

HOW SOFTWARE ENGINEERING TEAMS ARE USING AI TO BE MORE EFFECTIVE



WHITE PAPER

2024

CONTENTS

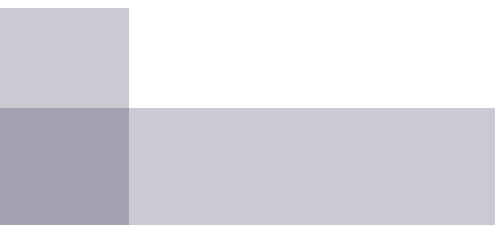
03 GENERATIVE AI: IS IT JUST HYPE OR THE NEXT INDUSTRIAL REVOLUTION?

04 WHO'S ALREADY USING AI IN THEIR DAILY WORKFLOW?

06 BEYOND GPT: WHY GENERAL-PURPOSE TOOLS AREN'T ENOUGH TO BOOST TEAM PRODUCTIVITY

07 USE CASES IN SOFTWARE ENGINEERING

10 CHALLENGES AND OPPORTUNITIES IN THE USE OF AI FOR COMPANIES



GENERATIVE AI: IS IT JUST HYPE OR THE NEXT INDUSTRIAL REVOLUTION?

It's no exaggeration to claim that nowadays, 9 out of 10 conversations touch on AI. Whether you're browsing social media or reading the news, you'll be bombarded with information about artificial intelligence.

Specifically, generative AI has become the internet's latest sensation, thanks in large part to OpenAI and ChatGPT.

Since its debut in 2022, half of the people involved believe we are nearing AGI (Artificial General Intelligence), while the other half see it as just a more sophisticated text generator.

Regardless of who's right, our goal is to provide a comprehensive view of how real teams are integrating AI into their daily workflows.

By the end of this, you'll be equipped to assess whether implementing this technology in your team is worthwhile, identifying both the benefits and the risks associated with its use.

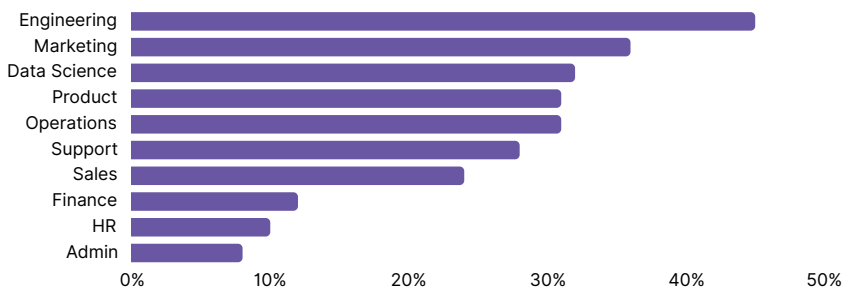


As of this writing, NVIDIA is the world's most valuable company, fueling speculation about whether AI might be the next bubble.

WHO'S ALREADY USING AI IN THEIR DAILY WORKFLOW?

According to a study by Retool, engineering, marketing, and data science teams are the most frequent users of AI in their daily operations.

This trend is largely driven by use cases, with teams involved in text generation (code and content) and task automation benefiting the most.



Source: Retool Study

If you think your team isn't using AI just because there isn't a clear policy or dedicated internal tool, you might be wrong.

28% are using AI without their bosses knowing.

Source: Retool Study



The situation becomes problematic when we consider that around **11% of the information entered into ChatGPT is confidential.**

Source: Cyberhaven

Devs Artificial Intelligence

In a recent study, GitHub surveyed companies in the United States with over 1,000 employees and found that 92% of developers are already incorporating artificial intelligence into their programming processes, whether at work or outside of it.

According to them, these tools improve code quality, increase speed, and reduce production errors.

The study highlights an interesting finding: when teams adopt AI for coding, their daily routines transform. Previously focused primarily on coding and bug fixing, now more time is dedicated to code reviews and security.



92% of developers say they are using AI daily.

