



AI vs. IP: The Legal Battle Over Generative Artificial Intelligence

Imagine the morning of November 30, 2022. The chipmaker NVIDIA had a market capitalization of roughly \$385 billion. Later that day, artificial intelligence (AI) research lab OpenAI first released ChatGPT, its AI-powered chatbot, to the public. By mid-June of 2023, NVIDIA's valuation had risen to \$1 trillion, propelled almost entirely by demand for its data center products — vital components in building the infrastructure needed to support widespread use of generative AI. In late February of this year, it broke \$2 trillion. Over that time, the usage and visibility of AI has seen a similarly dramatic rise.

However, while technology may move fast, law moves slow. The sudden and rapid ascent of AI to the forefront of both the public's and Wall Street's consciousness has outpaced established law, particularly with regards to intellectual property (IP). What the courts eventually find in this unknown legal territory will have profound implications for the development and use of AI and the future of creative industries.

WHAT GOES INTO GENERATIVE AI

Perhaps the most pressing of the numerous intellectual property issues around generative artificial intelligence programs centers on how they're made. Generative AI models are programs that are capable of outputting custom-made text, images, audio, and video in response to a user's inputs. While the details of how such programs work lie beyond the scope of this article, in broad strokes, they are made by feeding a massive number of examples of the types of content one wants them to produce into an algorithm.

This algorithm identifies common features of the samples given to it. For example, a program shown many, many pictures of cats, all of them labeled with the word 'cat,' would identify the common visual features of that set of images — all of the things that make a cat look like a cat — and associate them with the word 'cat.' Then, when a user inputs the word 'cat' back into the program, it would spit out an image with all of those same common visual features: a reconstituted picture of a cat.

Of course, popular generative AI programs can do far more than just produce pictures of animals. They also require the use of a massive amount of data, known as training data, to create. Fortunately for AI labs, they have access to the largest repository of available data in human history: the internet. Using automated programs, the makers of generative AI can harvest content from the internet and feed it into their platforms.

Unfortunately for AI makers, much of that data is protected by copyright laws. The holders of the copyrights for that content aren't compensated for its use, and with their property now serving as the foundation of somebody else's lucrative business model, many have decided to sue. Some of the earliest such cases were brought by independent illustrators fearful of being put out of work by programs trained on their own art, but more recently, larger enterprises such as the New York Times and stock image provider Getty Images have entered the fray.

WHAT COMES OUT OF GENERATIVE AI

AI companies have responded to suits by arguing that their programs fall under the fair use doctrine: because an individual piece of training data contributes only minutely to a given output, and because the result is substantially different from the original, their use is legal. This could be a strong argument, but the actual outputs of generative AI programs may undercut it.

In December, a team of researchers published a paper titled “Scalable Extraction of Training Data from (Production) Language Models,” in which they demonstrated that by prompting ChatGPT to output the word “poem” an infinite number of times, they could break the model and instead cause it to output unaltered sections of its training data. This proved definitively that ChatGPT retained its training data, unaltered, and that it could be made to reproduce it.

This study is not a unique point against the fair use case for generative AI. Although singular in that the researchers managed to extract training data entirely unaltered, previous methods had been successful in prompting generative AI programs into giving outputs that differed from their training data in only minor details — not enough to definitively head off claims of infringement.

CONCLUSION

Generative AI is not capable of reasoning. It can be restricted by its creators from giving certain outputs, such as not outputting unaltered training data, but it cannot understand why it shouldn't do so. If it is later prompted to produce an undesirable result in a way its creators failed to guard against, it will.

Due to the nature and complexity of these programs, it is likely impossible for their creators to anticipate every circumstance that might cause them to spit out infringing data, and as such, it will be impossible to ever completely prevent it. Thus, the only things that could ever fully shield AI developers from liability for such infringement would be specific legal protections, established either through legislature or the courts, or if they had the rights to all of the data used to train their programs.

Of course, paying to license such a volume of data risks being quite expensive — unless the amount of data needed can be reduced. Due to the constraints of IP law, the eventual winner of the AI wars may not be whoever can build the smartest model, but whoever can build the most efficient one.

At press time, the author did not hold shares in any of the companies mentioned.

Commentaries like this one, as well as a host of other features, such as regularly updated analyst commentaries for each of the most actively traded stocks on the North American exchanges in more than 90 industries, are included as part of *The Value Line Investment Survey*[®]. For more information, or to subscribe, visit valueline.com or call **1-800-VALUeline** (1-800-825-8354).

Originally published in the April 12, 2024 issue of *The Value Line Investment Survey*.

Factual material is obtained from sources believed to be reliable and is provided without warranties of any kind. THE PUBLISHER IS NOT RESPONSIBLE FOR ANY ERRORS OR OMISSIONS HEREIN OR ANY DAMAGES OR LOSSES ARISING FROM ANY USE OF THE INFORMATION CONTAINED HEREIN. This publication is strictly for subscriber's own, non-commercial, internal use. No part of it may be reproduced, stored or transmitted in any printed, electronic or other form, or used for generating or marketing any printed or electronic publication, service or product. Officers, directors, or employees of Value Line, Inc. and its affiliates and subsidiaries, and of EULAV Asset Management, may own stocks that are featured in this publication. Nothing herein should be construed as an offer to buy or sell securities or to give individual investment advice.

©2024 Value Line, Inc. All Rights Reserved. Value Line, the Value Line logo, The Value Line Investment Survey, The Most Trusted Name In Investment Research, “Smart research. Smarter investing.”, Timeliness, and Safety are trademarks or registered trademarks of Value Line, Inc. and/or its affiliates in the United States and other countries. All other trademarks are the property of their respective owners.