Benjamin Bremer

CSCI 472 - Historical and Current Computer Science

December 3, 2025

The Legal and Ethical Implications of GitHub Copilot

I. Introduction

GitHub Copilot (Copilot) is a generative AI tool developed by GitHub, a subsidiary of Microsoft. Copilot is specifically developed to help in the development of code, and this goal is aided by its integration into Microsoft Visual Studio Code, one of the most popular personal and professional integrated development environments (IDEs). Ethical and legal concerns are prevalent with all AI tools, but because Copilot is used to develop so many other applications, the ethical and legal implications also extend to all code generated by Copilot. This paper aims to shed light on a few such ethical and legal concerns. In section II, three points of the Association for Computing Machinery (ACM) Code of Ethics and Professional Conduct will be discussed. In section III, two legal aspects of Copilot will be analyzed.

II. Ethical Issues

The ethical use of Copilot and similar AI tools is a hot topic. As with any other computing innovation, the ethics surrounding Copilot for computing professionals should be governed by the ACM Code of Ethics and Professional Conduct. While Copilot touches on nearly every aspect of the Association for Computing Machinery's (2018) Code of Ethics, this section focuses on 3: ACM 1.6, which states "Respect privacy" (Section 1.6), ACM 2.6, which states "Perform work only in areas of competence" (Section 2.6), and ACM 3.7, which states "Recognize and take special care of systems that become integrated into the infrastructure of society" (Section 3.7).

ACM expands on ACM 1.6 in their Code of Ethics and Professional Conduct, stating that computing professionals are especially burdened with the protection of privacy because computing systems "enable the collection, monitoring, and exchange of personal information

quickly, inexpensively, and often without the knowledge of the people affected" (Association for Computing Machinery, 2018, Section 1.6). Thus, ACM 1.6 claims that it is the responsibility of the computing professional to know the responsibilities they have in regards to the collection and distribution of private information (Association for Computing Machinery, 2018). Copilot makes that task harder than ever. According to GitHub, Copilot is trained on a vast database of open-source code; however, by their own admission, Copilot did occasionally produce personally identifiable information, or PII, in testing (GitHub, n.d.). While GitHub claims that those instances were hallucinations, Niu et al. (2023) claim that GitHub's lack of systematic assessment and exclusion of PII training data is a significant privacy concern. In fact, Niu et al. (2023) were able to generate 43 prompts that generated privacy leaks from the Codex model, the model behind Copilot. In many cases, these leaks were considered indirect, meaning that GitHub's PII blocker is working, but not well enough (Niu et al., 2023). Szolderits (2025) claims that training code should be anonymized to prevent PII leaks, but that method could have legal implications, as discussed further on in the paper.

ACM 1.6 also states that only a minimal required amount of PII should ever be collected, and computing professionals should only use it for legitimate purposes (Association for Computing Machinery, 2018). Additionally, as stated above, ACM 1.6 is concerned with the distribution of PII without the corresponding individual's knowledge. Copilot is teetering on the edge of this ethical standard, as repository owners may be donating their code or personal information without their knowledge or consent. Then, this collected information is built into the model, which is released into various end-user products, and may regurgitate the individual's information, again without their knowledge or consent. Overall, privacy is a significant concern, not only for AI tools as a whole, but for Copilot specifically. By training on code contributed to

repositories by millions of users, the model is bound to, maybe only inadvertently, learn and provide private information without the owner's knowledge or consent.

Section 2.6 of the Association for Computing Machinery's (2018) Code of Ethics rightfully constrains computing professionals to working within areas of competence, and disclosing to relevant parties when the professional is beyond the scope of their knowledge and abilities. With the seamless integration of Copilot into an industry-leading IDE, Visual Studio Code, computing professionals may be more prone to overconfidence, failing to properly identify and disclose their shortcomings. According to a 2025 Forbes article, young professionals grossly overestimate their skills with AI while underperforming on tasks like prompting or evaluating results for accuracy (Galagali, 2025). According to the same article, almost 80% of Gen Z professionals use AI, such as Copilot, daily, and 38% of Gen Z professionals are fully reliant upon AI tools for their work (Galagali, 2025). The staggering number of professionals that are fully reliant upon tools like Copilot to complete their daily tasks are the same professionals that are drastically undereducated in its use. On a higher level, an article by The HR Digest claims that business leaders are overestimating their employees' skills in AI use, leading to a lack of support and training for the proper use of AI tools, and a gap in governance of their proper use (Coker, 2025). The improper use and total reliance on Copilot could lead to low-quality or even erroneous work, causing users to violate ACM 2.6.

According to Section 3.7 of the Association for Computing Machinery's (2018) Code of Ethics, computing professionals have a special responsibility to be good stewards of systems that are tightly integrated into society. Today, code is ubiquitous, impacting nearly every aspect of our daily lives. Copilot is becoming more ubiquitous in the development of such code. Its integration into Visual Studio Code, an industry-leading code development IDE, makes it more accessible

than ever. For example, a CNBC article cites Microsoft CEO Satya Nadella saying that 20-30% of Microsoft's code base has been AI generated, to which Meta CEO Mark Zuckerberg responded that Meta is hoping that half of development is done by AI tools by as soon as 2026 (Novet and Vanian, 2025). Despite the previously stated privacy concerns relating to Copilot, and the overreliance and overestimation of proficiency with tools like Copilot, it is becoming increasingly more ingrained into society. Computing professionals should be aware not only of how Copilot affects society, but of how the code developed by Copilot affects society. As more code is developed by Copilot, its reach into our lives grows increasingly strong, further highlighting the importance of upholding every aspect of the ACM Code of Ethics.

