









# Responsible AI in Agile Software Engineering - An Industry Perspective

Rasmus Ulfsnes<sup>1,7</sup><sup>(✉)</sup>, Nils Brede Moe<sup>1</sup>, Jostein Emmerhoff<sup>2</sup>,  
Marcin Floryan<sup>3</sup>, Anastasia Griva<sup>4</sup>, Jan Henrik Gundelsby<sup>5</sup>,  
Astri Moksnes Barbala<sup>1</sup>, and Kieran Conboy<sup>1,6</sup>

<sup>1</sup> SINTEF, 7034 Trondheim, Norway

Rasmus.Ulfsnes@sintef.no

<sup>2</sup> Schibsted Nordic Marketplace, Oslo, Norway

<sup>3</sup> Spotify, Stockholm, Sweden

<sup>4</sup> Bespot, Athens, Greece

<sup>5</sup> Knowit Solutions AS, Oslo, Norway

<sup>6</sup> University of Galway, Galway, Ireland

<sup>7</sup> NTNU, Trondheim, Norway

**Abstract.** There is a rapid emergence of tools, methods, and guidance for the use of AI across all parts of the software development process, from requirements gathering to code generation to testing and user feedback. However, AI raises many concerns regarding responsible use, and there is a need to understand and develop principles for what responsible software development entails in practice in an agile context, as well as carefully evaluate the incorporation of AI tools and methods in software engineering. We draw on experience from Bespot, Knowit, Schibsted, and Spotify to identify challenges faced by companies pioneering the use of AI in their software development efforts and start charting a roadmap for responsible AI in software engineering.

**Keywords:** GenAI · Responsible AI · Agile Software Engineering · Challenges · Responsible AI · Software Engineering · Industry Perspective

## 1 Introduction

The landscape of business is rapidly evolving with the integration of artificial intelligence, particularly in the realm of software development. The rise of Generative AI (GenAI) has been remarkable, offering transformative capabilities that enhance various segments of the development lifecycle—from requirements management [4] to code generation [7] and security testing [6]. The primary focus of these advancements has been to drive efficiency, automate routine tasks, and increase productivity [15]. Nonetheless, there is an increasing imperative to address the ethical dimensions of AI deployment. To date, however the literature on what constitutes responsibility within the software engineering field has

largely been discussed in a separate body of literature to the literature on AI tools in software engineering, and very few studies in either camp are built on empirical data. This paper aims to bring the two communities together, shedding light on which challenges software organizations face in terms of leveraging AI responsibly and outlining a way forward in solving these. Given the recent and rapid emergence of this area, we asked four key experts with extensive industry experience in large-scale agile organizations to provide written statements about challenges associated with responsible AI in their organizational context with regard to software engineering. Our selection was based on software-intensive companies impacted by and in the process of using AI in their engineering processes.

## 2 Background

In looking into what responsibility means in software engineering by reviewing the current understanding of ethical principles in the field, Ina Schieferdecker [12] asserts that software trustworthiness today hinges more on acceptance than technical quality, emphasizing that software and its features must be comprehensible and explainable. Software and its applications can only succeed if they garner public trust, Schieferdecker notes, which is tied to users' belief that products have been developed according to responsible principles.

Otherwise, literature on the topic in the software engineering field has largely focused on literature reviews. One such study, a rapid review study focusing on what responsible AI means in software engineering, was conducted by Barletta et al. [1]. They investigated frameworks that provide principles, guidelines, and tools designed to aid practitioners in the development and implementation of responsible AI applications. In analyzing each framework in relation to the various phases of the Software Development Life Cycle (SDLC), Barletta et al. found that the majority of these frameworks are focused primarily on the Requirements Elicitation phase, with minimal coverage of other phases. Barletta et al.'s findings thus indicate the absence of a comprehensive "catch-all" framework that effectively supports both technical and non-technical stakeholders in the execution of real-world projects. Similarly, Lu et al. [9] conducted a systematic literature review on responsible AI for software engineering to summarize the current state and identify critical research challenges. They present a research road map on software engineering to operationalize responsible AI. Some of the findings are proposed as tools, such as ethical risk assessments, or as product features embedded within AI systems to mitigate ethical risks and enhance trust in markets where it is currently lacking, for instance, an ethical black box.

