Mark S. Hewitt, Ph.D.
14599 Alabaster Ave Delray
Beach, FL 33446
mark@telephony.net
(941) 320-1676

Friday, September 1, 2023

U.S. Copyright Office
Attn: Fair Use and Artificial Intelligence Section
101 Independence Avenue SE
Washington, DC 20559-6000

Subject: Public Comment on Fair Use and Artificial Intelligence Training

Dear Sir/Madam,

I am writing to provide public comment on the intersection of fair use and artificial intelligence (AI) training as part of the ongoing inquiry conducted by the U.S. Copyright Office. I commend the Copyright Office's efforts to address the complex issues arising from using copyrighted materials in AI training and its impact on copyright liability and ownership.

As outlined in the Fair Use Research document, it is crucial to understand the difference between plagiarism and the learning process that makes AI a tool for human innovation. AI models, such as large language models like ChatGPT, operate by analyzing and processing vast amounts of data, including open and copyrighted materials, to learn and generate human-like responses.

Plagiarism, which involves the unauthorized and deliberate use of copyrighted materials for personal gain, is a clear violation of copyright law. However, the use of copyrighted materials by innovators in transformative ways, such as AI training, falls within the realm of fair use. Fair use enables commentary, criticism, and transformative benefits that contribute to the progress of science and the useful arts.

AI models, like humans, learn by observing and processing information from various sources, including copyrighted materials. Just as humans read books, articles, and other copyrighted works to expand their knowledge and understanding, AI models analyze and derive insights from these materials to improve their language generation capabilities.

It is important to note that AI models like ChatGPT are only at the early stages of development; they do not seek to profit directly from distributing copyrighted materials. They aim to learn the underlying rules of human language and assist users in various tasks, such as saving time at work, simplifying daily life, or providing entertainment. The goal is to enhance human productivity and facilitate creative expression rather than to exploit copyrighted works for financial gain. Any limited access to the same materials we use to learn would only hinder the tools we build to improve ourselves and our world.

Furthermore, copyright law protects the expression of ideas and creative works rather than the underlying opinions or factual information. As demonstrated in the precedent set by notable copyright cases, such as the Google Books case, statistical data, word frequencies, syntactic patterns, and thematic markers derived from copyrighted works are beyond the scope of copyright protection.

AI-generated outputs, produced through the learning process from copyrighted materials, should not be considered derivative works in every instance. Simple responses, factual information, or summaries of copyrighted works, as well as outputs that do not exhibit substantial similarities to the training materials, are akin to book reports or reviews and fall within the realm of fair use.

In conclusion, fair use principles should apply to AI training, allowing for the transformative use of copyrighted materials that facilitate innovation and progress in artificial intelligence. By distinguishing between plagiarism and the learning process inherent in AI models, we can ensure that AI remains a valuable tool for human creativity and advancement while respecting the rights of copyright holders.

Thank you for considering this public comment. A balanced approach to the intersection of fair use and AI training will foster continued innovation while upholding the principles of copyright law. If there are any further opportunities for discussion or clarification, I would appreciate the opportunity to contribute to this critical dialogue.

Sincerely,

Mark S. Hewitt, Ph.D.