It's a mistake to think that these tools are only present in more operational teams. Leadership is also increasingly incorporating AI technologies.

56%

of engineering and product leaders report using AI daily and recognize that teams utilizing this technology have a greater potential to add business value.

BEST LEADER!

Source: Gartner 2024

BEYOND GPT: WHY **GENERAL-PURPOSE TOOLS** **AREN'T ENOUGH** TO BOOST TEAM PRODUCTIVITY

Although GPT has brought significant advancements, its use is still limited to less complex tasks. This is because generic LLMs like GPT lack integration with the day-to-day tools used by engineering teams, as well as the context of the projects and code being developed.

For AI to truly drive productivity, it is essential that it integrates with the tools already used by teams, such as IDEs, version control systems, and project management tools.



USE CASES IN SOFTWARE ENGINEERING

IMPROVEMENTS IN CODE REVIEW

Google conducted an interesting study on the use of AI to streamline code reviews. The company developed a tool that uses machine learning to suggest changes based on reviewers' comments. Google's AI assistant proposes code modifications as soon as a reviewer comments.

The main benefit reported was the reduction in code review time, saving thousands of hours for the team over the course of a year.

AUTOMATED TEST GENERATION

Goldman Sachs is another major company that achieved significant progress by using AI, specifically for creating automated tests. With this technology, they were able to develop unit tests for their Java code, completing a year's worth of work in just one night.

The strategy involved integrating automation directly into their CI tool, which understood the code's behavior to generate test cases, covering a broad portion of the codebase.

The primary benefit was the significant time savings, allowing the team to focus on other important tasks.

USE CASES IN SOFTWARE ENGINEERING

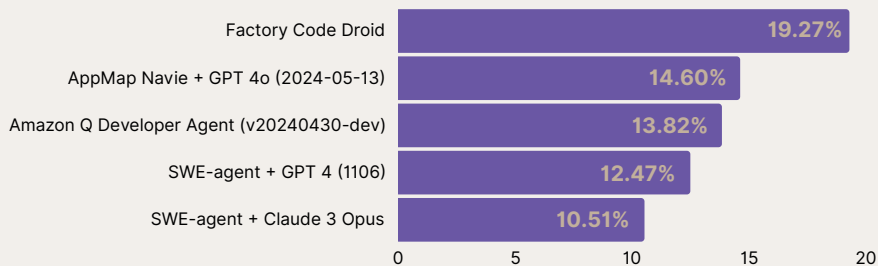
IDENTIFICATION AND BUG FIXING

In another interesting study, Google uses Artificial Intelligence to identify and fix simple vulnerabilities in code.

The results were also notable, with 15% of the bugs found by the tools being fixed. The most effective fixes addressed simpler errors, such as uninitialized variables, but also included corrections for more complex issues like memory leaks or data races.

Get to Know SWE-Bench

All the use cases mentioned above were applied in code interaction scenarios. For these cases, we use a specific evaluation system called SWE-Bench.



% refers to the percentage of SWE-bench instances (a total of 2,294) that were resolved by the model.

Source: SWE-bench

USE CASES IN SOFTWARE ENGINEERING

AUTOMATION OF STATUS REPORTS

Leaders spend an average of 1 hour per day responding to questions from the team and stakeholders. In this context, an LLM can be the team's best ally.

Voltz automated various "check-ins" to update the progress of teams at different levels. From daily check-ins to help the team prepare for stand-up meetings to sprint retrospective check-ins, describing the main challenges faced during the period and the work completed by the team.

This brought greater alignment to the team and freed up time for leadership to focus on more important initiatives rather than compiling status reports.