In regard to the literature on responsible AI technologies, a multitude of factors are prominent, covering both human, social, and organizational factors. For instance, Mikalef et al. [11] point to 8 dimensions for responsible AI. However, there are different approaches to achieving responsible AI, and one venue is the concept of explainable AI [10], wherein efforts are being made to outline how and at what level different stakeholders need and understand the outputs of AI.

Another venue is that of domain expertise [14] where the argument for bridging the experts of AI, with the experts of whatever domain the technology would assist. Collaborating with the AI-systems developers, however, is not that easy when you are purchasing off-the-shelf AI technology, e.g., Copilot or ChatGPT for software engineers, to assist in their programming tasks (see [13]).

### 3 Approach

As AI in software engineering is a novel phenomenon, and there are studies and research on the topic, we argue that utilizing a Delphi-type approach is appropriate [8]. The Delphi method can be used both quantitatively and qualitatively. We sourced four experts from different software organizations to elicit their take on responsible AI and how it affects their organization. To guide our inquiry, we utilize the eight dimensions proposed by Mikalef et al. [11]: *Fairness, Transparency, Accountability, Robustness and safety, Data governance, Laws and regulations, Human oversight* and *Societal and environmental well-being*. After eliciting the information, we analyzed the challenges the different organizations experienced in terms of grappling with responsible AI.

All four industry examples adhere to the key principles of Agile, which include incrementally developing the software in iterative cycles, implementing regular ceremonies to review and refine both the product and development methods, collaboratively responding to changes, and consistently engaging with users. Additionally, the software teams within these organizations are organized in a manner typical of agile teams.

## 4 Industry Perspectives on Responsible AI in Software Engineering

### 4.1 Bespot - Recruiting Skilled Expertise

In recent years, a significant challenge we've encountered is related to the hiring process for software developers and AI experts. Traditionally, companies have relied on assessing candidates' experience by reviewing their profiles on web-based platforms like GitHub and StackOverflow. This approach allowed us to initially assess their coding abilities, problem-solving skills, and overall expertise. However, with the rise of GenAI we have begun to question the efficacy of using developers' profiles on such platforms as part of our talent screening process.

One issue we identified is the potential for inaccuracies in candidates' profiles, which may not truly reflect their coding skills or contributions to the community. Some discrepancies are apparent upon closer scrutiny of platform data, such as sudden improvements in ratings, reputation, or badges. Still, efforts from the companies' side are required to detect such profiles. Also, it is difficult to figure out whether something is GenAI written in other cases. This situation seems to result in inequalities in evaluating and hiring talents.

While web-based platforms like GitHub and StackOverflow remain valuable resources for assessing candidates, of course software development companies do not only use these. There might also be internal coding tests/challenges, etc., as part of the hiring process. Still, even in such cases, GenAI was detected to be used profoundly or not, affecting hiring once again. This increasing prevalence of GenAI has prompted internal debates among companies regarding its responsible use in hiring practices.

On the one hand, some state that using GenAI to generate code is acceptable, and achieving an optimal equilibrium between automated processes and human intuition is essential in coding. This is reinforced by trends in certificates such as prompt engineering for GenAI. However, detractors caution against relying on AI as a collaborator, citing concerns about perpetuating inequalities and potential risks to the company's integrity. The latter can happen since we are not certain about where the data are stored, who has access, etc.

For example, our company, Bespot, develops location fraud detection and validation software solutions. The company has developed an AI solution utilizing tracking technologies (e.g., WiFi, GPS, cellular) to detect user locations with near-centimeter precision accurately. However, protecting our competitive advantages is crucial since these algorithms are proprietary and treated as black boxes. Consequently, hiring individuals who may inadvertently expose sensitive algorithms to GenAI collaboration poses a significant risk, particularly for companies operating in sectors requiring stringent data protection measures.

In conclusion, navigating the intersection of GenAI and hiring practices presents challenges for companies seeking to maintain a balance between human-GenAI collaboration and responsibility. While leveraging AI technologies offers potential benefits, careful consideration of ethical implications, data security concerns, and competitive interests is important in ensuring responsible decision-making within the hiring process.

## 4.2 Knowit - Security, Sustainability, and GenAI's Mental Models

While the potential of Generative AI is undeniable, its integration into practical, real-world applications comes with significant challenges. Knowit is a large consultancy firm focused on digital transformation. It combines IT, design, and management with an emphasis on security, cloud, and AI services. At Knowit, we are committed to sustainable practices and human rights. Despite over a year of democratized access to GenAI, our clients are still primarily in the exploration phase, hesitant to fully embrace its potential. We believe this hesitation stems from several fundamental issues, including concerns about security, transformative use of technology, and environmental and economic sustainability.

