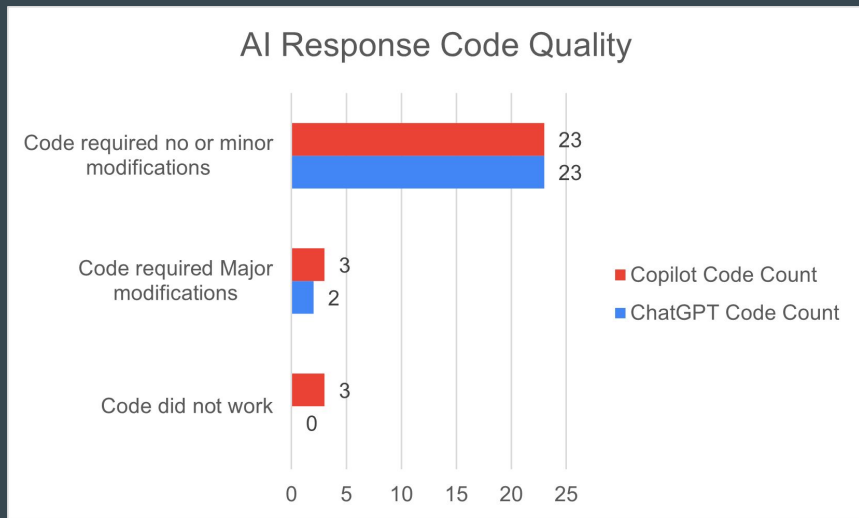


# Analyzing the Effectiveness of AI in Software Development Results

...

By: Emmy Voita

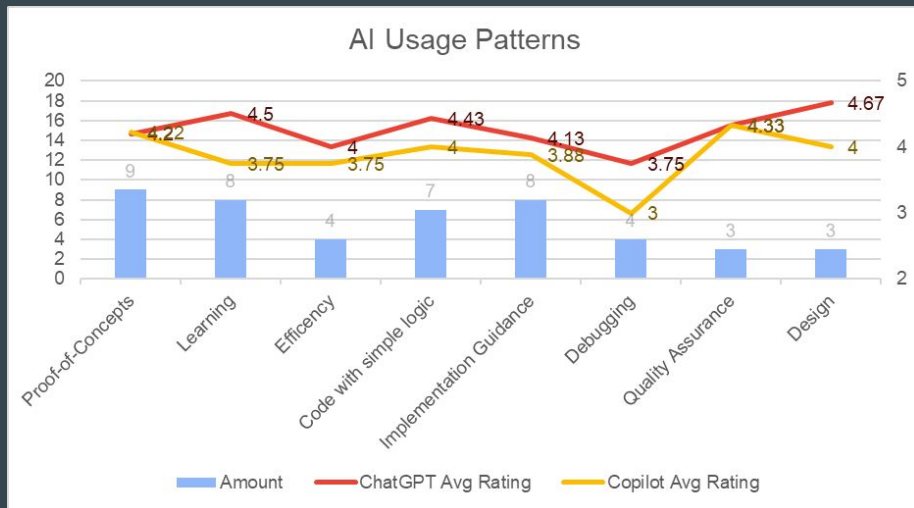
# AI Code Quality



- The code quality between Copilot and ChatGPT was very similar, despite ChatGPT not having full knowledge of the workspace, which supports the null hypothesis.
- However, ChatGPT slightly outperformed Copilot, as it did not have any instances of giving up on code that completely did not work.
  - This reliability contributed to a better overall experience with ChatGPT.

# AI Usage Patterns

Each AI chat was sorted by its usage type, and the average response rating was calculated for each usage type for both AIs.



## Usage Trends:

- Primarily used for implementation guidance, learning, and proof-of-concepts
- Rarely used for design and quality assurance

