermarks and remove any images, drawings, code, designs, et. al. that are illegally being repurposed.

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?*

Yes, many images are so unique and powerful, the shooting of the Vietnamese man in the head during Vietnam or the young girl running from the pain of agent orange during the same conflict, that any inferences or extrapolation of this content would clearly be infringing conflict. Subtler, but still recognizable variations of this unlawful re-use of information would be a Native American that was wearing a specific outfit during a photo shoot, like a leather jacket with very specific headwear, also constituting plagiarism.

Many of our families works are in museums and private collections, have received awards including, but not limited to, the *National Magazine Award*, *Aspen Design Conference Award*, a *CLIO*, etc., by juried peer of experts, consequently these images have been deemed unusual and uniquely distinguishable in relationship to other works of a similar nature.

8. *Under what circumstances would the unauthorized use of copyrighted works to train AI models constitute fair use?*

If a website and/or site specifies that all information is in the public domain, is open source and/or available for public use, then these instances might warrant fair use, but the owners of the sites should still have an opportunity to opt out if they are not-for-profits.

Please discuss any case law you believe relevant to this question.

George Washington University has a AI litigation database that is filled with a repository of emerging U.S. and international case law (UK, et. al.)

https://blogs.gwu.edu/law-eti/ai-litigation-database-search/

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?*

The authors of the underlying image should be credit, compensated and/or have their request to have the content removed if that is their wish.

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?

No, if the information is ultimately going to be used for commercial purposes that profit the parent company.

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?*

The same rules apply to them if they are complacent in the unlawful use.

8.3. *The use of copyrighted materials in a training dataset or to train generative the fair use analysis apply if AI models or datasets are later adapted for use of a commercial nature?*

If any copyrighted materials are later used for commercial purposes of any kind, then this is unauthorized infringement.

Does it make a difference if funding for these noncommercial or research uses is provided by for-profit developers of AI systems?

Yes, if the not-for-profit or noncommercial use is later converted into a for-profit use, then this is also unapproved and prohibited use.

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?

If a John Steinbeck novel is taken from a pirate book site, if the AI system later places a copy of that book on their platform, this is not fair use it is illegal and a violation of multiple U.S. statutory codes.

8.4. *What quantity of training materials do developers of generative AI models use for training?*

Billions of images, lines of programming code, videos, trademarks, patents, drawings, text, written material and other materials are used by the technology firms for training AI models.

Does the volume of material used to train an AI model affect the fair use analysis?

No, it is still unauthorized use.

If so, how?

N/A

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?*

Repurposing copyrighted material does not give the infringing corporation the right to reassign market value to something that they don't own.

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?

The original owner of the copyrighted work has the right under common law and U.S. statutory code to have a cease and desist issued for a particular copyrighted work, the body of works by the same author and the market for the entire general

class of works.

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)?*

Yes, they should have to opt in.

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

Yes, for all uses of copyrighted work.

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?*

The copyright owner would need be notified in writing of how their material had been removed from the AI system, what unauthorized use had occurred and/or any other potential libelous use.

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?

Robot.txt code and other firewalls can be incorporated on a website to ensure that the images and text on a website may not be repurposed by artificial intelligence. Additionally, AI could scan for the words copyright, trademark and patent and stop crawling a website with this information.

9.3. *What legal, technical, or practical obstacles are there to establishing or using such a process?*

None, technology companies with trillion dollar valuations should have no issue hiring programmers to ensure that the original copyright holders are protected properly.

Given the volume of works used in training, is it feasible to get consent in advance from copyright owners?

Yes, like class-action lawsuit notifications that affected parties receive in the mail.

9.4. *If an objection is not honored, what remedies should be available?*

The infringed on parties should be able to receive compensation per instance of misuse.

Are existing remedies for infringement appropriate or should there be a separate cause of action?

No, there should be a penalty for: (I) knowingly training the AI model on copyrighted material; and (II) deliberately commercializing the misused images.

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?*

Yes, they also should have a right to object to the AI model being trained on their work based on common law.

If so, how would such a system work?

In the same fashion as the aforementioned method regarding copyrighted materials.

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

Through the U.S. mail and notarized by the original copyright holder.

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

No.

10.2. *Is a voluntary collective licensing scheme a feasible or desirable approach?*

No. The original creators own the material.

