Ethical Considerations of Using Generative AI in Software Development

Tiina Tuomisto
Lasse Harjumaa
Kokkola University Consortium Chydenius
University of Jyväskylä
Kokkola, Finland
e-mail: tii2tuo@gmail.com | lasse.m.harjumaa@jyu.fi

***Abstract*—Generative Artificial Intelligence (GenAI) is gaining more and more popularity among different professions, and it is also transforming software development by enabling automated text, code, and image generation, enhancing productivity across various tasks. This study examines the most common use cases of GenAI tools in software development, alongside software professionals' experiences, ethical concerns, and awareness of ethical guidelines related to GenAI. Through qualitative semi-structured interviews with IT professionals in Finland, the research identifies that GenAI tools are primarily used for programming, software testing, and general problem-solving. While these tools accelerate workflows, challenges such as unreliable outputs, outdated training data, and ethical concerns - particularly regarding transparency, bias, and data security - remain significant. The findings align with previous research, highlighting both the benefits and risks of GenAI adoption. Ethical considerations, though acknowledged, are rarely integrated into daily practices, emphasizing the need for clearer guidelines and education. This study contributes to the discourse on the evolving role of AI in software engineering, underscoring the importance of ethical awareness and responsible AI utilization.**

Keywords-Generative artificial intelligence; ethics; software development; GenAI; GenAI tools.

I. Introduction

GenAI is one of the most influential technological innovations of recent years. While traditional systems utilizing artificial intelligence follow predetermined patterns and rules, GenAI systems can create more imaginative items, such as text, images, sounds and videos. This ability makes it an attractive tool for various purposes, including tasks related to software development. Instead of having to search through repositories of previously written questions and answers by other developers, such as searching through StackOverflow, GenAI tools will create an answer based on huge amounts of data, which has been used to train them. To improve the answers over time, they are further trained based on the feedback given on the generated responses.

Generative AI can improve productivity in various software development tasks, including coding, testing, debugging, documentation, reviewing and brainstorming. At best, the quality of the product and productivity of the developers can improve significantly [1].

It has been predicted that generative AI will have a significant impact on software practitioners' work, permanently changing the roles of software engineers. Developers will need to have new kinds of competencies to master and fully utilize the potential of GenAI systems [2]. Since generative AI does not have a comprehensive knowledge of the context in which it generates a solution, or human language, the engineers' who is authoring the prompt has the responsibility to verify the conformity of the generated outcome.

GenAI tools also pose numerous issues, many of which are related to ethics. For example, Weidinger et al. [3] identify six ethical and social risk areas related to language models: 1. Discrimination, Exclusion and Toxicity, 2. Information Hazards, 3. Misinformation Harms, 4. Malicious Uses, 5. Human-Computer Interaction Harms, 6. Automation, Access and Environmental Harms. Tanaka et al. [4] further decompose these risk classes into more detailed subcategories to help mitigate risks by either removing the risk sources, avoiding the hazard, or managing the impact.

Due to the potential ethical problems associated with GenAI tools, several entities have developed ethical guidelines for the developers and users of AI systems. An example of this is the High-Level Expert Group on Artificial Intelligence (AI HLEG) established by the European Commission [5], which has created ethical guidelines for developing trustworthy AI systems. However, adhering to various ethical guidelines and principles is not easy. As Ryan and Stahl [6] point out, they do not provide sufficient advice for implementing the principles in practice. Consequently, software practitioners encounter various ethical issues when using GenAI tools.

This study brings forward the experiences and opinions of software practitioners regarding the benefits and drawbacks of utilizing GenAI tools, as well as the awareness of AI ethical guidelines and the prevalence of ethical issues. Our study extends prior research by offering insights into how practitioners engage with ethical issues specifically in the context of generating software engineering artifacts and by examining the extent to which they follow relevant guidelines, if any. We also integrate perspectives on software quality, the ethics of software development more broadly, and existing ethical guidelines for AI to outline the foundations of guidelines specifically intended for the use of GenAI in software development tasks.

