

Getty Images (UK)—written evidence (FON0043)

House of Lords Communications and Digital Select Committee inquiry: The future of news: impartiality, trust, and technology

Thank you for calling for evidence in connection with your Future of News Inquiry. Getty Images agrees with the Committee that a healthy news ecosystem is vital to our democratic society. As a leading supplier of imagery to the news sector, we are well placed to comment on some of the key strategic challenges facing the sector. While all of the issues being investigated are important, our response is focused on the issue of trusted information and the impact caused by generative AI.

Established in the UK and now listed on the NYSE, Getty Images is a preeminent global visual content creator and marketplace. The content we distribute includes a wide range of news, sport, entertainment and archive images and video clips. Our editorial services have made us the partner of choice to major organizations in the global editorial space due to our photographic talent, expert editorial operations teams and our unrivaled global distribution platform. We are also custodians of an extensive and comprehensive historic archive with over 135 million images dating back to the beginning of photography. Our editorial and archival content is trusted to inform people, public policy, services and markets of past and present newsworthy events.

With regard to our editorial portfolio of content, we are committed to ensuring that high editorial standards are maintained¹. Getty Images does not produce or distribute editorial content that has been created or augmented using generative Artificial Intelligence models. Editorial content creators are held responsible for the truthful, impartial and accurate representation of the reality of the events and scenes they witness. They are held to be the first and principal fact checkers. The distribution as part of our editorial offering of imagery or caption text that has been synthetically created is strictly impermissible. Our editorial image partners and providers are also required to uphold the same standards and practices as our in-house editorial content teams.

As well as editorial imagery, Getty Images also distributes creative imagery, referred to by some as stock imagery, which is suitable for use in marketing and advertising. A relatively new subset of creative content is AI generated content, and last year Getty Images launched its own "AI Generator", now available from both its www.gettyimages.co.uk and www.istock.co.uk websites.

At Getty Images, we take care to ensure that our customers can easily distinguish between editorial and creative content. We include clear navigation on our websites and each item of content is clearly labelled. However, when content leaves our website, the extent to which it's applicable genre - editorial or creative - is able to be discerned, depends on the publisher of that content. At one end of the spectrum, viewers of newspaper websites that operate according

¹ See the Getty Images Editorial Standards here: <https://www.gettyimages.co.uk/company/editorial-policy>

to established principles of editorial integrity can generally be assured that what they are viewing is “real”, as they know that the newspaper’s reputation depends on it having done appropriate due diligence regarding the provenance of content. However, at the other end of the spectrum, publishers such as social media websites and aggregator websites, generally do not undertake such due diligence prior to publication and therefore the task of discerning what is “real” versus “unreal” is more difficult for the viewer, and this task is becoming exponentially more difficult with the advent of AI generated content.

Generative AI Content Compared to Primary Source Journalism Content

Generative AI has introduced significant risks to the news ecosystem and, as further discussed below, transparency regulation is a key tool that can be used to limit damage. We know that generative AI tools are widely available and may be used to produce and distribute synthetic content at mass scale and low cost. Many of these tools rely on foundation models trained on unlicensed proprietary content, including “primary source journalism” content, and are starting to compete directly with the producers of such content.² By “primary source journalism”, we mean professionally produced, first-hand witness-based journalism, of the type that we invest in producing and distributing, including from about 160,000 editorial events that we cover each year. Getty Images is a proud creator and distributor of primary source journalism and supports the response to this call for evidence submitted by the News Media Coalition (NMC)³. Unfortunately, deepfake content is easily created using generative AI and such content has already led to a rise in the distrust of media. Combining that technology with the ability to distribute such content via Internet platforms that are already struggling to identify which content is authentic, can meaningfully undermine public fluency in facts, trust in institutions, and national security. Although Getty Images has developed and distributes our own Generative AI models⁴, these are designed to generate only creative outputs and have safeguards built in to ensure that its outputs do not resemble real people. These Generative AI models are trained only on licensed, creative images. They have no knowledge of real-life celebrities, politicians or athletes, therefore are not susceptible to generating deepfake content.

Responses to Key Questions

Our responses are focused on issues raised by generative AI and are not intended to answer all of the questions raised.

² See US Federal Case NYT v OpenAI and Microsoft for evidence of how generative AI is competitive with trusted media industry . <https://www.theguardian.com/media/2023/dec/27/new-york-times-openai-microsoft-lawsuit>

³ See NMC response for a discussion of primary source journalism (PSJ)

⁴ See <https://www.gettyimages.com/ai/generation/about>

Trends over the next 12 months and 5 years

2. How is generative AI affecting news media business models and how might this evolve?

Democracy is underpinned by trust in a free, reliable, independent, and strong media ecosystem. If development and deployment of generative AI is left unregulated these technologies have the potential to irreparably damage the news industries and threaten the sustainability of the ecosystem as a whole, by significantly eroding the public's trust in the independence and quality of content available online, and also threatening the financial viability of news providers.

