October 30, 2023

Library of Congress
U.S. Copyright Office
Attn: Rhea Efthimiadis
Assistant to the General Counsel
101 Independence Avenue, S.E.
Washington, D.C. 20559-6000
Telephone: (202) 707-8350
Email: meft@copyright.gov

RE: Docket No. 2023-6

To Whom It May Concern:

Regulosity LLC ("Regulosity") and Pangea Tech Solutions Pvt Ltd ("Pangea") appreciate the opportunity to provide comments on the United States (U.S.) Copyright Office ("Copyright Office"), Library of Congress notice of inquiry and request for comments titled "Artificial Intelligence and Copyright," docket no. 2023-6.

Regulosity assists international and domestic companies obtain and maintain market access for products worldwide. We are experts at navigating the intersection of public policy, business, and technology. Regulosity uses our expertise in water and plumbing, energy and mechanical, environmental, public health, electrical and hazardous locations, and innovation policies to create solutions that support businesses and achieve compliance with public policies.

At Pangea, we embrace the power of artificial intelligence (AI) technology and data analytics to create value and shared success for our client partners, stakeholders, employees, and community. Combining our experience and specialized skills, Pangea enables our clients to apply AI insights to solve problems, derive insights through data analytics, and increase revenue through customer satisfaction.

Contact Information

Regulosity LLC
247 W Freshwater Way, Suite 528
Milwaukee, WI 53204
Office: +1 (414) 988-8613
Email: info@regulosity.com

Contact Person: Mrs. Misty A. Guard
Contact Person Title: President

Pangea Tech Solutions Pvt Ltd
A-303, Mantri Splendor
Hennur Main Road, Geddalahalli,
Bangalore, Karnataka, India 560077
Office: +9 18023683393
Email: info@pangeatech.net

Contact Person: Mr. Vivek Nithianand
Contact Person Title: CEO

Regulosity and Pangea respectfully provide comments on the following issues to inform the Copyright Office's study of copyright law and the policy issues raised by AI systems to assess whether U.S. legislation or regulations are warranted.

Our responses aim to contribute to a better understanding of copyright issues in the generative artificial intelligence ("GAI") domain and to encourage discussions that lead to balanced and sustainable solutions. To address the complexity of GAI and copyright law, our responses endeavor to provide context from multiple perspectives, where applicable to the questions asked by the Copyright Office:

- Generative content users and the general public ("Public Users")
- Generative AI system users ("GAI System Users")
- AI developers who build and deploy generative AI systems ("GAI Developers")
- Content creators with copyrighted works ("Copyright Owners")
- Researchers ("Researchers")

Currently, researchers use multiple software platforms, such as, but not limited to, ATLAS.ti, NVivo, indeemo, SAS/STAT, SPSS, Stata, R, and Knime, specialized for a particular analytical method to achieve outcomes. A human then combines the results to understand a topic comprehensively. Existing software platforms require the use of syntax language, often proprietary to the platform, to achieve analytical outcomes. In research terminology, GAI systems conduct a meta-analysis while simultaneously performing qualitative, quantitative, and mixed methods analytics, accelerating research and learning. GAI's advancement is the ability to turn human-based queries into analytical syntax to produce an outcome. GAI system outcomes are limited by the human-based query and, therefore, by GAI System Users. GAI systems can generally be classified as:

- Topic-specific GAI systems, like IBM Watson Health's AI[1]
- Generic GAI systems, like ChatGPT[2]
- Organization-specific GAI systems, like private software applications not for public use

As the foundation for our responses, we ascertain the following foundational statements:

- GAI may learn replicable patterns to produce derivative works[3].