The rest of the paper is arranged as follows. Section 2 describes the basic concepts of generative artificial intelligence and its utilization in the various phases of software development, as well as the ethical issues related to AI within the domain. Section 3 describes the research approach and the main findings of the study, and the contributions and limitations of the study are discussed in Section 4. Finally,

Section 5 concludes the paper.

II. Generative AI in Software Development

This section provides a brief overview of the tasks that typically comprise the software development process and examines how GenAI has been applied to these tasks, as reported in academic research. It then discusses how ethical issues related to GenAI have been addressed in guidelines issued by major organizations.

A. Software Development Tasks

Software development process models - such as the waterfall model, agile, spiral, and DevOps - provide structured guidance to build high-quality software efficiently. These models help developers, project managers, and stakeholders manage resources, improve predictability, and standardize outcomes. Regardless of the model, key development tasks remain consistent.

Development usually begins with planning and a feasibility study, where project goals, risks, and viability are assessed. Requirements engineering follows, focusing on gathering and documenting user needs, either through formal specifications in traditional models or iterative collaboration in agile methods. System design defines architecture, components, and interfaces. In the implementation phase, developers choose suitable tools and follow best practices to ensure maintainable, high-quality code. Testing and reviews are critical for quality assurance: testing includes unit, integration, and acceptance levels, often automated to enhance efficiency. Reviews help detect defects early. In agile development, testing is continuous and central to rapid feedback. Deployment moves the finished product to production, ensuring it is accessible and operational. Finally, maintenance and support address issues post-launch, keep the software secure, and adapt it to evolving needs. Together with project management tasks, these stages form a cycle that ensures user-focused, robust, and maintainable software delivery.

B. Applications of GenAI

GenAI tools have been widely studied for their ability to support software development tasks. These tools can generate code [7], fix coding errors, translate between programming languages [8], and assist with writing comments [9] or answering programming queries [10]. The effectiveness of GenAI varies by user experience. While Copilot-generated code provides a good starting point [11], it does not always speed up development. Moradi Dakhel et al. [12] noted that experienced developers gain the most, as they can spot and correct errors. Inexperienced users benefit from guidance and learning opportunities, provided they evaluate the output critically. Across skill levels, GenAI reduces time spent consulting documentation and offers real-time advice - though its suggestions should be validated, as it lacks full context awareness.

GenAI also supports software testing. Studies have explored its use in generating unit tests [13], for example. Although generated test data may contain errors, it forms a useful foundation when properly reviewed. In maintenance, GenAI enables anomaly detection and automatic code fixes, as shown in a study by Khlaisamniang et al. [14]. In design, GenAI aids in producing diagrams [1], UI development, and prototyping [15]. Project planning is another emerging area, where GenAI helps with scheduling, cost estimation, and risk analysis [16]. GenAI also assists in requirements engineering: language models help gather, summarize, and complete requirements [17], including user story generation. Despite occasional inaccuracies, GenAI tools offer significant value across the software development life cycle by providing inspiration, automating routine tasks, and enhancing productivity.

C. Generative AI Tools

Studies evaluating the code quality of GenAI tools show mixed results. GitHub Copilot and Microsoft Copilot have been rated highest in code generation [18][19]. ChatGPT has also demonstrated strong performance, though it is considered less effective in implementation and testing [1]. Google Gemini and Claude received lower evaluations in multiple studies, while Amazon Q Developer was also criticized [19]. GenAI tool performance varies by programming language, making it essential to align tool selection with project-specific needs.

Error correction capabilities have been studied primarily in ChatGPT and GitHub Copilot, showing moderate success but with limitations in complex problem-solving [8]. Effective prompting is crucial to improving output quality. Other tools, such as Claude and Gemini, lack extensive evaluation in this area. GenAI tools like ChatGPT and GitHub Copilot also assist with information retrieval and development guidance, though their utility beyond programming is limited. ChatGPT remains the most researched tool across software development tasks, including documentation and design, while others have been studied mainly for coding purposes.

