Published on *The National Law Review* https://natlawreview.com

Potential Pitfalls and Best Practices In Using AI Tools to Generate Code

Article By:

Dr. Christian E. Mammen

Fabio E. Marino

Daniel M. Grigore

Artificial Intelligence (AI) has become an even hotter topic since the introduction of generative AI tools like ChatGPT, a chatbot developed by OpenAI, as well as tools like Copilot and OpenAI Codex, which use generative AI to write computer code. The possibilities that these tools present appear endless, and users have begun to test the limits of AI in the workplace to see where the benefits stop and the consequences begin.

Reportedly, Samsung recently experienced three instances of corporate-secret leaks involving ChatGPT. Twice, semiconductor engineers fed confidential source code information into ChatGPT to fix and optimize the code. The third occurrence involved an efficiency-minded employee who asked ChatGPT to summarize meeting notes. Since ChatGPT retains user input data for machine learning to train itself, the workers effectively, and inadvertently, disclosed Samsung confidential information to OpenAI.

But inadvertent disclosure of confidential information is not the only potential pitfall of using ChatGPT, as demonstrated by the recently filed class action suit against GitHub, its parent company Microsoft, and its business partner OpenAI. The complaint alleges that Copilot, GitHub’s AI-powered coding assistant, pirates software that can be traced back to open-source repositories or open-source licensees. The complaint reports that GitHub has conceded that it trained Copilot with data from vast numbers of publicly accessible repositories of code stored on GitHub, much of which, allegedly, is published under licensing terms that require crediting the original authors. Yet, according to the complaint, Copilot regurgitates long sections of licensed code without providing credit. The complaint falls short of accusing GitHub and its compatriots of outright copyright infringement, but it does seek a permanent injunction to ensure that GitHub modifies Copilot to avoid producing uncredited work in the future, plus an array of damages to compensate the class for the alleged use of licensed code without crediting its source.

Turning to the realm of patents, a leading patent blog asked a hypothetical that mingles inventions with AI. In the hypothetical, an inventor, having just developed a core idea for a new product, hops onto ChatGPT and asks it to build on the product idea. ChatGPT expands on the product to the inventor's satisfaction and even provides detailed designs that the inventor had not thought of. The inventor includes the ChatGPT transcript in the disclosure documents. A patent search reveals that the original product idea, alone, would not be patentable, but is likely patentable combined with the ChatGPT input. Standing in the shoes of a patent attorney, you see value in claims directed solely to the features ChatGPT provided. The question: how would you advise your client?

These cautionary tales and thoughtful hypotheticals provoke interesting legal and pragmatic questions:

- Does running proprietary code or other confidential information through a generative AI tool count as disclosure?
    - Is generative AI different in this respect from either traditional search engines or other specialized tools like online legal research?
- Furthermore, in an increasingly first-to-file world, could using an AI tool potentially create invalidating prior art that disrupts the patentability of otherwise patent-worthy ideas?
- How can AI users with valuable intellectual property take steps to prevent revealing what they wish to keep hidden?
- How can AI users avoid infringing the IP rights of others?

For AI users specifically interested in generating code, double check the source. If a company utilizes AI to generate code, it should run the resulting code through open-source licensing tools to determine the code's origins and to check whether the code is licensed. While code written solely by AI *may not* be copyrightable (although source code is copyrightable, the Copyright Office has recently reiterated that to be copyrightable any work must have sufficient human creative involvement), code that has been written, conceived, or outlined by humans with the assistance of AI *may* be copyrightable, and if a third party's copyrighted code is used as training data for the AI, there are unanswered questions about whether that is infringement or fair use of the third party's copyrighted code. Companies that hire vendors or subcontractors to write code for them should include specific language allocating the risks of AI-based infringement, or alternatively, covenants, representations and warranties concerning the use of AI to draft the code. Additionally, when using AI tools, a company should consider whether (a) any *inputs* to the AI are considered trade secrets, and whether sufficient steps are taken to confirm and preserve confidentiality of those inputs, and whether (b) any *outputs* from the AI are, likewise, trade secrets.

While AI tools that generate code present exciting potential, users looking to take advantage of these technologies should pause and take precautionary measures to avoid hitting the rocks hidden beneath the surface of machine learning algorithms.

National Law Review, Volume XIII, Number 108

Source URL:https://natlawreview.com/article/potential-pitfalls-and-best-practices-using-ai-tools-to-

generate-code