[1] A decision-support system that ranks cancer therapeutic options based on machine learning algorithms.
[2] https://openai.com/chatgpt
[3] Derivative works are works based upon one or more preexisting works, such as a translation, musical arrangement, dramatization, fictionalization, motion picture version, sound recording, art reproduction, abridgment, condensation, or any other form in which a work may be recast, transformed, or adapted. A work consisting of editorial revisions, annotations, elaborations, or other modifications, which, as a whole, represent an original work of authorship. CIRCULAR 92: COPYRIGHT LAW OF THE UNITED STATES AND RELATED LAWS CONTAINED IN TITLE 17 OF THE UNITED STATES CODE, https://www.copyright.gov/title17/title17.pdf

- GAI is not able to create "original works of authorship[4,5]" without the aid of humans.

We respectfully suggest that the Copyright Office balance fostering innovation, economic impacts, and preserving intellectual property rights in future regulatory actions.

Copyright Office Feedback Request No. 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?"

Public Users

Public Users benefit from new forms of entertainment. A risk is that Public Users find it challenging to identify content produced by GAI systems instead of human Copyright Owners.

GAI System Users

GAI System Users benefit from access to advanced technology, critical thinking and creativity improvements, and assistance with creating personalized material, such as original songs, stories, or artwork. GAI System Users could unknowingly violate laws, such as the Federal Trade Commission Act[6] (FTC Act) and copyrights, by creating false information and engaging in unfair methods of competition and unfair or deceptive acts or practices in or affecting commerce. GAI System Users are more interested in the GAI system's efficiency and efficacy and may not be directly concerned with the intricacies of how the GAI system model was trained. Another risk is the removal of barriers in the creation of offensive or damaging information. GAI systems can be used to create works that maliciously exploit, deceive, or misinform Public Users.

GAI Developers

GAI is accelerating advancements in commercial, industrial, research and development, education, statistical, and research applications. GAI advancements accelerate market strategies, access, and entry worldwide by reducing analytical timeframe constraints. GAI advancements require moral and legal assessments by GAI Developers to ensure GAI systems are used to minimize harm to Public Users and ensure legal protections, such as copyrights and laws, are not violated in the jurisdictions where the GAI systems are deployed.

[4] Circular 92: Copyright Law of the United States and Related Laws Contained in Title 17 of the United States Code, (2022), https://www.copyright.gov/title17/title17.pdf.
[5] Memorandum Opinion: Stephen Thaler v. Shira Perlmutter, Register of Copyrights and Director of the United States Copyright Office, et al, (2023), https://storage.courtlistener.com/recap/gov.uscourts.dcd.243956/gov.uscourts.dcd.243956.24.0_2.pdf; Order: Stephen Thaler v. Shira Perlmutter, Register of Copyrights and Director of the United States Copyright Office, et al, (2023), https://storage.courtlistener.com/recap/gov.uscourts.dcd.243956/gov.uscourts.dcd.243956.23.0.pdf.
[6] 15 U.S.C. §§ 41-58, as amended

Copyright Owners

GAI is helpful throughout the creative process, from idea generation to task automation to creating new content. GAI can replicate a Copyright Owner's "writing style," verbal speaking patterns, and verbal tonalities, to name a few. If a Copyright Owner relies on replicable patterns to produce copyright works, GAI may learn the replicable patterns and produce a derivative work for the Copyright Owner. Therefore, Copyright Owners could increase production and revenues using GAI systems. In creative-based industries, such as television, that rely on established replicating formats for copyright works, GAI systems could diminish the value of human screenwriting work, resulting in job losses.

Researchers

Restrictions on GAI systems may hinder future research and development in GAI systems and reduce advancements beneficial to Public Users. For example, restricting the use of IBM Watson Health's AI in oncology research would negatively impact oncology research and the eradication of cancers. The implications of GAI systems to society are research topics in philosophy, public policy, political science, psychology, and many others.

Copyright Office Feedback Request No. 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?"