The reliability of GenAI tools depends heavily on the comprehensiveness of their training data. When the data is insufficient or misaligned with the problem, tools may produce vague or incorrect responses, requiring users to spend time validating and refining the output [8]. Developers must also provide clear prompts, as vague inputs can degrade response quality. Although tools like Copilot can reduce vulnerabilities [7], other studies show they may introduce more errors than inexperienced developers [12] and obscure error sources [20]. Programming language and problem familiarity affect response accuracy, with GenAI tools performing better on common tasks with ample training data [19]. While GenAI supports learning and boosts productivity for novices, overreliance may hinder skill development. Understanding a tool's inner workings is also vital, as misconceptions about how suggestions are generated can mislead users [20].

D. Ethics in Software Development And Issues Related to Generative AI

The IEEE-CS/ACM Joint Task Force (2021) outlines eight ethical principles for software developers, covering responsibilities to the public, clients, employers, products, professional

judgment, management, the profession, colleagues, and self. These guidelines, which are also adopted nationally (e.g., in Finland by TIVIA), emphasize that developers must ensure their work is safe, lawful, and ethical.

Software must not harm people, nature, or society, and any risks must be reported. Developers must communicate honestly about project feasibility, costs, and risks, and should not misuse confidential information or overstate their skills. They must deliver products that meet quality standards and are properly documented, tested, and maintained. Developers are expected to act impartially, avoid conflicts of interest, and reject unethical practices like bribery. Managers must ensure team members understand ethical practices, treat staff fairly, and handle project resources responsibly.

Professionalism involves educating others, complying with laws, and taking accountability for software quality and errors. Developers should also support and evaluate colleagues fairly, protect confidentiality, and continually improve their skills in areas like design, maintenance, and testing. Ultimately, ethical software development requires responsibility, transparency, fairness, and ongoing self-improvement.

Various organizations, including AI HLEG [5], IEEE [23][24][26][27], Adobe [25], Google [21], and IBM [22], have developed ethical guidelines for AI to address risks throughout its lifecycle. Despite this, no specific, comprehensive ethical standards exist solely for generative AI, highlighting a gap amid its rapid development and increasing use. These ethical guidelines and principles for AI are not legally binding. They therefore act as incentives for ethical behavior. Table 1 represents the summary of the ethical principles of these major organizations.

III. Research Setting and Key Findings

This section describes the research approach adopted in the study and summarizes the main findings that emerged from the interviews.

A. Research Approach

This study examines whether IT companies utilize any ethical guidelines when using GenAI tools and how familiar the guidelines for artificial intelligence are to IT-professionals. The research questions are as follows:

RQ1. What software development tasks are Generative AI tools used for in practice?

RQ2. What are the benefits and drawbacks of Generative AI tools as perceived by software developers?

TABLE I. Summary of the Ethical Principles.

Ethical Principle	Description	Organizations
Privacy	Users must be able to control their own data and be informed about why the data is collected and how it is used. User data must be kept intact and must not be misused.	[5], [21], [22], [23]
Security and resistance to threats	AI must withstand exceptional situations, such as attacks, without causing harm. It must perform as intended in varying conditions and its results must be repeatable. The results of AI must be accurate, and the probability of incorrect predictions must be disclosed openly. AI should also have security measures.	[5], [21], [22], [24]
Explainability	Explainability refers to the comprehensibility of information provided about artificial intelligence. In this case, the explanation provided should be formulated according to the level of expertise of the recipients of the information so that the explanation is understandable. This guideline is closely related to transparency.	[5], [22], [23]
Transparency	The user must gain an understanding of how artificial intelligence works, how data is used and stored. This helps to understand how and why the system reaches a certain result. Traceability must be ensured so that if an error occurs, the cause of the error can be determined and prevented in the future. The user must be informed whether they are working with another person or artificial intelligence. This guideline is closely related to explainability.	[25], [5], [21], [23]
Justice and equity	The aim is to treat different groups and individuals equally. It prevents bias, which refers to bias towards a position, object or person, which can lead to unfair results. Data sets, such as training data, can contain unintentional biases that exacerbate discrimination and prejudice against individuals and groups.	[25], [5], [21], [24]
Sustainability and well-being	Artificial intelligence should operate in an environmentally friendly manner and support the well-being of people and society. The aim is to ensure that the artificial intelligence being developed brings clearly more benefits to society than disadvantages and challenges.	[5], [21], [26]
Accountability	Assessing artificial intelligence and identifying and reporting its negative impacts. Impact assessments can be used to prevent negative impacts. Users have the right to appeal if a decision made by an artificial intelligence negatively affects them. The guideline also includes the principle of fairness.	[25], [5], [21]
Self-determination	This guideline aims to protect human rights. Artificial intelligence must not endanger people's right to self-determination and artificial intelligence can be monitored by a human to overturn decisions made by the artificial intelligence, if necessary. Also includes features of explainability, transparency and accountability.	[5], [21]
Avoiding damage	Artificial intelligence should not cause harm to people or nature and should not be able to be misused. This principle consists of several other principles, such as fairness and sustainable development and well-being.	[5], [21], [22], [23], [24], [26], [27]