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?

Only if the original copyright holder signs a legal agreement with them to represent them.

Should Congress consider statutory or other changes, such as an antitrust exception, to facilitate negotiation of collective licenses?

No.

10.3. *Should Congress consider establishing a compulsory licensing regime?*

No. Congress can't force a copyright holder to surrender their rights.

If so, what should such a regime look like?

N/A

What activities should the license cover, what works would be subject to the license, and would copyright owners have the ability to opt out?

Copyright usage is licensed on a per image use (i.e. a series of framed prints for a hotel chain in every room and then a separate license for international rights, television, movies, etc.).

How should royalty rates and terms be set, allocated, reported and distributed?

Royalty rates are set by the copyright holder.

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

No. Each instance of an image's use requires renegotiation and compensation.

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

Yes.

11. *What legal, technical or practical issues might there be with respect to obtaining appropriate licenses for training?*

None, it is possible for corporations to contact social media and/or cell-phone owners regarding a class action lawsuit regarding their technology, it is certainly possible for an artificial intelligence company like OpenAI to contact the original copyright holder.

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)?

Anyone, even if they are jointly involved, must endeavor to legally notify the original copyright holder.

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, some images are so distinctly unique that variations inferred from them would have been impossible make without utilizing the original material.

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

The economic impacts of a licensing agreement would be minimal because the Artificial Intelligence companies they have received millions of dollars in funding.

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

The speed at which the copyright infringement occurs (at a record pace).

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?*

Yes.

Should creators of training datasets have a similar obligation?

Yes.

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

The exact number of images, music, movies, text, etc. involved.

15.2. *To whom should disclosures be made?*

The copyright holder(s).

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

Both the third parties and the developers of AI should have the same obligation to obtain the consent and compensation and/or opt out from the original creators.

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?*

With cloud based storage it is relatively easy to store an near infinite number of names, addresses and other information to notify intellectual property owners of their right to opt in and receive compensation and/or out.

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

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?*

Yes, laws pertaining to medical records, minors, et. al. Additional, technology companies are required to provide sales records for each image sold to the government for accounting purposes.

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?*

No.

If so, what factors are relevant to that determination?

The information was plagiarized from multiple sources.

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?

No.

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. It is strict fraud by AI systems.

20. *Is legal protection for AI-generated material desirable as a policy matter?*

No.

Is legal protection for AI-generated material necessary to encourage development of generative AI technologies and systems?

No. The companies will move ahead with working on AI generative.

Does existing copyright protection for computer code that operates a generative AI system provide sufficient incentives?

Yes.

20.1. *If you believe protection is desirable, should it be a form of copyright or a separate sui generis right?*

Additional protection to an AI company beyond their Terms & Services is not warranted.

If the latter, in what respects should protection for AI- generated material differ from copyright?

Protecting something that is taken from a musician, author or artist without permission or compensation is analogous to stealing a car, someone's salary and/or anything else.

21. *Does the Copyright Clause in the U.S. Constitution permit copyright protection for AI-generated material?*

No. It is not possible to copyright something that is an unauthorized appropriation of another person's work.

Would such protection "promote the progress of science and useful arts"?

No.

If so, how?

N/A

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?*

No.

If so, in what circumstances?

N/A

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?*

Yes, it appears to be sufficient.

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?*

For unique photos, like the Vietnam assassination at point blank range or the young woman burned by agent orange in the same conflict; or Geronimo, the Native American posing with his gun.

Are existing civil discovery rules sufficient to address this situation?

Yes. However, the Artificial Intelligence companies should be held liable for the destruction of the records used in generating the content.

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?*

All of the aforementioned parties, including any re-use of AI models by end users to create new material from the material previously misappropriated (i.e. 3rd derivation).

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

Yes, it is more difficult to track open source downloads by members of the general public.

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

The original copyright holders still maintain their rights and title to their work.

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

The speed of the copyright infringement is extremely rapid.

LABELING OR IDENTIFICATION

28. *Should the law require AI-generated material to be labeled or otherwise publicly identified as being generated by AI?*

Yes, AI material, advertisements, movies, text needs to have a label identifying that they were not generated by a human.

If so, in what context should the requirement apply and how should it work?