GAI systems raise serious concerns in the Policy Analysis and Technology Industries. GAI system models are trained on real-world data compiled as training materials. The training materials can contain copyrighted works, publicly available information, such as from governments, or free information, such as open source or social media posts. The veracity of the training materials used for GAI system models is at the discretion of GAI Developers and GAI System Users. If the training materials contain biases, deepfakes, or other misrepresentations, such as political leanings, the GAI system will produce works with the same biases, inherent deepfakes, and misrepresentations. Both sectors rely on accurate, reliable, and repeatable training materials for robust recommendations. In the technology industry, disinformation can result in security flaws and other architecture issues. In the policy analysis industry, disinformation can result in policy recommendations with adverse Public User outcomes.

<<*Intentional blank space. See the next page.*>>

Copyright Office Feedback Request No. 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."

1. "Research for CULT Committee - The Use of Artificial Intelligence in the Cultural and Creative Sectors"[7]
2. "The Art in the Artificial AI and the Creative Industries"[8]
3. "Artificial intelligence: Development, risks, and regulation by UK Parliament – House of Lords Library"[9]
4. "So what if ChatGPT wrote it?"[10]
5. Canada's Statutory Review of the Copyright Act[11]

Copyright Office Feedback Request No. 4

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

International harmonization of GAI legislation and regulations is imperative for trade and importing or exporting products and services using GAI. International consistency fosters global competition while supporting innovation and fair competition between established businesses, start-up businesses, and entrepreneurs. Harmonization of GAI system regulations aligns with provisions of the General Agreement on Tariffs and Trade 1994,[12] Agreement on Technical Barriers to Trade,[13] Agreement on Trade-Related Aspects of Intellectual Property Rights,[14] General Agreement on Trade in Services,[15] and the Berne Convention.[16] GAI systems are used by all service sectors listed in the General Agreement on Trade in Services.

[7] https://www.europarl.europa.eu/RegData/etudes/BRIE/2020/629220/IPOL_BRI(2020)629220_EN.pdf
[8] https://creative-pec.files.svdcdn.com/production/assets/publications/PEC-and-Nesta-report-The-art-in-the-artificial.pdf
[9] https://lordslibrary.parliament.uk/artificial-intelligence-development-risks-and-regulation/
[10] https://www.sciencedirect.com/science/article/pii/S0268401223000233
[11] Statutory Review of the Copyright Act: Report of the Standing Committee on Industry, Science, and Technology, (2019), https://www.ourcommons.ca/content/committee/421/indu/reports/rp10537003/indurp16/indurp16-e.pdf.
[12] General Agreement on Tariffs and Trade 1994, (1994), https://www.wto.org/english/docs_e/legal_e/legal_e.htm#services.
[13] Uruguay Round of Multilateral Trade Negotiations, *Agreement On Technical Barriers To Trade*, (1994), https://www.wto.org/english/docs_e/legal_e/17-tbt.pdf.
[14] Agreement on Trade-Related Aspects of Intellectual Property Rights, (2017), https://www.wto.org/english/docs_e/legal_e/31bis_trips_e.pdf.
[15] Uruguay Round of Multilateral Trade Negotiations, *General Agreement On Trade In Services*, (1994), https://www.wto.org/english/docs_e/legal_e/26-gats.pdf.
[16] Berne Convention for the Protection of Literary and Artistic Works, (1979).

To facilitate the development of international harmonized GAI system legislation and regulations, we suggest the Copyright Office review the following regulatory framework and national strategy:

6. Artificial Intelligence Act, 2021/0106(COD), European Parliament.[17,18]
7. National AI Strategy of the United Kingdom.[19,20]

Copyright Office Feedback Request No. 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."

New legislation is not required to address copyright issues and GAI systems. We respectfully suggest the Library of Congress pursue a rulemaking to amend 37 Code of Federal Regulations (C.F.R) § 200-235 to incorporate GAI systems.

Copyright Office Feedback Request No. 6

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

Training resources for GAI system models are collected and curated in various ways. The copyright status of the training materials used is not known and should not be assumed to be copyrighted works. Training materials must be standardized into a specific structure that the GAI system can understand, which generally requires human intervention. Therefore, the use of GAI systems requires GAI System Users and GAI Developers with the ability to manipulate, clean, and modify training material data to meet the specific training material requirements of the GAI systems models.