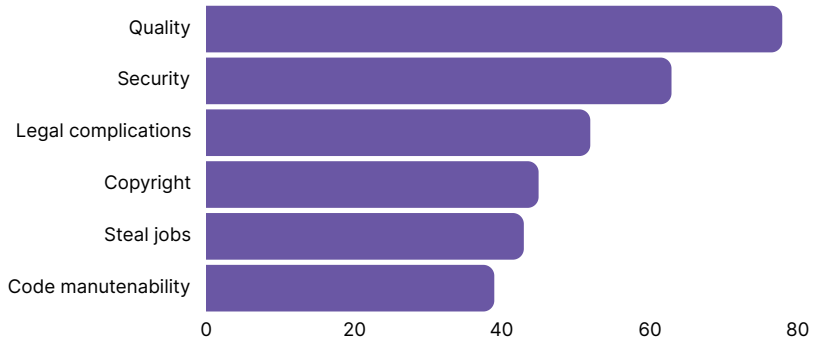
METRICS DEEP DIVE

Many teams fall into the trap of having dashboards with dozens of metrics without being able to extract useful insights. By properly contextualizing the LLM, it is possible to summarize what happened over a specific period, helping to identify root causes.

Voltz adopted this approach and significantly reduced the time spent analyzing metrics while also correlating the root causes of issues with greater precision.

CHALLENGES AND OPPORTUNITIES IN THE USE OF AI FOR COMPANIES

Source: The Engineering Leadership Report 2024 (Leadev)



Main Concerns of Leadership Regarding the Use of AI by Their Teams

Quality of the Generated Output

Affectionately called hallucination, the architecture of an LLM is intrinsically based on probabilities conditioned by the input text and the model's learning during training, rather than being a deterministic model. These models are trained to predict the next token based on the previous ones.

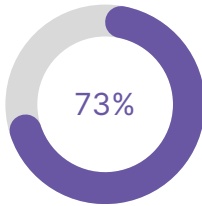
This approach presents challenges in ensuring the quality of the output, especially in more complex tasks, raising concerns.

However, by employing appropriate techniques, it is possible to significantly reduce the occurrences of hallucinations. For example, improving and diversifying the training data, using multiple agents to validate assertions, and applying specific training techniques.

Pinecone

Pinecone, known for its vector database, developed the first hallucination-free model, named Luna. Although Luna does not exhibit hallucinations, this innovation resulted in a drop in performance. Luna often responds with "I don't know" to most questions.

Privacy and Security



of leaders use AI tools that have **not been approved by their company's security team.**

Source: Glean CIO Report

One of the biggest concerns regarding security and privacy is the risk of leaking confidential information through models. This can happen when sensitive data is used to train the model, and it generates texts that include this information, potentially exposing private data to unauthorized users.

There are also concerns about the collection, storage, and management of data used to train LLMs. The integrity and confidentiality of training data are essential, as any compromise can affect not only the model's security but also its effectiveness and reliability.

OpenAI's Privacy Policy and Terms of Use

- ChatGPT Usage: In this case, OpenAI states that it may use your data to improve and train models.
- OpenAI API Usage: The company claims that it does not use any information to train or enhance the model.

REFERENCES

- <https://retool.com/reports/state-of-ai-h1-2024>
- <https://www.cyberhaven.com/blog/4-2-of-workers-have-pasted-company-data-into-chatgpt>
- <https://www.gartner.com/en/newsroom/press-releases/2024-05-16-gartner-identifies-the-top-five-strategic-technology-trends-in-software-engineering-for-2024>
- <https://www.glean.com/resources/guides/software-eng-future-trends>
- <https://leaddev.com/engineering-leadership-report-2024>
- <https://research.google/pubs/resolving-code-review-comments-with-machine-learning/>
- <https://research.google/pubs/ai-powered-patching-the-future-of-automated-vulnerability-fixes/>



AUTOMATE REPETITIVE TASKS AND FREE UP TO 8 HOURS A WEEK FOR YOUR TEAM

Kody is your **AI-powered agent for the entire engineering workflow**—from management to execution. It handles code reviews, test creation, release notes, and other crucial tasks, so your team can focus on what matters.

TRY IT NOW



We have been using Kody in various scenarios, and the results have been extremely positive.

The tool gives us clear visibility into metrics, supports leaders with actionable recommendations, and provides daily insights for management. This reduces the operational workload, allowing us to focus on more significant initiatives.

Leonardo Maia