The main challenge is related to security, uncertainty of regulations, privacy concerns, and a large unknown attack surface through a plethora of chatbots. All this makes it difficult and too risky for our customers to put the technology to production use. Additionally, 'hallucinations'-incorrect or nonsensical information generated by these systems-pose another significant challenge. Our mental models of computer technology usually let us think about data as a fact or

something that is deterministic, predictable, and reliable. However, GenAI operates differently; it is based on statistics and probabilities. This unpredictability requires us to rethink the way we understand and use this technology in our systems and daily work. For instance, while tools like GitHub Copilot offer coding assistance, concerns about energy consumption, code quality, and socio-technical impacts on team collaboration continue to raise doubts about their long-term productivity benefits. Another major sustainability challenge is the substantial energy consumption associated with GenAI. For example, a single Chat GPT query consumes fifteen times more energy than a standard Google search, highlighting the environmental impact of this technology. Additionally, the lack of clear revenue generation from GenAI investments raises concerns about its long-term economic sustainability. For instance, the venture capital firm Sequoia estimated that the AI industry spent \$50 billion on Nvidia chips to train advanced AI models last year but generated only \$3 billion in revenue. Knowit recognizes GenAI's transformative potential but also acknowledges the significant challenges associated with its adoption in real-world applications and the sustainability challenges it brings. Addressing these issues is essential for leveraging GenAI effectively and responsibly, ensuring both environmental and economic viability.

### 4.3 Spotify on Algorithmic Responsibility

Every new technology should be approached with a healthy dose of skepticism. This becomes harder when you see everyone around you jumping on the bandwagon. Fortunately, at Spotify, we have over a decade of experience using machine learning and artificial intelligence to enhance our products, especially in the recommendation space. As a result, Spotify has been exposed to some of the challenges inherent in using this technology, specifically in terms of algorithmic bias. For example, we want to avoid recommendations that skew towards the artist's gender or towards more popular songs from certain artists. As part of acting responsibly in this space, we have invested to avoid unintended algorithmic harm. Our research into [algorithmic responsibility](#) is helping us to avoid the challenges. As AI tools become more popular and start powering more features such as AI DJ or the AI Playlist Generation, we work to ensure that we build a fair product, respects inclusion and diversity, and does not lead to discriminatory outcomes. Another aspect relevant for Spotify, in the area of responsible use of AI, is to consider the environmental impact, especially in the view of our [climate action](#) and responsibility towards the climate crisis. This applies both to Spotify using AI as part of our product portfolio and our use of tools such as Large Language Models that help with the day to day tasks of our employees.

### 4.4 Schibsted Nordic Marketplaces on Governance and Learning

Schibsted Nordic Marketplaces (NMP) offers digital marketplaces for real estate, job listings, mobility services, and classified ads. It is the leading company in the Nordics, with significant market shares in Norway, Finland, Sweden, and

Denmark. We see AI as fundamental in two aspects. 1) The use of AI services will be incorporated in new products based on our large data sets, and 2) AI tools will also be integrated into the company’s development practices through the likes of Copilot and other GenAI tools. Just as Apple revolutionized digital marketplaces with the iPhone, AI technologies can bring about similar significant changes to our products and the way we deliver the products. The new technology will change how we operate and affect the daily lives of employees. NMP needs to develop insights and knowledge about how to use commercial GenAI models and deploy AI solutions responsibly.

In practice, “Responsible AI” involves establishing guidelines, processes, and mechanisms to ensure that AI technology is available, easy to use, and implemented in line with the organization’s values and goals while adhering to regulations and ethical perspectives for fairness and sustainability. This can be seen as a lesson learned from the move to cloud services, which do not work without a defined governance structure. As with cloud services, this means taking responsibility seriously in the procurement process and will influence the vendors we choose for such products. This will pose a challenge for our software development teams with a high degree of autonomy regarding technology choice and how they work.

AI tools can influence collaboration and knowledge sharing within the organization. For example, internal communication and coordination within teams may change, so governance structures are needed to support collaboration and knowledge sharing. A key aspect will be investing in raising employees’ competencies to leverage AI effectively. It’s not enough for individuals to learn, as learning together is necessary to develop new practices for the use of technology in a responsible way. And without knowledge of the technology and how to use it responsibly, you won’t be able to do your work well.