Copyright Office Feedback Request No. 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)?"

Various training materials are gathered and curated from multiple sources to train GAI system models, which are discussed below.

GAI Developers

GAI Developers generally use publicly accessible information from the internet or other sources to gather a wide variety of training materials, such as, but not limited to, text, photos,

[17] https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex:52021PC0206
[18] https://oeil.secure.europarl.europa.eu/oeil/popups/ficheprocedure.do?lang=en&reference=2021/0106(OLP)
[19] https://www.gov.uk/government/publications/national-ai-strategy
[20] https://assets.publishing.service.gov.uk/media/614db4d1e90e077a2cbdf3c4/National_AI_Strategy_-_PDF_version.pdf

audio, video, songs, movies, articles, books, publicly accessible datasets, data generated by the GAI Developers, or training materials obtained from third-party data aggregators. GAI Developers frequently use web scraping to acquire information from internet sources. Additionally, GAI Developers frequently use third-party businesses to collect training materials. A new trend in training material acquisition for GAI system model training is for services and platforms to write a privacy policy that permits using a service or platform's usual and customary use data for GAI system model training.

GAI System Users

In general, GAI System Users are unaware of the origin of all the training materials used in the training of the GAI system. Technologically savvy GAI System Users generally use publicly accessible information from the internet or other sources to gather a wide variety of training materials, such as, but not limited to, text, photos, audio, video, songs, movies, articles, and books.

Copyright Office Feedback Request No. 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?"

The extent to which copyrighted works are licensed as GAI system model training materials varies. While Copyright Owners are beginning to license their copyrighted works for GAI system model training purposes, it is not a common transaction. Some of the licensing models that may be used include:

8. Non-exclusive license: This license permits the GAI System User or GAI Developer to use the copyrighted work for GAI system model training purposes, and the Copyright Owner retains the right to license the copyrighted work to others.
9. Perpetual, non-exclusive, sublicensable, no-charge, royalty-free, irrevocable copyright license to reproduce and prepare derivative works for public display, public performances, sublicense, and distribution, including the use of the prompts given by the GAI System User to train GAI system models.
10. Creative Commons Licenses[21]

A worldwide licensing system for copyrighted works would facilitate transparency in copyrighted works' use and access to copyrighted works to train GAI system models. We respectfully suggest a worldwide licensing system should be similar to the Madrid International Trademark System managed by the World Intellectual Property Organization (WIPO).

[21] https://creativecommons.org/licenses/list.en

Copyright Office Feedback Request No. 6.3

"To what extent is noncopyrighted 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 are the most common training material for GAI system model training, such as the Common Crawl dataset[22] and the Wikipedia dataset.[23] Public domain works in GAI system model training can vary depending on the specific use case and training material availability. Most GAI Developers use public domain works to train their GAI system models for tasks such as natural language processing and machine translation. GAI Developers also create or commission training materials for topic-specific or organization-specific GAI systems. For example, a GAI Developer building a GAI system model to diagnose diseases may make or commission a training material dataset of medical records.

Copyright Office Feedback Request No. 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."

It is common practice for GAI Developers to save some or all training materials. The rationale for document retention can be grouped into the following categories:

11. Model Improvement: Retaining the training materials allows for the continual retraining and enhancement of GAI system models. GAI Developers can retrain the GAI system model using augmented training materials when new training materials become available. Additionally, if the GAI system model's performance is subpar, the GAI Developers can use retained training materials for retraining.
12. Debugging and troubleshooting: If a GAI system model does not behave as expected, GAI Developers can use the retained training materials to analyze the error or bug to make the necessary corrections.
13. Compliance and auditing: Some service sectors require GAI Developers to retain training materials for auditing and compliance purposes. Retained training materials are used as evidence that the GAI system model was developed in accordance with all applicable service sector regulations and laws.
14. Replicability and transparency: Retaining training materials ensures the integrity and robustness of the GAI system model, which is proven through replicability of the GAI system model.