RQ3. Have developers encountered ethical issues when using Generative AI tools?

RQ4. How are ethical guidelines related to AI taken into consideration in software development when using Generative AI tools?

The research methodology employed was a qualitative semi-structured interview, incorporating elements of thematic interviewing as introduced by Hirsjarvi and Hurme [28]. Invitations for interviews were sent to randomly chosen IT companies with offices in the middle part of Finland, in a region known for active software industry. A total of nine individuals with long experience in software engineering participated in the interviews, which were conducted remotely. The gathered research data were transcribed, coded and analyzed. The full interview template including all questions is available at [29].

Interview requests were sent to 74 different sized companies, of which seven companies participated in this study. From these seven companies, a total of nine IT professionals agreed to be interviewed. Seven of the interviewees have worked in IT for at least 15 years, which allows them to analyze the changes that the GenAI tools have brought to the field. Eight of the interviewees have experience as software developers, while one interviewee's responsibility seems to be the quality control of software products.

The data analysis begins with open coding in which Urquhart [30] suggests analyzing the data systematically by reading paragraphs one by one to create codes. Once the text has been coded, they can then be categorized in a phase called axial coding. According to Williams and Moser [31] in this phase, relations between the codes found in open coding are examined and then they can be categorized into groups. Williams and Moser [31] also showcase that these categories can then be combined into themes in the selective phase. To create the final themes, the relations between codes are actively compared as were the relations between categories. The research questions were then answered based on the found themes. An example of formulating codes, categories and themes is represented in Table 2. The table shows issues that were mentioned concerning the code that GenAI tools produce. Both issues pertain to the trustworthiness and quality of the generated responses and are therefore categorized under the same theme.

TABLE II. AN EXAMPLE OF CODING

Comment	Open code	Axial coding: Categories	Selective phase: Themes
"...if I just copy paste the generated response and use it as it is in my code it never works. They rarely work." (I)	Errors in code	Faulty responses	GenAI tools generate low quality and untrustworthy responses
"... they [GenAI tools] generate repetitive amateur level code..." (E)	Amateurish code	Low quality responses	

B. Findings

1) Tasks for Which GenAI Tools Are Utilized: The most popular tools used among the interviewees are GitHub Copilot and different versions of ChatGPT. Based on the interviews, the tools are used for different tasks according to their strengths and weaknesses.

While ChatGPT is used for various tasks, GitHub Copilot is a popular choice for programming and testing. Eight out of nine interviewees mention that they have used GenAI tools for programming tasks while two interviewees have also used them for testing. Three of the interviewees also compliment the quality of GitHub Copilot's autocompletion ability. GitHub Copilot's autocompletions and integration to an IDE helps to streamline different coding tasks. ChatGPT is also considered a user-friendly tool because it allows the users to have conversational interactions with the tool.

Six interviewees mention that GenAI tools can also be useful for different word processing tasks. Especially ChatGPT is used for many tasks that require word processing. The interviewees have used it for content creation for websites and documentation. One interviewee also mentions that Microsoft Copilot has potential in making presentations. Creative work. Three of the interviewees have used GenAI tools in visual and creative tasks. They have been used to generate process images, promotional videos and ideas. The tools lack understanding of causality, which is the reason they are not suitable for creating pictures that showcase logic. One of the interviewees mentions that GenAI tools seem promising in generation of game graphics, but they seem to be lacking in animation tasks.