## 5 Discussion and Future Research

Based on the industry experts’ statements, we have identified several challenges related to responsible AI in agile software engineering. The most prominent are:

**Finding a Balance Between Human and AI.** As of now, the impact of GenAI, e.g. ChatGPT, is not understood in terms of its long-term effects on software engineering practices and the social processes involved in these practices. This raises concerns for our industry partners, already from before the engineers are hired, raising questions about the eligibility of the candidates and how to manage this from a recruiting standpoint. This might affect the fairness of the hiring processes [11] as new hires are no longer selected on equal terms. However, there is also a need to utilize AI’s positive effects, e.g., productivity [7], and finding this balance while remaining responsible is challenging.

**Unclear Effects on Communication and Collaboration.** While balancing human and AI automation is challenging, some effects go outside the individual use of AI tools. One notable concern is how this will affect teams and organizations and how they deal with learning [13]. This is particularly concerning as

large-scale agile organizations are dependent on the communication, collaboration, and knowledge sharing that occurs in and between teams. One approach here is to use governance that limits and sets boundaries on tools and practices for using AI, but this has a cost in terms of reduced autonomy in large-scale agile contexts [2].

**Managing Data Governance and Hallucinations.** Data governance and privacy issues do not just create challenges regarding AI, but the interest and accessibility of the tools are making it particularly challenging to leverage the technology. Individual developers must manage the data governance themselves [13], which can be challenging for the developers [4]. Additionally, the uncertainty of regulatory bodies makes it difficult for organizations to make good decisions. Moreover, the data that comes out might be the effects of hallucinations, which require developers to learn how to deal with bad code suggestions [7] and advice [13]. Nevertheless, there are also positive effects of using AI, in achieving greater security posture of the software developed [6].

**Managing Responsible AI in Software Products.** Managing and dealing with the practical issues of using various AI tools are quite challenging. Companies also want to embed these technologies into their products, providing new interactive interfaces or recommendations. This means making development processes that especially consider the potential of algorithmic harm to ensure fairness, transparency, and accountability [11]. While the companies aim to avoid these issues, there is a lack of frameworks and processes for managing this in the software development life cycle [1].

## 6 Conclusions and Future Work

Organizations are being met with ever-increasing pressure to allow individuals to use GenAI for their activities while also wanting to exploit and explore the potential of both GenAI and AI in their products and services.

According to our findings, organizations need to deal with challenges on different organizational levels: 1) Organization, 2) Team, and 3) Individual, as these challenges are interrelated between the different parts of an organization and need to be managed simultaneously.

What remains, however, is a clear approach to dealing with human-AI collaboration for agile organizations. There are five different ways to look at human-AI collaboration in the organization according to Kolbjörnsson [5]: 1) Individuals working without AI, 2) Collective, multiple people working together, 3) Automated, when work is done without human interference, 4) Augmented individuals, doing work together with AI, and 5) Augmented teams, when multiple people collaborate with AI.

What recent studies have shown, both experimental [3] and real-life settings [13], is that the exploration and subsequent use of GenAI is largely done by individuals and organizations seem to have a goal of automating work, and thus becoming more efficient.

As more and more organizations race towards more automated work, and thus becoming more efficient, there is a risk that we lose out on the decades of research on agile in organizations, putting a focus on collaboration and coordination. We, therefore argue that organizations and researchers should look into how collectives, such as agile teams, and organizations together can collaborate with Artificial Intelligence, be it generative or otherwise.