[22] https://commoncrawl.org/
[23] https://en.wikipedia.org/wiki/Wikipedia:Database_download

Copyright Office Feedback Request No. 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."

Training materials are made into datasets, which are loaded into the GAI system model for training purposes. During GAI system model training, the same dataset is used multiple times. GAI system model training does not result in the reproduction of training materials nor the storage of training materials. GAI system model training does not benefit from reproducing a specific copyrighted work. Additionally, a single copyrighted work cannot train a GAI system model.

15. Training Material Usage. GAI system models learn using datasets. Similar to students reading from textbooks, GAI system models read datasets. GAI system models read the datasets by loading them into the computer memory.
16. Nature and Duration of Training Material Reuse. Similar to students studying for an exam, GAI system models reload specific sections of datasets into the computer memory to review and analyze the information. This is similar to a student re-reading a textbook chapter in exam preparation. This duplication is temporary and is only for the duration of the GAI system model training.

Copyright Office Feedback Request No. 7.2

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

Inferences from the GAI system model training process are stored within the GAI system model as weights and biases. The GAI system model parameters are adjusted in the GAI system model training to capture and identify patterns and relationships in the training materials. Once the GAI system model is trained, the GAI system model parameters contain the learned knowledge, which allows the GAI system model to make inferences. Inferences store the high dimensional, abstract representation and mapping of the training materials or patterns learned, not direct replicas of the training materials. These inferences result in predictions or the ability to generate content.

<<Intentional blank space. See the next page.>>

Copyright Office Feedback Request No. 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?"

"Unlearning" is difficult for GAI system models. GAI system models are trained to build a high-dimensional, abstract representation and mapping of patterns while not retaining the original training materials. Therefore, it is technically challenging to identify a single data source responsible for an abstract representation. Retraining a GAI system model with modified training materials excluding a single data source might also not result in "unlearning" because the GAI system model may retain residual abstract representations from previous training. "Unlearning" would require identifying and isolating the exact influence of a single data source, which is not straightforward in complex GAI system models. Ongoing research into techniques, such as differential privacy, aims to achieve single data source identification and mitigate the influence of single data sources. However, these techniques have limitations and cannot offer complete "unlearning."

Copyright Office Feedback Request No. 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?"

It is challenging to identify with a high level of certainty whether a single copyright work was incorporated into GAI system models training materials. Ongoing research into techniques, such as model inversion[24] and membership inference attacks, aims to create inferences about training materials by analyzing the GAI system model's behavior; however, their accuracy and applicability can vary, and these techniques don't guarantee conclusive evidence of the presence or absence of a single data source.

Copyright Office Feedback Request No. 8.1

"In light of the Supreme Court's recent decisions in Google v. Oracle America and Andy Warhol Foundation v. Goldsmith, how should the ''purpose and character'' of the use of copyrighted works to train an AI model be evaluated? What is the relevant use to be analyzed? Do different stages of training, such as pre-training and fine-tuning, raise different considerations under the first fair use factor?"

The use of copyrighted works in accordance with the Fair Use factors to train GAI system models should be permitted. Different stages of training do not raise additional

[24] Model inversion is a technique where an individual tries to extract information about the training materials by analyzing a trained GAI system model. For instance, imagine someone who can guess the ingredients of a dish just by tasting it. If given a mysterious dish, they might not know the exact recipe, but they can identify its key ingredients. Similarly, with model inversion, an individual interacts with a model to deduce details about the training materials it was trained on, even without access to the original training materials dataset.

considerations under the first Fair Use factor, as the methodology used is the same for different stages. Generic GAI systems may experience challenges meeting the Fair Use factors. Topic-specific GAI systems and organization-specific GAI systems met the Fair Use factors through transformative uses to add something new with a different character while not substituting for the original use of the work; reliance on factual work; use of patterns and abstract representations; and do not harm the current market.