III. Legal Issues

When considering the legal aspects of Copilot, two issues are immediately prevalent: copyright, licensing, and intellectual property rights, and liability. These are common concerns with any created work or product. The use of Copilot in the development of software exacerbates these issues, as the lines of copyright infringement and liability assessment are blurred by Copilot's AI-generated code integrated into applications.

Copilot is trained on a dataset of publicly available code, according to claims made by GitHub (GitHub, n.d.). Responses to user queries can generate code that resembles code from the training dataset, which may be under copyright or other intellectual property (IP) protections (Szolderits, 2025). Because Copilot-generated code might infringe on existing copyright, there are legal questions about the ownership of the generated code (Szolderits, 2025). Additionally, the source code used to train Copilot is protected under various licenses. In order to be in compliance with such licenses, users need to check the code generated for similarity to source

code, and comply with the source code license (Szolderits, 2025). This can raise issues in corporate realms, as some open-source licenses restrict the use of source code in closed-source software products.

Additional questions may be raised about copyright, licensing, and IP infringement, when considering how Copilot disseminates information. Similar to the privacy discussion, Copilot may not directly generate copyrighted source code, but it may indirectly infringe on open-source licenses or copyright through generating similar results to the source training code. Additionally, Basanagoudar and Srekanth (2023) claim that failing to provide proper attribution amounts to copyright infringement and a violation of open-source licenses. Copilot doesn't, and often can't, cite which sources its generated code is pulled from, since the generated code is likely some combination of several sources. So, the question arises: How similar to source code does generated code need to be in order to require a citation to avoid copyright infringement? Basanagoudar and Srekanth (2023) point to the EU AI Act as a model for regulating attribution in AI model training. Overall, copyright, licensing, and IP rights are a significant legal concern for Copilot.

A significant legal concern with all AI tools is how to assess liability for damages made as a result of generated output. Because Copilot is being used to generate code in so many applications, there are many areas where liability could be an issue. According to Szolderits (2025), the proper assignment of liability for damages is a critical step that needs to be taken to properly protect users and companies. Current liability assertions can raise questions about whether the developer, GitHub, or even OpenAI should be liable for damages incurred by code developed with or generated by Copilot (Szolderits, 2025). In order to clarify legal concerns

moving forward, effort should be put into regulation and legal precedent to determine who is liable for damages incurred as a result of code generated by Copilot.

IV. Conclusion

Overall, there are many ethical and legal concerns surrounding Copilot, its use, and the code it generates. In fact, Copilot touches on even more legal and ethical issues than the ones covered in this short paper. This work serves to shed light on three points from the ACM Code of Ethics and Professional Conduct: ACM 1.6, ACM 2.6, and ACM 3.7. Additionally, this paper discusses two legal issues involving Copilot: copyright and intellectual property, and liability. In the future, more effort should be aimed toward how Copilot affects every aspect of the ACM Code of Ethics and Professional Conduct. Furthermore, an in-depth legal analysis of both points discussed in this paper could yield solutions to the issues discussed. Copilot is a powerful new tool, enabling software engineers to develop applications faster than ever before; however, as with the advent of any new tool, much ethical and legal deliberation should take place to ensure Copilot is properly governed to enable safe and effective use for all users, as well as any users affected by Copilot-generated code.

References

Association for Computing Machinery. (2018). *ACM code of ethics and professional conduct*. https://www.acm.org/code-of-ethics

Basanagoudar, V, & Srekanth, A. (2023). Copyright conundrums in generative ai: github copilot's not-so-fair use of open-source licensed code. Journal of Intellectual Property Studies, 7(2), 58-68.

Coker, D. (2025, October 28). Why AI overconfidence could be corporate America's biggest risk. *The HR Digest*. https://www.thehrdigest.com/why-ai-overconfidence-could-be-corporate-americas-biggest-risk/

Galagali, T. (2025, August 25). Gen Z's relationship with AI. *Forbes*. https://www.forbes.com/sites/tarungalagali/2025/08/25/ais-impact-on-generation-z/

GitHub. (n.d.). *GitHub Copilot*. Retrieved December 3, 2025, from https://github.com/features/copilot

Niu, L., Mirza, S., Maradni, Z., & Pöpper, C. (2023). CodexLeaks: Privacy leaks from code generation language models in GitHub Copilot. In *Proceedings of the 32nd USENIX Security Symposium* (pp. 2133–2150). USENIX Association. https://www.usenix.org/conference/usenixsecurity23/presentation/niu

Novet, J., & Vanian, J. (2025, April 29). *Satya Nadella says as much as 30% of Microsoft code is written by AI*. CNBC. https://www.cnbc.com/2025/04/29/satya-nadella-says-as-much-as-30percent-of-microsoft-code-is-written-by-ai.html

Szolderits, C. M. (2025). *A new era of coding: Investigating GitHub Copilot's influence on productivity, code quality and ethics in software development* [Master's thesis, Alpen-Adria-Universität Klagenfurt]. NetLibrary. https://netlibrary.aau.at/obvuklhs/content/titleinfo/11700723/full.pdf