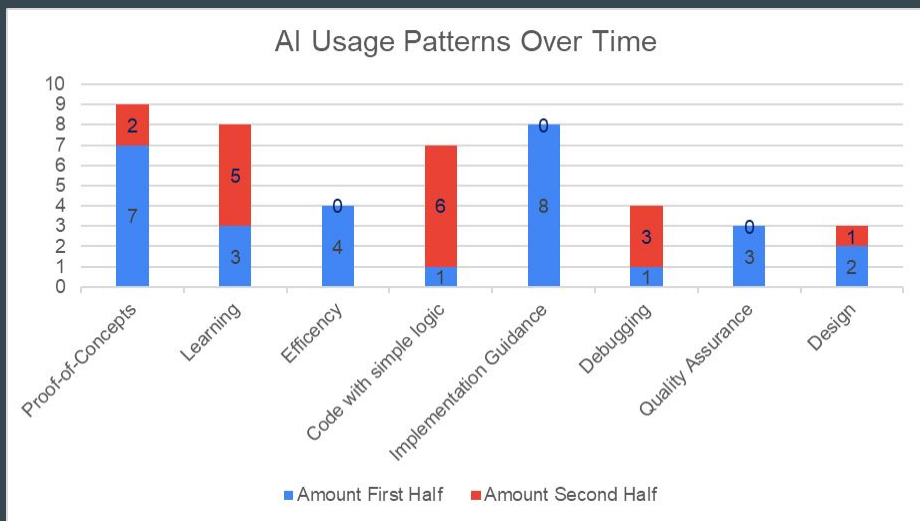
## Performance Highlights:

- Debugging challenged both AI tools
- ChatGPT strengths: Learning, simple code generation, design
- Copilot strengths: Quality assurance and proof-of-concepts

## Comparative Performance:

- ChatGPT performed significantly better than Copilot for Design, Learning, and Debugging topics.
- ChatGPT outperformed Copilot in nearly every category.

# AI Usage Over Time



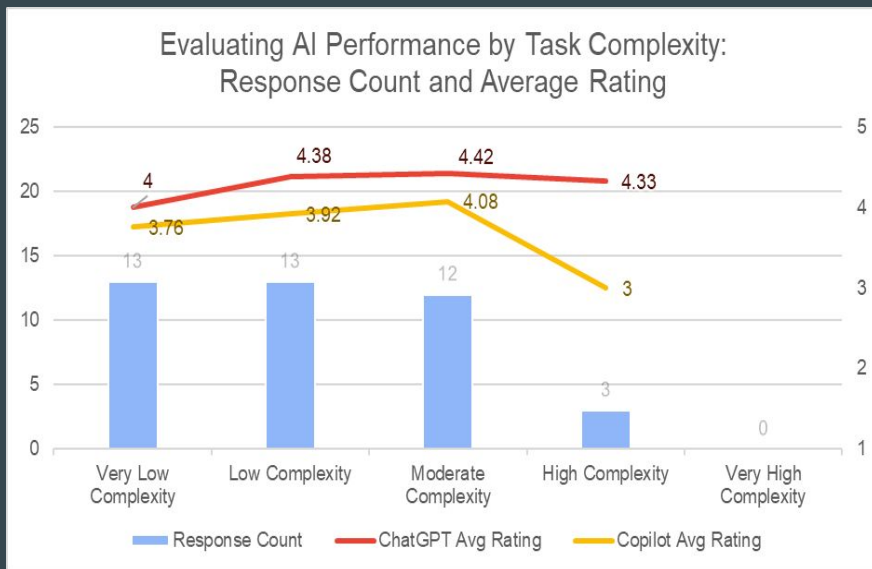
## First Half:

- 59% of total chats
- AI was primarily used for code with simple logic, learning, and implementation guidance in the first half of the project development

## Second Half:

- 41% of total chats
- usage moved more towards learning and code with simple logic in the later half

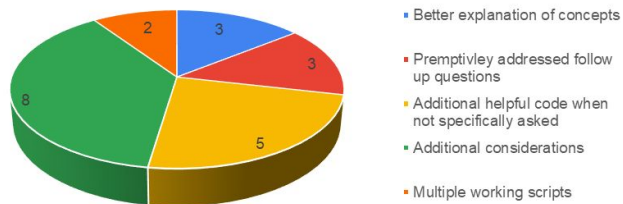
# Evaluating AI Performance by Task Complexity



- ChatGPT outperformed Copilot across all levels of task complexity, demonstrating notably higher performance, particularly for “high complexity” tasks (44.3% increase)
- 92.6% of chats discussed topics of moderate complexity or less, and there were no instances of AI being used for very high complexity tasks.
  - Not many “very high complexity” tasks due to the nature of the project.

# AI Desired Response Characteristics

ChatGPT: Distribution of 5-star Response Characteristics by Count



ChatGPT scored significantly higher 5-star responses compared to Copilot (19.5% to 51.2%)

Most frequent high-valued characteristics:

- ChatGPT:
  - additional considerations
  - helpful code when it wasn't specifically requested
- Copilot:
  - offering multiple working scripts
  - providing great explanations for concepts

Copilot: Distribution of 5-star Response Characteristics by Count