Among the interviewees, GenAI tools were most frequently used for tasks that are already well-documented, such as programming, testing, and word processing. Despite this, the tools appear to hold considerable potential, even though significant development is still required. The participants appeared interested in experimenting with the tools to reduce the amount of repetitive or monotonous tasks, for example.

2) Benefits and Drawbacks: The interviewees identify several pros and cons of using GenAI tools. They streamline a variety of tasks and are easy to use but can generate incorrect or untrustworthy answers. Users must assess and fix low-quality outputs and know enough about the subject to create effective prompts. If problems are frequent, the tools may be seen as unnecessary.

Interviewees agree that GenAI tools can speed up tasks, though their usefulness depends on the quality of training data. They support testing, coding, and writing by generating templates and offering simple integration. GitHub Copilot, for instance, integrates into programming environments and provides automatic suggestions, which three interviewees rated as high-quality. These tools help with repetitive tasks, allowing users to focus on more interesting work, increasing job

satisfaction.

ChatGPT can substitute for Stack Overflow by helping users solve programming problems efficiently. Information retrieval is quicker, and tools provide concise answers without web browser distractions. ChatGPT is especially effective in word processing, and interviewees generally viewed it more favorably for this than for coding. Two interviewees noted that GenAI tools can outperform people in some tasks. They possess broader domain knowledge and are unaffected by fatigue, potentially reducing users' mental burden.

Despite the benefits, two interviewees do not use GenAI tools due to poor output quality. All interviewees noted the risk of incorrect answers, which require user verification and correction. Prompt quality impacts output, and ineffective prompts may force users to restart conversations. Thus, users need enough expertise to evaluate answers and formulate prompts.

Even with good prompts, hallucinations—confident but incorrect outputs—are common. If answers are unreliable, as one interviewee experienced, users may abandon the tools. Quality also varies across programming languages, with some tools struggling with niche tasks or using outdated libraries. Poor training data may lead to overly generic or unusable outputs. Three participants raised concerns about censorship and moderation, limiting use in ethically sensitive projects. One interviewee mentioned accessibility issues and memory limitations. Two interviewees noted the tools can't generate diagrams showing causality and context. Since many limitations are outside user control, it's important not to over-rely on GenAI tools and to retain traditional methods when needed.

3) Ethical Issues: Lack of knowledge on ethical issues related to GenAI and how they emerge during software development may shape interviewees' experiences. Developers often focus on their own ethical conduct and may overlook ethical problems caused by GenAI tools. Although some interviewees encountered issues while using GenAI, they might not recognize them as ethical in nature. Even without identifying specific ethical concerns, users may still avoid these tools, as in the case of one interviewee who stopped using them due to potential ethical risks. Only two interviewees explicitly mentioned ethical concerns. However, when considering inequality, such as GenAI's varying performance across programming languages, the number rises to six. Four of these six did not identify this inequality as an ethical issue, highlighting a lack of awareness.

Six interviewees bring up that GenAI tools lack transparency and three of these interviewees also mention that AI models can be black boxes. The generated answers may also be derived from copyright-protected and licensed material. The interviewees are aware of this risk, even though they have not faced it themselves while using GenAI tools. The interviewees agree that the GenAI tools can produce untrustworthy answers. GenAI tools need to be more explainable and transparent so that they can be trusted. Because the tools can generate inaccurate and incorrect information users should develop their critical thinking abilities.

Two interviewees noticed that the tools may repeat the same answers while solving programming tasks. They reported that the tools tend to replicate the same programming mistakes and problem-solving methods, which may indicate a bias.

Six of the nine interviewees also recognize that GenAI tools may cause information security risks. The tools may gather information from customers and companies to be used for its training. Users must also consider carefully what information can be given to the tool via prompts. Two interviewees also mention the communication platform, Slack, having a potential ethical issue where they use the user data to train their AI models.

