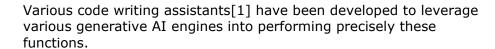# Beware The Legal Risks Of Using AI In Software Development

By **Andrew Freyer and Palash Basu** (September 20, 2023)

Generative artificial Intelligence has garnered tremendous public attention with its potential to transform many industries, potentially chief among which is the software development industry.

Andrew Freyer

Specifically, modern expectations for the pace of software development may signal that software companies — especially those embracing continuous integration and continuous delivery practices — are among the most motivated and excited by the promise that trained generative AI engines can be prompted to provide, as output, purpose-configured code that is syntactically correct, compilable and executable.

Various code writing assistants[1] have been developed to leverage various generative AI engines into performing precisely these functions.

Some assistants enable a developer to describe software functionality in prose and generate purpose-configured code as output to be inspected by the developer and inserted into an existing codebase.

Palash Basu

In other cases, an assistant may provide real-time, in-line, code suggestions to a developer as the developer works.

According to a recent survey by GitHub Inc., "92% of U.S.-based developers working in large companies report using an AI coding tool either at work or in their personal time,"[2] while "70% [predict] AI coding tools will offer them an advantage at work and cite better code quality, completion time, and resolving incidents as some of the top anticipated benefits."[3]

However, substantially all generative AI platforms including code writing assistants are trained against massive data lakes of billions of lines of publicly available text including open-source code that — despite being publicly available — may nevertheless be subject to copyright protections and/or license-specific use restrictions.

As a result, although generative AI output often appears to be nondeterministic, original and unique, such output often can include code, comments or other content repeated verbatim from an original source location in a manner that infringes copyright or runs afoul of a use restriction associated with that source.

More simply, incorporating generative AI output into a software product is not without legal risk.

Certain Risks of Incorporating Output of Generative AI Into Software Products

As an initial matter, developers should review and understand terms of use of any code writing assistant used in the course of developing a software product.

Some terms disclaim any ownership of output,[4] whereas others expressly assign any and

all right, title and interest of the provider of the code writing assistant to the user of the system.[5]

Still, other terms of use may retain for the provider of a code writing assistant partial ownership or royalty-free nonexclusive licensing rights to output of that assistant. Simply put, software development companies are well advised to be aware of any retained rights or ownership claims resulting from use of code writing assistants in development workflows.

A more prominent concern results from use of generative AI output that duplicates, or substantially duplicates, code or other content appearing in training data.

Often, users of code writing assistants do not have any visibility into whether a particular output, or portion thereof, is original to a code writing assistant, or whether that output is repeated from an identifiable source.

Further, in many cases, output of a code writing assistant may include sections copied from multiple sources, each associated to different copyright holders. This opacity presents a significant infringement liability issue for users of code writing assistants as incorporating any suggested code may expose direct or indirect copyright infringement liability in respect of an unknowable number of copyright holders.

Further still, even if a particular output is not identical to copyright-protected content, it nevertheless may constitute a derivative work.[6]

Courts have noted that not all copying is copyright infringement, but it has nevertheless long been established that intention to infringe is not essential under the Copyright Act.[7]

More simply, a developer is unlikely to successfully argue that use of copyright-infringing generative AI output is an excusable innocent infringement.

To mitigate risks of unintended copyright infringement, some code writing assistants include filtering features or suppress generative outputs that include potentially copyrighted works.[8]

However, such features are unlikely to be accompanied by indemnity or accuracy guarantees and thus users of code writing assistants should be aware of the potential that any use of generative output carries copyright infringement liability risk.

Further, many open-source software products are released with copyright management information identifying creators, authors, owners, and/or providing use restriction information — e.g., identifying an open-source license that applies.

The Digital Millennium Copyright Act specifically protects against the removal or alteration of copyright management information.[9] Problematically, code writing assistant output is typically not accompanied by any copyright management information, introducing another potential liability under the DMCA for developers using such tools.

However, notwithstanding the foregoing academic arguments, copyright holders face a significant challenge discovering and asserting infringement whether against users or providers of code writing assistants.

As against users of code writing assistants, discovery of infringement may be effectively impossible absent publication of infringing code by those developers. As against providers of

code writing assistants, courts are signaling that pleading standards likely require a showing by copyright holders that holder-owned code has actually been provided as output.[10]

Certain Risks of Prompting Generative AI Output

The content of prompts provided to code writing assistants — whether provided manually in natural language or automatically by operation of an assistant within a software development environment — requires consideration of confidentiality agreement obligations, trade secret and confidential business information protection policies.

Many code writing assistants specifically note in terms of use that prompt content may be stored and used as training data to further refine underlying generative AI engines.

This introduces a possibility that information provided within a prompt may be included in an output to another user, potentially exposing confidential information.

Simply, absent confidentiality agreements with providers of code writing assistants, developers should presume that prompts, product specifications, business information, personal health information, personally identifiable information and any other information provided intentionally or inadvertently to a code writing assistant is neither private nor confidential.

Developers are also well advised to ensure that prompts are either not retained or are siloed for use only within the developer's own environment.

Andrew J. Freyer is a shareholder and Palash Basu is a senior attorney at Brownstein Hyatt Farber Schreck LLP.

The opinions expressed are those of the author(s) and do not necessarily reflect the views of their employer, its clients, or Portfolio Media Inc., or any of its or their respective affiliates. This article is for general information purposes and is not intended to be and should not be taken as legal advice.

[1] Examples include GitHub Copilot, Amazon CodeWhisperer, Tabnine, and the like.

[2] https://github.blog/2023-06-13-survey-reveals-ais-impact-on-the-developer-experience/. (Accessed Sep. 1, 2023).

[3] Id.

[4] Product Specific Terms for GitHub Copilot expressly states that GitHub does not "claim any ownership" in suggestions generated by the system. GitHub Copilot Product Specific Terms Section 2 (Ver. Jun 2023); Available at https://github.com/customer-terms/github-copilot-product-specific-terms (Accessed Sep. 1, 2023).

[5] Section 3(a), Open AI Terms of Use (Ver. Mar. 14, 2023); Available at https://openai.com/policies/terms-of-use,. (Accessed Sep. 1, 2023).

[6] 17 U.S.C. §101.

[7] Feist Publ'ns, Inc. v. Rural Tel. Serv. Co., Inc., 499 U.S. 340, 361 (1991) and Buck v.

Jewell-Lasalle Realty Co., 283 U.S. 191, 198 (1931).

[8] e.g., GitHub Duplicate Detection

[9] 17 U.S. Code § 1202.

[10] J. DOE 1 et al. v. GitHub, Inc. et al., Case No. 4:22-cv-06823-JST (N.D. Cal.).