

Back UK creative sector or gamble on AI, Getty Images boss tells Sunak

Publication Date: 2024-01-14

Author: Alex Hern

Section: Technology

Tags: Artificial intelligence (AI), Computing, New York Times, Newspapers & magazines, US press and publishing, Intellectual property, Photography, news

Article URL: <https://www.theguardian.com/technology/2024/jan/14/back-uk-creative-sector-or-gamble-on-ai-getty-images-boss-tells-sunak>



Rishi Sunak needs to decide whether he wants to back the UK's creative industries or gamble everything on an artificial intelligence boom, the chief executive of Getty Images has said. Craig Peters, who has led the image library since 2019, spoke out amid growing anger from the creative and media sector at the harvesting of their material for "training data" for AI companies. His company is suing an AI image generator in the UK and US for copyright infringement. "When I look at the UK, probably about 10% of its GDP is sitting in the creative industries, whether that's movies, music, television. I think making that trade-off is risky. If I'm the UK, betting on AI, less than a quarter point of GDP within the UK today, significantly less than the creative industries, is a bit of a perplexing trade-off." In 2023, the government set out its goal to "overcome barriers that AI firms and users currently face" in using copyrighted material in response to a consultation from the intellectual property office, and it committed to support AI companies "to access copyrighted work as an input to their models". That was already a step back from an earlier proposal for a broad copyright exception for text and data mining. In a response to a Commons committee on Thursday, Viscount Camrose, the hereditary peer and parliamentary under-secretary of state for artificial intelligence and intellectual property, said: "We will take a balanced and pragmatic approach to the issues that have been raised, which helps secure the UK's position as a world leader in AI, whilst supporting our thriving creative sectors." The role of copyrighted work in AI training has come under increased pressure. In the US, the New York Times is suing OpenAI, the maker of ChatGPT, and Microsoft for using its news stories as part of the training data for their AI systems. Although OpenAI has never revealed what data it used to train GPT4, the newspaper was able to get the AI system to spit out verbatim quotes of NYT articles. In a court filing, OpenAI said it was impossible to build AI systems without using copyrighted materials. "Limiting training data to public domain books and drawings created more than a century ago might yield an interesting experiment, but would not provide AI systems that meet the needs of today's citizens," the organisation added. Peters disagrees. Getty Images, in collaboration with Nvidia, has created its own image generation AI, trained exclusively on licensed imagery. "I think our partnership speaks exactly counter to some of the arguments that are put out there that you couldn't have these technologies with a licence requirement. I don't think that's the case at all. You need to take different tacks, different approaches, but the notion that there isn't the capability to do that, that's just smoke." Even within the industry, the tide is turning. A dataset of pirate ebooks called Books3, hosted by an AI group whose copyright takedown policy was at one point a video of a choir of clothed women pretending to masturbate their imaginary penises while singing, was quietly removed from download

after an outcry from the authors contained in it – but not before it had been used to train, among others, Meta’s LLaMa AI. As well as lawsuits by Getty Images and the New York Times, a host of other legal actions are progressing against AI companies over potential infringement in their training data. John Grisham, Jodi Picoult and George RR Martin were among 17 authors who sued OpenAI in September alleging “systematic theft on a mass scale”, while a group of artists filed a suit against two image generators in January last year, one of the first such cases to enter the US legal system. Ultimately, how courts or even governments decide to regulate the use of copyrighted material to train AI systems may not be the final word on the matter. A number of AI models, both text-generating LLMs and image generators, have been released “open source”, free to download, share and reuse without any oversight. A bar on using copyrighted material to train new systems will not scrub those from the internet, and will do little to prevent individuals from using new material to retrain, improve and re-release them in the future. Peters is optimistic that the result is not a foregone conclusion. He said: “Those that produce and distribute the code, they ultimately have legal entities and they are subject to that. The question of what you’re running on your laptop or your phone may be a bit more of a question, but there’s individual responsibility there.”