Volume 6, No. 4 | July–August 2023

RAIL

The Journal of Robotics, Artificial Intelligence & Law

Editor's Note: AI
Victoria Prussen Spears

AI as Inventor or Author—Developing Trends
Paul Ragusa and Nick Palmieri

AI Systems Versus Copyright Protection: Exactly Where Should We Draw the Lines?
Randall K. McCarthy

Trusting AI Systems
James W. McPhillips, Elizabeth Zimmer, Sandro Serra, and Mia Rendar

AI-Created Content of the Future: Helpful or Harmful?
Neil Sahota

Everyone's Talking AI, Including the Federal Trade Commission: Key Takeaways from the FTC's 2023 AI Guidance
Christiana State, Preetha Chakrabarti, Dalton Hughes, and Sarah Rippy

Using AI for Competitive Advantage
David L. McCombs, Dina Blikshteyn, Eugene Goryunov, and Nicolette Nunez

Why a Data Scientist Needs a Lawyer to Correct Algorithmic Discrimination
Bradley Merrill Thompson

"Take Two MOBAs and Call Me in the Morning"—Video Games May Soon Be Prescribed to Treat Medical Conditions
Stuart Irvin, Rujul Desai, Olivia Dworkin, and Beth Braiterman

Treating Crypto Fairly: New UK Government Consults on a Comprehensive Regulatory Regime for Cryptoassets
Richard Frase, Daniel Natoff, and Simon Wright

fastcase FULL COURT PRESS

RAIL
The Journal of Robotics, Artificial Intelligence & Law
Volume 6, No. 4 | July–August 2023

229 **Editor's Note: AI**
Victoria Prussen Spears

233 **AI as Inventor or Author—Developing Trends**
Paul Ragusa and Nick Palmieri

239 **AI Systems Versus Copyright Protection: Exactly Where Should We Draw the Lines?**
Randall K. McCarthy

247 **Trusting AI Systems**
James W. McPhillips, Elizabeth Zimmer, Sandro Serra, and Mia Rendar

257 **AI-Created Content of the Future: Helpful or Harmful?**
Neil Sahota

265 **Everyone's Talking AI, Including the Federal Trade Commission: Key Takeaways from the FTC's 2023 AI Guidance**
Christiana State, Preetha Chakrabarti, Dalton Hughes, and Sarah Rippy

271 **Using AI for Competitive Advantage**
David L. McCombs, Dina Blikshteyn, Eugene Goryunov, and Nicolette Nunez

277 **Why a Data Scientist Needs a Lawyer to Correct Algorithmic Discrimination**
Bradley Merrill Thompson

287 **"Take Two MOBAs and Call Me in the Morning"—Video Games May Soon Be Prescribed to Treat Medical Conditions**
Stuart Irvin, Rujul Desai, Olivia Dworkin, and Beth Braiterman

293 **Treating Crypto Fairly: New UK Government Consults on a Comprehensive Regulatory Regime for Cryptoassets**
Richard Frase, Daniel Natoff, and Simon Wright

EDITOR-IN-CHIEF

Steven A. Meyerowitz
President, Meyerowitz Communications Inc.

EDITOR

Victoria Prussen Spears
Senior Vice President, Meyerowitz Communications Inc.

BOARD OF EDITORS

Melody Drummond Hansen
Partner, Baker & Hostetler LLP

Jennifer A. Johnson
Partner, Covington & Burling LLP

Paul B. Keller
Partner, Allen & Overy LLP

Garry G. Mathiason
Shareholder, Littler Mendelson P.C.

Elaine D. Solomon
Partner, Blank Rome LLP

Linda J. Thayer
Partner, Finnegan, Henderson, Farabow, Garrett & Dunner LLP

Edward J. Walters
Chief Executive Officer, Fastcase Inc.

John Frank Weaver
Director, McLane Middleton, Professional Association

THE JOURNAL OF ROBOTICS, ARTIFICIAL INTELLIGENCE & LAW (ISSN 2575-5633 (print) /ISSN 2575-5617 (online) at $495.00 annually is published six times per year by Full Court Press, a Fastcase, Inc., imprint. Copyright 2023 Fastcase, Inc. No part of this journal may be reproduced in any form—by microfilm, xerography, or otherwise—or incorporated into any information retrieval system without the written permission of the copyright owner. For customer support, please contact Fastcase, Inc., 729 15th Street, NW, Suite 500, Washington, D.C. 20005, 202.999.4777 (phone), or email customer service at support@fastcase.com.

Publishing Staff
Publisher: Morgan Morrissette Wright
Production Editor: Sharon D. Ray
Cover Art Design: Juan Bustamante

Cite this publication as:

The Journal of Robotics, Artificial Intelligence & Law (Fastcase)

This publication is sold with the understanding that the publisher is not engaged in rendering legal, accounting, or other professional services. If legal advice or other expert assistance is required, the services of a competent professional should be sought.