Copyright Office Feedback Request No. 8.5

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

The use of copyrighted works in accordance with the Fair Use factors to train GAI system models should be permitted. For the fourth Fair Use factor, the GAI system model outputs may be assessed for market competition of a particular copyrighted work or the body of works of the same author. Topic-specific GAI systems and organization-specific GAI systems may overlap in the same market for a general class of works due to the transformative uses of the high dimensional, abstract representation and mapping of training materials patterns.

Copyright Office Feedback Request No. 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)?"

To facilitate affirmative consent (opt-in) and restricted use (opt-out), a worldwide licensing system for copyrighted works would promote transparency in copyrighted works use and access to copyrighted works to train GAI system models. We respectfully suggest a worldwide licensing system should be similar to the Madrid International Trademark System managed by WIPO.

Copyright Office Feedback Request No. 9.1

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

To facilitate affirmative consent (opt-in) and restricted use (opt-out) regardless of use, a worldwide licensing system for copyrighted works would be needed to enable transparency in copyrighted works use and access to copyrighted works to train GAI system models. We respectfully suggest a worldwide licensing system should be similar to the Madrid International Trademark System managed by WIPO. A new trend in training material acquisition for GAI system model training is for services and platforms to write a privacy policy that permits using a service or platform's usual and customary use data for GAI system model training.

Copyright Office Feedback Request No. 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?"

To facilitate restricted use (opt-out), a worldwide licensing system for copyrighted works would promote transparency in copyrighted works use and access to copyrighted works to train GAI system models. We respectfully suggest a worldwide licensing system should be similar to the Madrid International Trademark System managed by WIPO.

Copyright Office Feedback Request No. 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?"

No, it is not generally feasible to get consent from copyright owners. While mid to large-size businesses have the financial means and workforce to hire legal teams to track down and obtain copyright use permissions. Entrepreneurs, start-ups, and small businesses do not have the financial means or workforce to obtain permission from copyright owners. To facilitate affirmative consent (opt-in) and restricted use (opt-out), a worldwide licensing system for copyrighted works would promote transparency in copyrighted works use and access to copyrighted works to train GAI system models. We respectfully suggest a worldwide licensing system should be similar to the Madrid International Trademark System managed by WIPO.

Copyright Office Feedback Request No. 10

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

To facilitate affirmative consent (opt-in) and restricted use (opt-out), a worldwide licensing system for copyrighted works would promote transparency in copyrighted works use and access to copyrighted works to train GAI system models. We respectfully suggest a worldwide licensing system should be similar to the Madrid International Trademark System managed by WIPO.

<<*Intentional blank space. See the next page.*>>

Copyright Office Feedback Request No. 10.2

"Is a voluntary collective licensing scheme a feasible or desirable approach? 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?"

To facilitate affirmative consent (opt-in) and restricted use (opt-out), a worldwide licensing system for copyrighted works would promote transparency in copyrighted works use and access to copyrighted works to train GAI system models. A voluntary collective licensing scheme does not align with the General Agreement on Tariffs and Trade 1994, the Agreement on Technical Barriers to Trade, the Agreement on Trade-Related Aspects of Intellectual Property Rights, the General Agreement on Trade in Services, and the Berne Convention. We respectfully suggest a worldwide licensing system should be similar to the Madrid International Trademark System managed by WIPO. Waiving the Federal Trade Commission antitrust laws will negatively impact entrepreneurs, start-ups, and small businesses who do not have the financial means or workforce to compete in negotiations of collective licenses.

Copyright Office Feedback Request No. 10.3

"Should Congress consider establishing a compulsory licensing regime? 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?"

