

gettyimages®

iStock™

Unsplash

European Commission
B-1049 Brussels Belgium
COMP-GENERATIVE-AI@ec.europa.eu

11 March, 2023

Re: Competition in Generative AI – Call for Contributions

Best Regards,

Getty Images appreciates the opportunity to provide our perspective on competition in the context of generative AI. We are a preeminent global visual content creator and marketplace. With more than 825,000 active customers from almost every country in the world, Getty Images works with businesses of all types and sizes to connect and compete in an increasingly digital and visual world. We offer a growing library of over 551 million visual assets (image and video) that delivers unmatched depth, breadth, and quality. Our library represents the work of more than 551,000 contributors and we are the partner of choice to major companies and organizations in the global editorial and corporate sectors. Each year we cover more than 160,000 news, sport and entertainment events around the globe.

Getty Images is also an active participant in the generative AI ecosystem. Our large library of high-quality visual assets paired with high-quality metadata are valuable assets in the context of generative AI and we act as a licensor in the market for AI training data. We have seen growing demand for Getty Images content by generative AI developers and we have entered into dataset licensing deals with generative AI developers who choose to source their training data fairly. In 2023, we launched “Generative AI by Getty Images”, a commercially safe generative AI tool, trained exclusively using Getty Images best-in-class creative library, and combined with the latest AI technology, and proven computing infrastructure from NVIDIA. The tool, available on Gettyimages.com and via API, allows customers to use text-to-image generation to ideate and create compelling visuals, that include uncapped indemnification and perpetual, worldwide usage rights. We share the revenue generated by

the commercialization of “Generative AI by Getty Images” with our world-class content creators, allowing them to continue to create more of the high-quality pre-shot imagery for which Getty Images is known and on which our customers depend.

We believe that generative AI Models can bring significant benefits to consumers and creators alike. However, we also recognize that the technology introduces risks to consumers, creators, and the public interest. A key risk that needs to be considered in the context of competition is liability for unauthorized use of protected data as model training data. Unfortunately, many generative AI developers train their models on unauthorized web-scraped data and infringe intellectual property and other third-party rights. Such brazen disrespect for intellectual property illegally cuts development costs and unfairly puts generative AI developers who source training data responsibly at a clear competitive disadvantage.

Our submission focuses on how the use of copyright protected content as training data for generative AI models without authorization unfairly harms competition. We have chosen to provide answers to questions 2 and 6 of the Commission’s questionnaire to illustrate the issues that are most relevant to our business.

We thank the Commission for the opportunity to participate in this call for contributions. Please feel free to contact us via the information provided in the cover e-mail if you have any questions or would like us to provide additional evidence. In addition, we are eager to participate in relevant workshops that the Commission may organize. Any invitation to contribute would be greatly appreciated.

COMMISSION PROMPTS – GENERATIVE AI

2) What are the main barriers to entry and expansion for the provision, distribution or integration of generative AI systems and/or components, including AI models? Please indicate to which components they relate.

There is currently a major barrier to entry in the market for a critical AI component, that of the licensing of training data. This is due to the unequal bargaining power between leading generative AI developers and the owners of training data, especially where this data has been scraped from the internet and therefore originates from a multitude of sources. Such developers have commonly taken the position that authorization of data owners is not necessary as it is either too difficult to obtain or is not needed. This position is unsupported by case law and is the subject of numerous copyright infringement actions, including actions being brought by Getty Images against a major

generative AI model provider in the UK and the US. However, pending trial of these actions, which may take several years to conclude, major AI model developers are brazenly continuing to use unauthorized web-scraped, which is having the effect of normalising such activity. This makes it difficult for creators of valuable training data to participate in the generative AI ecosystem and to benefit from the exploitation of products that are trained on, and in fact compete with, their work. This is despite such well-funded AI developers paying vast sums for other main key components including computer processing and engineering talent. Such developers of generative AI models who brazenly use web-crawled data for training are effectively foreclosing data licensing markets, by restraining them from growing at the same pace as the underlying AI technology.

In a fair market, suppliers of a valuable resource should be able to obtain a fair price from users who demand that resource. In the context of generative AI, developers who use unauthorized web-scraped data are essentially stealing that resource and unfairly benefiting from it.

