

Peer Review

Review of: "Challenges and Paths Towards AI for Software Engineering"

Jaime Díaz¹

1. Universidad de La Frontera, Chile

This manuscript offers a thoughtful and timely position paper on AI's current state and future directions for Software Engineering (AI4SE). The authors do a commendable job surveying a wide range of relevant work, weaving in practical examples and referencing benchmarks that illustrate current progress and persistent challenges. Given the rapid evolution of large language models (LLMs) and their growing role in software engineering workflows, this paper is relevant and well-placed.

The structure is clear and purposeful. The three-part framework—task taxonomy, existing challenges, and research directions—gives the paper a solid backbone. The proposed taxonomy, which categorizes tasks by scope, logical complexity, and level of human intervention, is particularly useful and could help shape future evaluations. Including detailed real-world case studies, such as the React Fiber refactor or Scala migration, also strengthens the work's practical grounding.

One of the standout aspects is how the paper identifies and organizes the key cross-cutting challenges: issues like benchmark contamination, tool integration, human-AI collaboration gaps, and the difficulties of operating in low-resource or long-context settings are well-articulated and deserving of deeper research.

That said, the manuscript also has some notable weaknesses that limit its impact and clarity in its current form.

First, the paper lacks methodological transparency. It is unclear how the literature was selected or how the taxonomy and benchmark coverage were determined. As a result, some sections read more like curated commentary than a systematically constructed review.

Second, while the simplified autonomy model (low/medium/high) is a helpful conceptual tool, its application across different tasks is vague. No clear metric or rationale for assigning or comparing these autonomy levels weakens its practical utility.

Third, the paper sometimes blurs the line between AI4SE (using AI for software engineering tasks) and

SE4AI (engineering AI systems). This distinction is important—especially when discussing tool integration or verification pipelines—and the paper would benefit from more precisely separating these perspectives.

Another issue is the sheer volume of references and benchmarks. While the coverage is impressive, it can become overwhelming. Readers are left without a critical synthesis or hierarchy of what matters most. Tables, taxonomies, or some form of prioritization would go a long way in making the discussion more digestible and actionable.

Lastly, while Section 4 is rich in ideas, many proposed research directions are presented without enough feasibility discussion. Suggestions like symbolic tool integration or agent-based debugging frameworks are intriguing, but the paper does not fully address the engineering or computational challenges they entail.

Declarations

Potential competing interests: No potential competing interests to declare.