Copyright © 2023 Full Court Press, an imprint of Fastcase, Inc.

All Rights Reserved.

A Full Court Press, Fastcase, Inc., Publication

Editorial Office

729 15th Street, NW, Suite 500, Washington, D.C. 20005
https://www.fastcase.com/

POSTMASTER: Send address changes to THE JOURNAL OF ROBOTICS, ARTIFICIAL INTELLIGENCE & LAW, 729 15th Street, NW, Suite 500, Washington, D.C. 20005.

Articles and Submissions

Direct editorial inquiries and send material for publication to:

Steven A. Meyerowitz, Editor-in-Chief, Meyerowitz Communications Inc., 26910 Grand Central Parkway, #18R, Floral Park, NY 11005, smeyerowitz@meyerowitzcommunications.com, 631.291.5541.

Material for publication is welcomed—articles, decisions, or other items of interest to attorneys and law firms, in-house counsel, corporate compliance officers, government agencies and their counsel, senior business executives, scientists, engineers, and anyone interested in the law governing artificial intelligence and robotics. This publication is designed to be accurate and authoritative, but neither the publisher nor the authors are rendering legal, accounting, or other professional services in this publication. If legal or other expert advice is desired, retain the services of an appropriate professional. The articles and columns reflect only the present considerations and views of the authors and do not necessarily reflect those of the firms or organizations with which they are affiliated, any of the former or present clients of the authors or their firms or organizations, or the editors or publisher.

QUESTIONS ABOUT THIS PUBLICATION?

For questions about the Editorial Content appearing in these volumes or reprint permission, please contact:

Morgan Morrissette Wright, Publisher, Full Court Press at mwright@fastcase.com or at 202.999.4878

For questions or Sales and Customer Service:

Customer Service
Available 8 a.m.–8 p.m. Eastern Time
866.773.2782 (phone)
support@fastcase.com (email)

Sales
202.999.4777 (phone)
sales@fastcase.com (email)

ISSN 2575-5633 (print)
ISSN 2575-5617 (online)

AI Systems Versus Copyright Protection: Exactly Where Should We Draw the Lines?

Randall K. McCarthy*

In this article, the author advocates changes to the U.S. copyright law to provide clear guidance on what can constitute training data, how copyrighted works can be used by an internal artificial intelligence (AI) system, and the extent to which AI-generated output content can be protected by copyright.

Artificial intelligence (AI) systems are becoming increasingly common in areas relating to human creativity. Many forms of content can be generated by AI systems, including text-based content, audio, video, images, programming code, technical papers, and so on. AI systems have the ability to learn from vast amounts of data, identify patterns, and generate new content based on those patterns. As AI systems continue to become more sophisticated, these systems will continue to raise fundamental issues regarding copyright protection for both the output content generated by the systems and the input data used to generate such content.

U.S. copyright law[1] covers original works of authorship created by a human being. Only a small amount of originality and creativity is necessary, but court cases and U.S. Copyright Office guidelines alike uniformly require the originality and creativity must come from a human. Content created exclusively by an AI system, like other non-human sources, is not eligible for copyright protection.

However, AI-generated content is seldom created without the use of some form of copyrighted material that was created by a human. This is true whether such use is authorized or not. Copyrighted material can exist in the training data used to train an internal AI algorithm, in the input data supplied by a particular user to generate a desired output, within the programming structure of the AI algorithm itself, in the stored parameters (including diffusion models) utilized by the AI algorithm, in adjustments made by a user to modify or tweak an interactive AI process to generate a final desired output, and in post-curation processes where AI-generated

Robotics, Artificial Intelligence & Law / July–August 2023, Vol. 6, No. 4, pp. 239–245.
© 2023 Full Court Press. All rights reserved.
ISSN 2575-5633 (print) /ISSN 2575-5617 (online)

content that subsequently modified by a human user to create the final desired work.

In these and other areas, the question becomes where to draw the line between the rights of owners of copyrighted works and the rights of individuals to use such works to create new works. The existing copyright laws are helpful to a point, but can only take us so far in resolving these issues. Further action is desperately needed.

Fair Use and Transformation

A copyright grant includes rights to display the copyrighted work and the right to make derivatives of the work.[2] These rights held by the copyright owner are offset by the rights of others to use the copyrighted work in an authorized way, such as through the rights of parody, First Amendment expression, fair use, and transformation. Parody and First Amendment concerns are present but are not issues that often arise in an AI context. Fair use and transformation, however, are central to this analysis. Fair use is codified at 35 U.S.C. 107, and the doctrine of transformation has been established via case law as an expansion of the first element of fair use. These areas of existing law can be helpful in determining where to draw the line between copyright owners and AI system–generated content.