AI development can also impact societal well-being and the environment. Two of the nine interviewees expressed concern about the impact that AI can have on the environment because their development and usage require a lot of resources. Three interviewees also believe that the tools may cause inequality between societies, companies and people. Four interviewees noticed that the code produced by GenAI tools varies in its quality according to the used programming language. This may also cause inequality among software developers because the training data does not represent all programming languages extensively.

In general, the interviewees are not worried about AI replacing them. This might partly be because of their long work experience in the IT field. Though, they do agree that AI can replace people to some extent in software development, especially in software testing and possibly in coding. Three interviewees are also worried about the impact that the tools might have on the professional skills of the software developers. These worries are especially related to programming when the tools can generate code quickly and effortlessly. As a result, users may be less likely to think about the problem independently but instead strives to solve the problem as fast as possible.

4) Ethical Guidelines: Five of the nine interviewees had little to no awareness of the existing ethical guidelines for AI. Though, even if the guidelines were unknown for these interviewees, the ethical terms and subjects related to ethics of AI were familiar to them. The interviewees have gotten information about the ethics of AI through the internet, colleagues and training but only two interviewees have gotten information from official sources, for example, the ethical guidelines made by AI HLEG. Four of the interviewees mention that they get most of their information about the ethics of AI from the internet by reading and listening to discussions. The responsibility of understanding and learning about the ethics of AI seem to fall on the employees.

Seven of the companies do not comply with any ethical guidelines meant for AI. Because of this the ethical discussions and decisions are left to the users of GenAI tools. This means that the users are most likely guided by their own morals and other guidelines and requirements in software development.

Two interviewees mention that their companies have methods to avoid possible ethical issues. One of the companies has appointed a group that handles ethics and informs and

guides the employees on these matters. The other company is aware of different ethical guidelines provided by the Digital and Population Data Services Agency of Finland, for example. Two interviewees also mention that it's possible to forbid the use of GenAI tools in a project to avoid ethical issues. Though, it's possible that the clients are not aware or informed about the use of these tools in their projects.

Three interviewees mention that their companies are currently updating their ethical guidelines to include AI. Companies seem to develop their processes according to the changes brought by AI but updating the guidelines and understanding the different ethical aspects takes time. One interviewee also brings up that the use of ethical guidelines of AI not only help with understanding how to use these tools ethically, but they also help the user to decide whether these tools should be used at all when making a product.

IV. Discussion

This section summarizes the main contributions of the study and discusses its limitations and threats to validity.

A. Main Contributions

The study reveals that GenAI tools are primarily used for programming, testing, and problem-solving, with occasional use in text processing. Although earlier research suggests broader applicability, real-world use remains limited, especially in visual tasks due to poor output quality. GitHub Copilot is preferred for coding, while ChatGPT is assigned more varied tasks. GenAI tools help accelerate work, but users must critically evaluate outputs due to frequent hallucinations, errors, and outdated code - especially with newer libraries. Effective use requires strong prompt design skills, as tools struggle with context, particularly in visual logic tasks.

Ethical considerations are largely overlooked in practice. Most professionals are unfamiliar with formal AI ethics guidelines, though they recognize concepts like transparency and explainability. Ethical concerns raised include data security, response repetition, and inequality due to uneven support for different languages and frameworks. While many companies lack formal ethical policies, internal discussions are increasing. Additionally, broader concerns such as unauthorized data use, data leaks, and bias are acknowledged. The findings align with earlier studies and highlight the growing need for ethical awareness as adoption increases across development workflows.

The results show that ethical guidelines related to artificial intelligence are poorly known. Based on this, it could be useful to know how companies that develop or specialize in developing artificial intelligence systems follow ethical guidelines. These companies could also provide practical advice on preventing ethical problems. The concern is that ethical guidelines related to artificial intelligence may also be ignored in companies that develop artificial intelligence systems. An interesting topic for further research would therefore be to examine the knowledge of ethics and related practices of companies that provide artificial intelligence systems.

B. Limitations and Threats to Validity

Because the sample is small, the results cannot provide a comprehensive picture of the entire software development field. The results of the study can provide an indication of typical problems, benefits and uses, as well as compliance with ethical guidelines, but they lack more specific perspectives on, for example, the use of tools in game development.