Generative AI has unleashed a flood of disinformation and synthetic content that has already started to undermine the public's trust in institutions and each other. Globally, there have been numerous examples of deepfakes that have effectively deceived the media and the public eroding trust.⁵ There is also a rising concern that deepfakes can be used to maliciously influence elections and attack democratic processes, for example deepfake robocalls imitating US President Biden have already been used to try and influence the US Presidential election.⁶

Responsible media companies have been forced to bear an additional burden when verifying sources. In order to maintain the high level of editorial integrity that our customers demand, we are continually investing in the creation and sourcing of trustworthy editorial content. But there is a risk that such content will be modified using generative AI tools to tell a different story. We expect that the generative AI arms race between the makers of disinformation and purveyors of newsworthy content is only just getting started.

4b) What impact do concerns around disinformation have on trust in the information environment? (And to what extent does this differ between different sections of society?)

A flood of AI generated content can distort facts and leave the public with no basis to discern what is true and what is made up. Even absent malicious intent, many generative AI applications that are trained indiscriminately on scraped content including editorial content produce factual errors and fictional information, in addition to propagating long-standing biases. Internet platforms that invite the public to upload content compound the problem by enabling widespread distribution and promotion of disinformation, further eroding trust. Over the last three decades, such platforms have demonstrated their inability to take responsibility for the content they distribute. With the advent of generative AI, the laissez-faire approach to content moderation has caused the erosion of trust to accelerate.

Generative AI does not have to be this way. Generative AI models can be responsibly trained on non-editorial licensed content and be used for

⁵ See list of deepfakes that went "viral" in 2023 by Deanna Ritchie <https://readwrite.com/deepfake-ai-generated-images-that-went-viral-in-2023/>

⁶ See <https://www.npr.org/2024/02/08/1229641751/ai-deepfakes-election-risks-lawmakers-tech-companies-artificial-intelligence>

commercially safe purposes. When developing our “AI Generator”, Getty Images partnered with leading AI developer Nvidia to train on only our creative content that has been cleared for commercial non-editorial use.⁷ Because our model was not trained on editorial content, it is not possible to use our generative AI tools to create deep fakes that portray public figures in a deceiving way.

Evaluation

1. How well are news organisations responding to factors affecting their business models, and are any changes needed?

The UK news media ecosystem has a history of embracing and successfully navigating new technology, from the introduction of the printing press, to broadcast media, to the internet and social media. However, the pace of development of generative AI far exceeds that of prior technological leaps, and the impact has been quick and far reaching. Transparency regulation that requires transparency standards and encourages trust (as further described below) are needed to support the news ecosystem and encourage continued investment in the creation of high-quality and trustworthy news media content.

5. Are there any actions the Government should take to address concerns around due impartiality, trust, and the influence of technology platforms?

Getty Images recommends that the Committee considers the following areas as it seeks to identify short and long term solutions to address the threat of generative AI to the news industry:

Transparency:

Transparency is key. In the context of Generative AI, we would like to see developers and distributors being obligated to meet transparency standards related to training data as well as the identification of generative output.

All participants in the digital services ecosystem, from the start of the AI life cycle onwards, should be required to uphold transparency best practices. Getty Images would urge UK regulators to require all persons who make generative AI models, platforms, or technologies available on the market, to develop their tools so that all content generated or modified by an AI tool, not only deepfakes, is clearly identified as such. An example of how this can work in practice is by use a metadata signal indicating that content is AI generated. For example, the IPTC “Digital Source Type” property is an emerging standard for this type of metadata signal.

In addition to regulation that requires the marking of AI generated content, it will also be helpful if regulation can be introduced that legally incentivises the use of licensed content, rather than unlicensed scraped content, when used as training data. Moreover, developers of AI models and creators of training

⁷ <https://www.theverge.com/2024/1/8/24027259/getty-images-nvidia-generative-ai-stock-photos>

datasets should be required to collect, retain and disclose auditable records regarding the sources of materials used to train AI models.

At a minimum, such auditable records could include:

- Disclosure of all public and private datasets that are likely to contain copyright protected work that have been used to train a generative AI model. Disclosure should include records of the name of the datasets used and a short description of each. Where a dataset includes unlicensed scraped content, the URL from where each item of content as scraped should also be disclosed.
- Grounds upon which copyright protected work is used, including:
 - whether licenses have been acquired to use any datasets that include copyrighted work; and
 - any legal justification for use without a license

Fair & Responsible Licensing:

Producers of editorial news content need to be able to negotiate on a level playing field with AI model operators and developers regarding access to and use of their intellectual property. Those participants in the news ecosystem that produce primary source journalism must have the choice about whether to allow their content to be used to train generative AI. This will enable them to insist on safeguards to ensure that generative AI outputs complement rather than compete with their own news services. If they elect to allow their content to be used, they should be compensated for the use of their works. In addition, in view of the fact that generative AI models are often trained on data obtained from a variety of sources, UK AI regulators should promote an environment where rights holders are able to work together to fairly license data, including with regard to licensing structures, without being held back by UK competition law.

Down Stream Regulation:

All generative AI model and digital platform providers should have a responsibility to address misinformation, bias and misuse from and of their services. Without regulation requiring this, technology companies are unlikely to voluntarily regulate their own actions when often the most outrageous and controversial content is what attracts the most views and advertising revenue, irrespective of whether it is real.

12 February 2024