Fair use involves a number of factors, but primarily relates to the nature and extent of the use and the impact that the use has on the market for the copyrighted work. The U.S. Supreme Court is currently evaluating an important fair use case, *Andy Warhol Foundation for the Visual Arts v. Goldsmith*;[3] this case is not specifically directed to AI-generated content, but it will provide further guidance in the area of digital images and fair use.

Transformation relates to the extent to which a copyrighted work has been modified. If the existing work has been modified sufficiently, it is said to be transformed into a new work and is not a violation of the existing copyrighted work. An instructive transformation case *Author's Guild v. Google*,[4] in which the U.S. Court of Appeals for the Second Circuit determined that Google's scanning and cataloging of numerous copyrighted books as digital content was "highly transformative" and constituted fair use. Again, the issues were not related specifically to an AI system, but did deal more generally with digital copies of input data.

Despite this guidance, AI systems present unique challenges that cannot be fully resolved using existing fair use and transformation principles, even those set forth by the *Andy Warhol Foundation* and *Author's Guild* cases.

Training Data

Large data sets of training data are needed to train an AI algorithm. This allows the system to set various internal parameters. Sometimes the training data is merely evaluated and then discarded. In other cases, the training data may be incorporated into the AI model and used during subsequent content generation.

Developers of AI systems usually attempt to only use authorized training data in order to remove the use of unauthorized copyrighted works as an issue. However, the amounts of training data used by modern AI systems is staggering and will only increase in time. Except in extremely limited circumstances, it is inevitable that AI systems will be trained using data for which the developers did not obtain an explicit license. Indeed, it can be difficult or nearly impossible to identify what training data were used at all, unless the developers reveal this information (or even know).

There are cases currently pending in various courts where AI system developers have been sued for the unauthorized use of copyrighted material as part of the training package. One such case is *Getty Images v. Stability AI*,[5] filed in the U.S. District Court for the District of Delaware. It is not yet clear how these and other cases will be handled, and a legislative solution may be required.

An analogy may be helpful in evaluating the specific issues involved in training an AI system. Suppose that a novice artist visits a museum gallery and observes numerous paintings made by a famous painter. Care is taken by the novice artist to consider the subjects of the paintings (e.g., lily pads and bridges) as well as the style used (the brush techniques and level of detail). The novice artist then goes home and creates a painting based on this trip to the museum. The famous painter subsequently alleges that the painting by the novice artist violates one of the copyrights held by the famous painter to one of the displayed works at the museum. This is illustrative of a traditional copyright conflict.

Under existing copyright law, the famous painter would need to show that the novice artist had access to the copyrighted

painting (which the novice artist freely admits), and that there is a substantial similarity between the novice artist's painting to the copyrighted work. Such analysis will be based on the fair use and transformation factors discussed above.

It could be argued that, with respect to training data, an AI system that accesses (such as over the internet) publicly available images as training data is analogous to the novice painter visiting the museum. Both the novice artist who observes the paintings by the master painter and an AI system that uses training data to create original images are engaged in a process of learning and creativity that is inspired by existing works. In both cases, the result is a new work that draws on the influence of the existing works, but is also distinct and original. Whether the work (either the painting or the AI-generated content) is transformative will depend on the work itself, and not the process used in generating the work.

However, there are also some differences between the two processes. The novice artist who observes paintings by the master painter is engaging in a process of study and observation, which involves a degree of interpretation and personal expression. The novice artist must use their own skills and judgment to create a new work that is inspired by the master's work but also reflects their own artistic vision.

The AI model, on the other hand, is using statistical patterns and algorithms to generate new images. While the model may learn from the images in the training data, it does not have the same degree of personal interpretation and expression as a human painter. The output of the model is also more constrained by the data it was trained on, whereas a human painter has more freedom to depart from their sources of inspiration and create something truly original. Hence, the factors may be different when evaluating training data by an AI system as compared to human observation.

A final complicating factor is that copyright infringement focuses on the substantial similarity of the resulting work, not the internal process that was used to make the work. In the case of the novice artist, it is immaterial what mental processes took place; what matters is what does the painting look like. The question is whether this same analysis should be applied to AI system–generated content.

User Input Data

One powerful feature of AI systems is the ability to generate complex content based on relatively simple input. Because of this, the U.S. Copyright Office has recently published updated guidelines addressing copyright registrations to AI-generated works.[6] In these guidelines, the Copyright Office states "the Office will consider whether the AI contributions are the result of 'mechanical reproduction' or instead of an author's 'own original mental conception, to which [the author] gave visible form.' The answer will depend on the circumstances, particularly how the AI tool operates and how it was used to create the final work. This is necessarily a case-by-case inquiry."