Some interviewees reported having only limited experience with GenAI tools, which may have constrained the depth of their engagement. Nonetheless, even these participants were able to articulate a range of challenges they had encountered, with many indicating that such difficulties constituted the primary reason for their discontinuation of tool use. Furthermore, the specific purposes for which study participants used GenAI tools do not necessarily reflect typical usage patterns, but rather serve as illustrative examples of the tools' potential applications.

V. Conclusion and Future Work

Addressing ethical issues related to AI and large language models is vital for achieving trustworthiness of the tools and requires a multidisciplinary approach involving professionals from the fields of technology, ethics, law and policy. However, it seems that even though the software professionals are familiar with the concepts of ethical principles related to AI, they do not know the ethical guidelines provided by several organizations, and the employer companies do not utilize the guidelines.

The issues related to transparency and explainability of the generated answers were particularly familiar to the interviewees. In addition, it seems that recurring responses generated by the tools seem to be an issue. Concerns were also raised about the security of the tools. Furthermore, many felt that the quality of the responses provided by the tools was insufficient.

This study illustrates and raises discussion on ethical issues that may arise in software development, as well as examines the awareness of AI ethics among software professionals. Currently, GenAI tools support various work tasks, but they have several weaknesses that require users to actively evaluate the quality and reliability of responses. Additionally, not all users benefit equally from these tools. While guidelines published by different organizations aim to prevent potential ethical issues and enable the development of reliable AI systems, it appears that GenAI tools still have significant room for improvement.

Since it is expected that awareness of ethical problems caused by AI systems will increase among users, companies will likely create their own guidelines to avoid them. In the future, it may be useful to find out how considering the ethical perspective is developing in companies that utilize AI tools, or whether the developers of the tools have created their own guidelines.

References

[1] Z. Özpolat, Ö. Yildirim, and M. Karabatak, "Artificial intelligence-based tools in software development processes: Application of chatgpt," *Eur. J. Technic*, vol. 13, no. 2, pp. 229–240, 2023.

[2] C. Ebert and U. Hemel, "Technology trends 2023: The competence challenge," *IEEE Softw.*, vol. 40, no. 3, pp. 20–28, 2023.

[3] L. Weidinger *et al.*, *Ethical and social risks of harm from language models*, https://arxiv.org/abs/2112.04359, retrieved: August, 2025, 2021.

[4] H. Tanaka *et al.*, "Taxonomy of generative ai applications for risk assessment," in *Proc. IEEE/ACM 3rd Int. Conf. AI Eng.-Softw. Eng. AI*, 2024, pp. 288–289.

[5] AI HLEG, *Ethics guidelines for trustworthy ai*, https://www.europarl.europa.eu/meetdocs/2014_2019/plmrep/COMMITTEES/JURI/DV/2019/11-06/Ethics-guidelines-AI_FI.pdf, retrieved: June, 2025, Apr. 2019.

[6] M. Ryan and B. C. Stahl, "Artificial intelligence ethics guidelines for developers and users: Clarifying their content and normative implications," *J. Inf. Commun. Ethics Soc.*, vol. 19, no. 1, pp. 61–86, 2021.

[7] O. Asare, M. Nagappan, and N. Asokan, "Is github's copilot as bad as humans at introducing vulnerabilities in code?" *Empirical Software Engineering*, vol. 28, no. 129, 2023.

[8] M. A. Haque and S. Li, "The potential use of ChatGPT for debugging and bug fixing," *EAI Endorsed Trans. AI Robot.*, vol. 2, 2023.

[9] T. Kumari and A. Das, "Towards development of an effective ai-based system for relevant comment generation," in *Proc. 15th Annu. Meet. Forum Inf. Retrieval Eval.*, 2023, pp. 118–120.

[10] G. Giunti and C. P. Doherty, "Cocreating an automated mhealth apps systematic review process with generative ai: Design science research approach," *JMIR Med. Educ.*, vol. 10, e48949, 2024.

[11] P. Vaithilingam, T. Zhang, and E. L. Glassman, "Expectation vs. experience: Evaluating the usability of code generation tools powered by large language models," in *CHI Conf. Hum. Factors Comput. Syst. Ext. Abstr.*, 2022, pp. 1–7.