A labeling system similar to movie ratings: G, PG-13, R, NC-17, etc.; music: Parental Advisory, etc.; and additional warnings on comic books, books, etc.

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

The companies that have developed and are commercializing the technology.

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

None, AI systems can be integrated in cloud base architecture databases that can include extensive labeling or identification requirements.

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?*

An additional penalty should be imposed by a regulatory body, FTC, Better Business Bureau, FBI, et. al.

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

The same tools and technology that generated Artificial Intelligence (AI) can also be used to identify AI-generated material.

How accurate are these tools?

Very accurate.

What are their limitations?

There are instances when the AI systems can misinterpret parameters and make identifications that are false and/or misleading creating a libelous situation.

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?*

None.

31. *Should Congress establish a new federal right, similar to state law rights of publicity, that would apply to AI-generated material?*

No. If the original material was plagiarized without authorization, then no federal rights are conferred on AI generated content.

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?

No. The mechanics of unlawful use at the state level are analogous to those at the federal level. No state rights are awarded to stolen content by an AI system.

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)?*

No.

Who should be eligible for such protection?

N/A.

What form should it take?

N/A

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?*

Sound recordings should continue to be afforded the same degree of protection that they have traditionally held.

No, artificial intelligence systems that illegally appropriate elements of music, should continue to be held liable, when and/or where applicable.

Does this issue require legislative attention in the context of generative AI?

Please refer to 33. Any legislation should favor artists and creators and decades of copyright law, not large entities seeking to circumvent the law by creating derivations and inferences of

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

Each creative endeavor by a content generated required the some element of the following: equipment (cameras, film, computer software and hardware, lighting, sets, paints (oil, latex, enamel, acrylic), fabric, thread, brushes, canvases, paper, ink, printers, travel, transportation (car, van or truck), instruments (guitars, drums, etc.), sound engineering, record pressing (or digital file), advertising, etc.

All of these items add up to significant resources for the artist, photographer, painter, designer, videographer, filmmaker, musicians, etc.

A large corporation would expect their software, hardware, lines of code and the expense of hiring thousands of engineers with salaries and fringe benefits, chip manufacturing, machining and plastic casting **not to be illegally appropriated** when developing a program and/or new laptop, cell phone, et. al.

VII. Glossary of Key Terms

The Office has included definitions of key terms as they are used in this Notice to clarify the technical processes involved in generative AI systems. The following definitions are used for purposes of this Notice only; they do not necessarily reflect the government's legal position with respect to any particular term.

Artificial Intelligence (AI): A general classification of automated systems designed to perform tasks typically associated with human intelligence or cognitive functions. Generally, AI technologies may use different techniques to accomplish such tasks. This Notice uses the term "AI" in a more limited sense to refer to technologies that employ machine learning, a technique further defined below.
AI Model: A combination of computer code and numerical values (or "weights," which is defined below) that is designed to accomplish a specified task. For example, an AI model may be designed to predict the next word or word fragment in a body of text. Examples of AI models include GPT–4, Stable Diffusion, and LLaMA.

AI System: A software product or service that substantially incorporates one or more AI models and is designed for use by an end-user. An AI system may be created by a developer of an AI model, or it may incorporate one or more AI models developed

by third parties.

Generative AI: An application of AI used to generate outputs in the form of expressive material such as text, images, audio, or video. Generative AI systems may take commands or instructions + from a human user, which are sometimes called "prompts." Examples of generative AI systems include Midjourney, OpenAI's ChatGPT, and Google's Bard.

Machine Learning: A technique for building AI systems that is characterized by the ability to automatically learn and improve on the basis of data or experience, without relying on explicitly programmed rules. Machine learning involves ingesting and analyzing materials such as quantitative data or text and obtain inferences about qualities of those materials and using those inferences to accomplish a specific task. These inferences are represented within an AI model′s weights.

Training Datasets: A collection of training material (as defined below) that is compiled and curated for use in machine learning. Examples of training datasets include BookCorpus, ImageNet, and LAION.

Training Material: Individual units of material that are used for purposes of training an AI model. They may include a combination of text, images, audio, or other categories of expressive material, as well as annotations describing the material. An example of training material would be an individual image and an associated text "label" that describes the image.

Weights: A collection of numerical values that define the behavior of an AI model. Weights are stored within an AI model and reflect inferences from the training process.