The U.S. should not establish a U.S. compulsory licensing regime. A worldwide licensing system for copyrighted works would facilitate transparency in copyrighted works use and access to copyrighted works to train GAI system models and facilitate affirmative consent (opt-in) and restricted use (opt-out). Any licensing regime must align with the General Agreement on Tariffs and Trade 1994, the Agreement on Technical Barriers to Trade, the Agreement on Trade-Related Aspects of Intellectual Property Rights, the General Agreement on Trade in Services, and the Berne Convention.

We respectfully suggest a worldwide licensing system should be similar to the Madrid International Trademark System managed by WIPO. A worldwide licensing system would facilitate notifications to Copyright Owners and support fair compensation for Copyright Owners. A worldwide licensing scheme would enable streamlined negotiations for opt-in use and royalty arrangements between GAI Developers or GAI System Users and Copyright Owners. This benefits all parties by facilitating fair compensation for the use of copyrighted works for GAI system model training and derivative works and a mechanism for GAI Developers and GAI System Users to comply with copyright laws with minimal regulatory overhead.

Copyright Office Feedback Request No. 10.4

"Is an extended collective licensing scheme a feasible or desirable approach?"

Collective licensing schemes are not a desirable approach. Collective licensing schemes will negatively impact entrepreneurs, start-ups, and small businesses that do not have the financial means or workforce to compete in negotiations for collective licenses.

Copyright Office Feedback Request No. 10.5

"Should licensing regimes vary based on the type of work at issue?"

No, licensing regimes should not vary based on the type of copyright work at issue. This creates regulatory burdens and will negatively impact entrepreneurs, start-ups, and small businesses who do not have the financial means or workforce to negotiate a multitude of copyright work-type licensing regimes. A worldwide licensing system for copyrighted works would facilitate transparency in copyrighted works use and access to copyrighted works to train GAI system models and facilitate affirmative consent (opt-in) and restricted use (opt-out). Any licensing regime must align with the General Agreement on Tariffs and Trade 1994, the Agreement on Technical Barriers to Trade, the Agreement on Trade-Related Aspects of Intellectual Property Rights, the General Agreement on Trade in Services, and the Berne Convention. We respectfully suggest a worldwide licensing system should be similar to the Madrid International Trademark System managed by WIPO.

Copyright Office Feedback Request No. 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."

GAI system models can track the training material sources used to create output. Using "differentiable attribution[25]" allows the GAI system model to trace the contribution of each training material data source to the final output. This approach is used in natural language processing tasks, such as machine translation, where the GAI system model can determine which training materials are the most important in generating the output. A second approach would be to program the GAI system model to develop a list of the sources used to generate outputs. For example, a GAI system model that creates news articles could keep a bibliography of the training materials used to compile a list of contributors.

<<*Intentional blank space. See the next page.*>>

[25] Differentiable attribution is a technique for tracing the contribution of each input to the final output of a neural network. It is based on the idea of calculating the derivative of the output with respect to each input. This derivative represents how much the output would change if the input were changed slightly.

Copyright Office Feedback Request No. 13

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

The economic impacts of licensing requirements will vary. Increases in the cost of building GAI systems will result in higher prices for GAI systems goods and services. Without a worldwide licensing system, the development of GAI systems would be delayed. Additionally, licensing requirements will negatively impact entrepreneurs, start-ups, and small businesses that do not have the financial means or workforce to navigate complex and fragmented licensing schemes. If licensing schemes are complex and fragmented, GAI Developers may only be able to use training materials that contain biases, deepfakes, or other types of misrepresentations, such as political leanings, resulting in content with the same biases, inherent deepfakes, and misrepresentations.

Copyright Office Feedback Request No. 14

"Please describe any other factors you believe are relevant with respect to potential copyright liability for training AI models?"

17. Privacy and Data Protection Laws. GAI system model training may involve training materials that include personal or sensitive information. A new trend in training material acquisition for GAI system model training is for services and platforms to write a privacy policy that permits using a service or platform's usual and customary use data for GAI system model training.
18. Cross-Border Data Transfer. Data protection laws must be addressed if GAI system model training involves moving personally identifiable information across jurisdictional borders.