[12] A. M. Dakhel *et al.*, "Github copilot ai pair programmer: Asset or liability?" *J. Syst. Softw.*, vol. 203, p. 111 734, 2023.

[13] M. Schäfer, S. Nadi, A. Eghbali, and F. Tip, "An empirical evaluation of using large language models for automated unit test generation," *IEEE Trans. Softw. Eng.*, vol. 50, no. 1, pp. 85–105, 2024.

[14] P. Khlaisamniang, P. Khomduean, K. Saetan, and S. Wonglapsuwan, "Generative ai for self-healing systems," in *Proc. 18th Int. Joint Symp. Artif. Intell. Nat. Lang. Process. (iSAI-NLP)*, 2023.

[15] V. Bilgram and F. Laarmann, "Accelerating innovation with generative ai: Ai-augmented digital prototyping and innovation methods," *IEEE Eng. Manag. Rev.*, vol. 51, no. 2, pp. 18–25, 2023.

[16] A. Barcaui and A. Monat, "Who is better in project planning? generative artificial intelligence or project managers?" *Project Leadership and Society*, vol. 4, p. 100 101, 2023.

[17] D. Luitel, S. Hassani, and M. Sabetzadeh, "Improving requirements completeness: Automated assistance through large language models," *Requirements Eng.*, vol. 29, pp. 73–95, 2024.

[18] H. Hochmair, L. Juhász, and T. Kemp, "Correctness comparison of chatgpt-4, gemini, claude-3, and copilot for spatial tasks," *Trans. GIS*, vol. 28, no. 7, pp. 2219–2231, 2024.

[19] B. Idrisov and T. Schlippe, "Program code generation with generative ais," *Algorithms*, vol. 17, no. 2, p. 62, 2024.

[20] S. Barke, M. B. James, and N. Polikarpova, "Grounded copilot: How programmers interact with code-generating models," *Proc. ACM Program. Lang.*, vol. 7, pp. 85–111, 2023.

[21] Google, *Our principles*, https://ai.google/responsibility/principles/, retrieved: August, 2025, Oct. 2023.

[22] IBM, *IBM artificial intelligence pillars*, https://www.ibm.com/policy/ibm-artificial-intelligence-pillars/, retrieved: August, 2025, Aug. 2023.

[23] IEEE Standard 7001, *IEEE standard for transparency of autonomous systems*, IEEE, 2022.

[24] IEEE Standard 7002, *IEEE standard for data privacy process*, IEEE, 2022.

[25] Adobe, *Adobe's Commitment to AI Ethics*, https://www.adobe.com/content/dam/cc/en/ai-ethics/pdfs/Adobe-AI-Ethics-Principles.pdf, retrieved: June, 2025, Oct. 2023.

[26] IEEE Standard 7009, *P7009 standard for fail-safe design of autonomous and semi-autonomous systems*, IEEE, 2024.

[27] IEEE Standard 7010, *IEEE recommended practice for assessing the impact of autonomous and intelligent systems on human well-being*, IEEE, 2020.

[28] S. Hirsjärvi and H. Hurme, *Tutkimushaastattelu : teemahaastattelun teoria ja käytäntö*, H. Hurme, Ed. Helsinki: Yliopistopaino, 2000, retrieved: February, 2025. [Online]. Available: https://jyu.finna.fi/Record/jykdok.829757.

[29] T. Tuomisto, "Generatiivisen tekoälyn käyttö ja etiikka ohjelmistokehityksessä," retrieved: August, 2025, M.S. thesis, University of Jyväskylä, 2024. [Online]. Available: https://urn.fi/URN:NBN:fi:jyu-202411287512.

[30] C. Urquhart, *Grounded Theory for Qualitative Research: A Practical Guide*. 55 City Road: SAGE Publications, Ltd, 2013.

[31] M. Williams and T. Moser, "The art of coding and thematic exploration in qualitative research," *Int. Manag. Rev.*, vol. 15, no. 1, pp. 45–55, 2019.