## References

1. Barletta, V.S., Caivano, D., Gigante, D., Ragone, A.: A rapid review of responsible AI frameworks: how to guide the development of ethical AI. In: Proceedings of the 27th International Conference on Evaluation and Assessment in Software Engineering, pp. 358–367 (2023)
2. Bass, J.M., Haxby, A.: Tailoring product ownership in large-scale agile projects: managing scale, distance, and governance. *IEEE Softw.* **36**(2), 58–63 (2019). <https://doi.org/10.1109/MS.2018.2885524>, [https://ieeexplore.ieee.org/abstract/document/8648277?casa\\_token=QQesxZ-4c\\_4AAAAA:F80w6k0IDwalLeXRO2QnHGnIMb4oZQV7JDkwh-pyenQR3DdZwtwmnxgF8XnHHZtEzfrVE-9c9ng](https://ieeexplore.ieee.org/abstract/document/8648277?casa_token=QQesxZ-4c_4AAAAA:F80w6k0IDwalLeXRO2QnHGnIMb4oZQV7JDkwh-pyenQR3DdZwtwmnxgF8XnHHZtEzfrVE-9c9ng)
3. Bubeck, S., et al.: Sparks of artificial general intelligence: early experiments with GPT-4. *arXiv:2303.12712 [cs]* (2023)
4. Ebert, C., Louridas, P.: Generative AI for software practitioners. *IEEE Softw.* **40**(4), 30–38 (2023). <https://doi.org/10.1109/MS.2023.3265877>, <https://ieeexplore.ieee.org/abstract/document/10176168>
5. Kolbjørnsrud, V.: Designing the intelligent organization: six principles for human-AI collaboration. *Calif. Manage. Rev.* **66**(2), 44–64 (2024). <https://doi.org/10.1177/00081256231211020>, <http://journals.sagepub.com/doi/10.1177/00081256231211020>
6. Li, J., Meland, P.H., Notland, J.S., Storhaug, A., Tysse, J.H.: Evaluating the impact of ChatGPT on exercises of a software security course. In: 2023 ACM/IEEE International Symposium on Empirical Software Engineering and Measurement (ESEM), pp. 1–6. IEEE (2023). <https://ieeexplore.ieee.org/abstract/document/10304857/>
7. Liang, J.T., Yang, C., Myers, B.A.: A large-scale survey on the usability of AI programming assistants: successes and challenges. In: Proceedings of the IEEE/ACM 46th International Conference on Software Engineering, ICSE 2024, pp. 1–13. Association for Computing Machinery, New York (2024). <https://doi.org/10.1145/3597503.3608128>, <https://dl.acm.org/doi/10.1145/3597503.3608128>
8. Lilja, K.K., Laakso, K., Palomäki, J.: Using the Delphi method. In: 2011 Proceedings of PICMET 2011: Technology Management in the Energy Smart World (PICMET), pp. 1–10 (2011). <https://ieeexplore.ieee.org/abstract/document/6017716>. iSSN 2159-5100
9. Lu, Q., Zhu, L., Xu, X., Whittle, J., Xing, Z.: Towards a roadmap on software engineering for responsible AI. In: Proceedings of the 1st International Conference on AI Engineering: Software Engineering for AI, pp. 101–112 (2022)
10. McDermid, J.A., Jia, Y., Porter, Z., Habli, I.: Artificial intelligence explainability: the technical and ethical dimensions. *Philos. Trans. Roy. Soc. A: Math. Phys. Eng. Sci.* **379**(2207), 20200363 (2021). <https://doi.org/10.1098/rsta.2020.0363>, <https://royalsocietypublishing.org/doi/full/10.1098/rsta.2020.0363>



11. Mikalef, P., Conboy, K., Lundström, J.E., Popovič, A.: Thinking responsibly about responsible AI and ‘the dark side’ of AI. *Eur. J. Inf. Syst.* **31**(3), 257–268 (2022)
12. Schieferdecker, I.: Responsible software engineering. *Future Softw. Qual. Assur.* 137–146 (2020)
13. Ulfesnes, R., Moe, N.B., Stray, V., Skarpen, M.: Transforming software development with generative AI: empirical insights on collaboration and workflow. In: Nguyen-Duc, A., Abrahamsson, P., Khomh, F. (eds.) *Generative AI for Effective Software Development*, pp. 219–234. Springer, Cham (2024). [https://doi.org/10.1007/978-3-031-55642-5\\_10](https://doi.org/10.1007/978-3-031-55642-5_10)
14. Waardenburg, L., Huysman, M.: From coexistence to co-creation: blurring boundaries in the age of AI. *Inf. Organ.* **32**(4), 100432 (2022)
15. Ziegler, A., et al.: Productivity assessment of neural code completion. In: *Proceedings of the 6th ACM SIGPLAN International Symposium on Machine Programming, MAPS 2022*, pp. 21–29. Association for Computing Machinery, New York (2022). <https://doi.org/10.1145/3520312.3534864>, <https://dl.acm.org/doi/10.1145/3520312.3534864>

**Open Access** This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter’s Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter’s Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