Copyright Office Feedback Request No. 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?"

The collection, retention, and disclosure of GAI system model training materials should be a best business practice (BBP) for GAI Developers. Works from GAI system models should provide references or citations to copyrighted works, similar to the bibliographies of research and academic articles.

Copyright Office Feedback Request No. 15.1

"What level of specificity should be required?"

References or citations should use the standard level of care used for bibliographies of research and academic articles. Additionally, GAI system models should further state whether

the reference or cited source was used directly, served as a base for additional modifications, or used for fine-tuning to achieve the GAI system model's expected outcomes.

Copyright Office Feedback Request No. 16

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

In cases where the use of copyrighted works falls under fair use doctrines or safe harbor provisions, notification to copyright owners may not be necessary. Without a worldwide licensing system, notifications would be difficult to complete. Notifications to Copyright Owners should be the responsibility of the licensing regime or scheme, not the GAI Developers or GAI System Users.

Copyright Office Feedback Request No. 27

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

Policymakers should consider the implications of GAI system models evolving into artificial general intelligence (AGI) that mirrors human-like creativity. AGI could produce a piece of original artwork indistinguishable from any human-made art, which introduces questions about originality and ownership. Additionally, protecting the rights of Copyright Owners and fostering technological innovation will require a balance of policy initiatives.

Copyright Office Feedback Request No. 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?"

Yes, AI-generated material from Generic GAI systems for use by Public Users should be labeled or otherwise publicly identified as being generated by GAI systems. Organization-specific GAI systems and topic-specific GAI systems for Non-Public Users should not be required to label materials generated by GAI systems. Labeling AI-generated material intended for Public Users addresses the need to quickly and confidently identify content produced by GAI systems instead of human Copyright Owners. Labeling AI-generated materials generated for use by Public Users creates an environment of transparency and trust for Public Users, GAI System Users, Copyright Owners, GAI Developers, Researchers, and technology platforms.

Copyright Office Feedback Request No. 28.1

"Who should be responsible for identifying a work as AI-generated?"

GAI System Users are responsible for identifying a work as AI-generated. However, the GAI System Users may not always know GAI generated the work as some platforms do not disclose the integration of GAI.

Copyright Office Feedback Request No. 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?"

There are a variety of tools[26] used to identify AI-generated material. However, none of these tools are managed by standard-setting bodies. The accuracy of these tools ranges from 38 percent to 84 percent, with reported false positives of human-written texts flagged as GAI; technological limitations for different GAI system models, such as GPT-4 versus GPT-3.5; difficulty detecting GAI texts combined with human text or paraphrased; difficulty detecting the use of paraphrasing tools on human-written text; difficulty detecting GAI text on specialist topics versus general topics; and the tools provide a false sense of confidence due to the misrepresentation of yes-no binary analysis as percentages. The list of tools we identified are:

19. Winston AI
20. Originality.ai
21. GPTZero
22. Sapling AI Detector
23. Writer.com AI Content Detector
24. Copyleaks
25. CrossPlag AI Content Detector
26. Content at Scale AI Text Detector
27. AISEO Content Detection

Regulosity and Pangea sincerely appreciate the Copyright Office's consideration of our comments on the notice of inquiry and request for comments titled "Artificial Intelligence and Copyright," docket no. 2023-6. If there are any questions regarding the comments or information provided, please contact Mrs. Misty Guard, Regulosity LLC, at +1 (414) 988-8613 or Mr. Vivek Nithianand, Pangea Tech Solutions Pvt Ltd, at +9 18023683393.

Sincerely

Misty Guard
President
Regulosity LLC

Vivek Nithianand
CEO
Pangea Tech Solutions Pvt Ltd

[26] https://www.scribbr.com/ai-tools/best-ai-detector/