These guidelines provide direction that a copyright application for a work that was generated using an AI system should disclose the same in the comments field using a statement that identifies which elements were human-authored content and which elements were AI-generated content. While helpful in addressing these issues, further clarification is needed for a longer-term solution.

Consider a user that prompts an AI system to create a children's story about a frog that meets a rabbit and they have an adventure. In one case, this is the extent of the input. In another case, the user proceeds to provide explicit details and aspects that would normally be viewed as copyrightable material. In both cases, the AI system generates an output with images, text, and dialog. Notice that in both cases, it is not apparent that an AI system generated the content at all, and the amount of contribution by the user cannot be determined by a direct examination of the AI-generated content itself.

A legislative solution may be needed where a copyright registration can be granted, provided that disclosure is made that the work was generated with the assistance of an AI system, and the input supplied by the user is also disclosed. The Copyright Office already accepts sections of source code used as part of a copyright to a software system, and an analogous system should be considered where similar program listings are made as part of the input.

Diffusion Models and Other AI Algorithms

A variety of AI algorithms have been developed. One particular type of model uses stable diffusion, which is a technique that

involves the gradual diffusion of information across nodes in a network. It is a way of allowing information to spread across the network, while keeping it stable and avoiding excessive noise.

In AI systems that use stable diffusion techniques, portions of the training data are incorporated into the AI algorithm by combining the copyrighted works (such as images) with noise that is then used during the generation of AI content. The *Getty Images v. Stability AI* case is an example where these issues are being addressed.

In some ways, a traditional fair use and transformation analysis would appear to be an appropriate framework for resolving whether the internal use of a copyrighted work in a stable diffusion model is transformative or derivative. However, as mentioned previously, the existing copyright laws are set up to examine the resulting output, not the process used to make the resulting output.

Post-Curation Processing

A final area relates to post-curation processing where a user takes the output content from an AI system and modifies this output to generate a final work. There are a number of ways in which such post-curation processing can be carried out. A text-based output in the form of a story can be generated by an AI system, and the story is edited or rewritten by a human editor. In a graphical arts context, a human can physically or digitally modify an original image created by the AI system. In another case, an AI system generates background or character movements that are combined with human-generated elements into a full-length motion picture.

In terms of copyright, the work would be protectable to the extent of the originality and creativity supplied by a human author. Copyrightable elements are those that are distinctive and are directed to particular expressions, such as discussed in the famous Sherlock Holmes case *Klinger v. Conan Doyle Estate*[7] where specific details in the story, such as the demeanors, habits, backgrounds, and proclivities of Holmes and Watson, were protectable.

Once again, a legislative solution may be required. It will not be apparent from an examination of the final work what elements were generated using an AI system. Disclosure of the use of AI-assisted processing, as well as disclosure of the elements provided by the human author (including possibly a copy of the original AI system image), may be a workable solution that protects both the public and the copyright owner.

Conclusion

Copyright law is central to society and is often changed in significant ways due to advances in technology. The Supreme Court resolved the issues relating to copyright ownership and home video recordings in the *Sony Corp. of America v. Universal City Studios, Inc.*, case in 1984,[8] and Congress resolved the issues relating to copyright ownership and the internet via the Digital Millennium Copyright Act in 1998. The unique challenges and opportunities provided by the advent of AI systems requires a similar undertaking.

Such legislative changes should provide clear lines on what can constitute training data, how copyrighted works can be used by an internal AI system, and to what extent AI-generated output content can be protected. Attribution notices that a particular work was generated using an AI system should be an important part of the final solution. In this way, everyone involved—including the copyright owners, other creators of new works, and the general public—will be able to gain the greatest benefits from this exciting technology.

Notes

* Randall "Randy" McCarthy, a shareholder at Hall Estill, is a registered U.S. patent attorney who practices in all areas of intellectual property law prosecution, litigation, and counseling. His experience, which has a focus in the high-tech arena, includes development and management of client intellectual property (IP) assets, licensing, IP strategy formulation, and portfolio valuation.

1. Title 17 U.S.C.
2. 35 U.S.C. 106.
3. Andy Warhol Foundation for the Visual Arts v. Goldsmith, No. 19-2420 (U.S.).
4. Author's Guild v. Google, 721 F.3d 132 (2d Cir. 2015).
5. https://www.reuters.com/legal/getty-images-lawsuit-says-stability-ai-misused-photos-train-ai-2023-02-06/.
6. Copyright Registration Guidance: Works Containing Material Generated by Artificial Intelligence, Federal Register 2023-05321, March 16, 2023.
7. Klinger v. Conan Doyle Estate, 755 F.3d 496 (7th Cir. 2014).
8. Sony Corp. of America v. Universal City Studios, Inc., 464 U.S. 417 (1984).