It does not have to be this way. Getty Images has partnered with NVIDIA to launch “Generative AI by Getty Images”, a commercially safe generative AI tool, trained exclusively off Getty Images best-in-class creative library, where images featuring people are model-released. This model and other generative AI models that were fairly trained on authorized data, are evidence that it is possible to develop innovative generative AI models and systems while also respecting 3rd party rights. Technology companies should not be permitted to use their market power to keep rights holders from participating in data licensing markets in the context of generative AI.

6) Do open-source generative AI systems and/or components, including AI models compete effectively with proprietary AI generative systems and/or components? Please elaborate on your answer.

There is clear evidence that participants in the ecosystem for open-source generative AI systems and/or components are not competing fairly with developers, deployers and users of proprietary systems. This is because open-source pre-trained foundation generative models are more commonly trained on unauthorized web-scraped data, which gives their deployers a distinct cost advantage over any developer/deployer of a proprietary AI generative system who does chose to responsibly source and pay for training data.

The demand for high-quality visual content paired with high-quality metadata is not unique to developers of open-source generative AI systems. In the context of generative AI models used to generate synthetic visual content, both proprietary and open-source models rely on training data that is often copyright protected and contains personal data, e.g. in the form of photos depicting real people. While both proprietary and open-source AI systems are known to be trained on unauthorized content, open-source foundation generative AI models pose a unique risk because liability for illegally using protected data is diffused throughout the open-source ecosystem. The open-source ecosystem facilitates cheap and easy deployment and distribution of foundation models, making it even more difficult for the owners of unlicensed training data (consisting of copyright works and personal data) to enforce their rights against downstream deployers and users of such models. An entity who deploys a pre-trained open-source generative AI model benefits from the training data but, the diffused and anonymous nature of the ecosystem makes it difficult for damaged third parties to enforce their rights against such a deployer. In addition, developers of foundation open-source generative AI models often claim that they are not liable if they are not directly monetizing the models. Accordingly, and by design, downstream participants in the open-source ecosystem also claim that they are not liable for infringing on 3rd party rights, leaving damaged parties without remedy and developers who responsibly and fairly source and use data at a distinct disadvantage. Finally, if an open-source model is held to be infringing, it is nearly impossible to stop continued use and/or dissemination. Unlike with a proprietary model, where owners may be enjoined, once an open-source model has been released there is no single entity that can be targeted and made to cease and desist.

Again, it does not have to be this way. It is possible for participants in the open-source generative AI ecosystem to act in a lawful manner. Data licensing markets do exist and have the potential to grow in line with the advancement of AI systems. However, until all participants take responsibility for the related costs, the open-source ecosystem is being unjustly built at the expenses of creators, consumers, and the public interest.

Fortunately, the Commission's recent adoption of the Artificial Intelligence Act (AI Act) lays the groundwork for the creation of a fairer playing field. In particular, the transparency obligations for providers of general-purpose AI models under Article 52c are expected to promote the legal sourcing and use of training data. Even though the AI Act exempts AI models made accessible to

the public under a free and open license from certain transparency obligations, Sections 1(c) and 1(d) apply to developers of both proprietary and open-source generative AI systems and obligate them to put in place a policy to respect copyright law as well as make available a detailed summary about the content used for training. While not a complete solution, such transparency should make it easier for rightsholders to control how their IP and private data is used and to obtain appropriate licensing compensation. In the context of the development and deployment of open-source generative AI models, transparency obligations will give participants in the ecosystem greater visibility into how generative AI models were trained and they will be on notice of potential related liability. These insights will enable members of the open-source ecosystem to proactively enter into licensing deals that give them the rights they need to safely deploy commercial generative AI systems.

To be clear, implementation of the AI Act is not enough to solve the unique problems posed by open source generative AI. The next logical step is to introduce competition law policy that is cognizant and complementary to existing intellectual property laws such that participants are encouraged to obtain necessary rights and to compensate creators for the essential contributions they have made. Policy that discourages the anti-competitive nature of using unauthorised web-scraped data will ultimately benefit the related data licensing market and encourage innovation that is responsible, legal, and fair.