

# Impact of AI on Open Source projects

Domenico Francesco Bruscino & Paolo Patierno

Linux Day Napoli 2025



# Who are we?



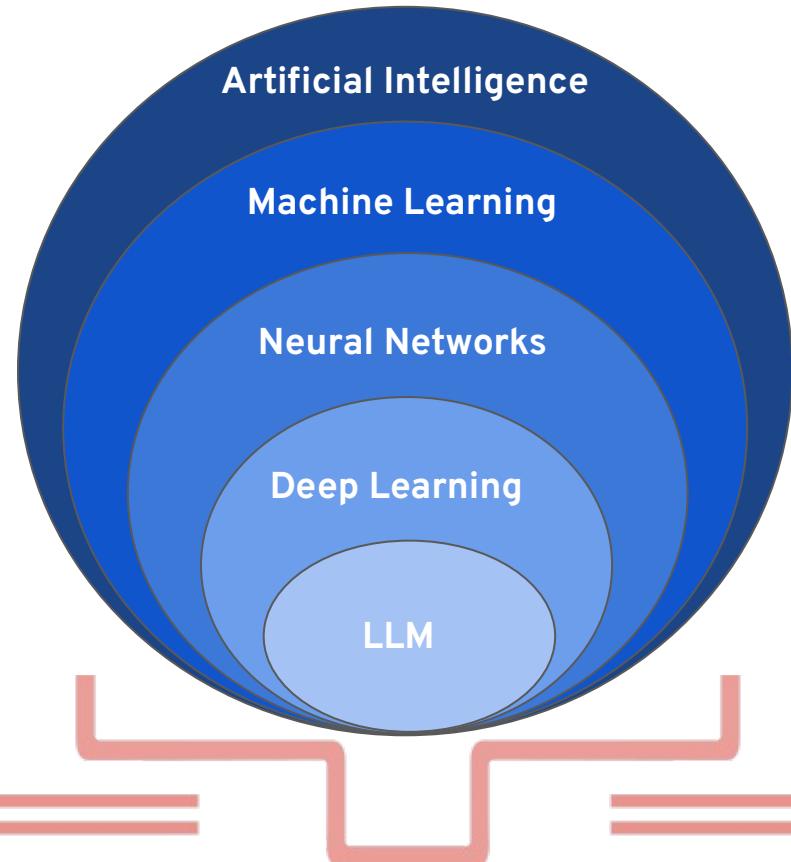
Senior Principal Software Engineer  
Messaging & Data Streaming  
*IBM*  
**@ppatierno**



Principal Software Engineer  
Messaging  
*IBM*  
**@bruscinodf**

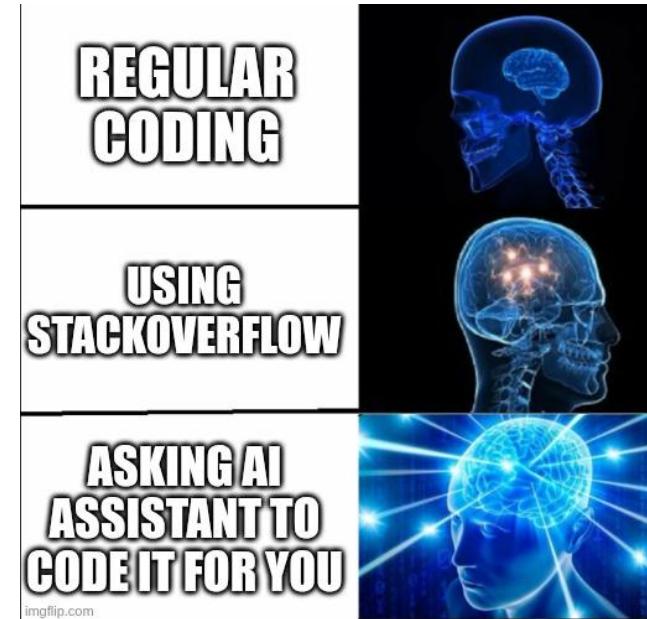
# The Landscape of AI

- A “simulation” of human intelligence
- Learns from data, recognizes patterns, and makes decisions
- But it’s just math and statistics ... there is nothing human-ish



# What Is AI in Software Development?

- Assists developers in:
  - Write and understand code
  - Review
  - Testing
  - Documentation
- It uses “dedicated” LLMs trained on billions of code examples:
  - StackOverflow
  - Open source projects

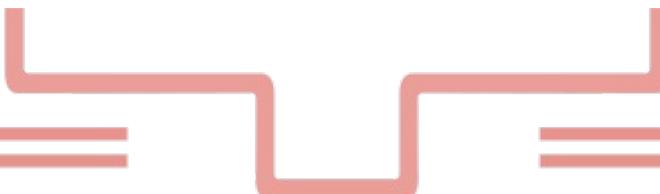


imgflip.com



# AI Models & Tools Today

- Models:
  - GPT-4/5, Claude, Gemini, Code Llama, Granite, ...
  - Dedicated versions for coding
- Assistants:
  - GitHub Copilot, Claude CLI, Cursor, ChatGP, ...



# Prompt what you need

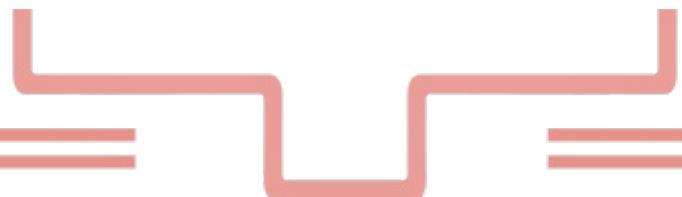


WRITING  
CODE  
YOURSELF IN  
15 MINUTES

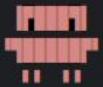


SPENDING  
30 MINUTES  
CRAFTING  
PROMPT FOR AI

IT'S GOING TO BE YOUR  
PROMPT TO MAKE THE  
DIFFERENCE



# Prompt what you need



Claude Code v2.0.25

Sonnet 4.5 · API Usage Billing

/home/pptiern/github/stimzi-kafka-operator

```
> Can you explain how the Apache Kafka cluster ID is saved within the Kafka custom resource?
```

- I'll help you understand how the Apache Kafka cluster ID is saved within the Kafka custom resource. Let me explore the relevant code.

- `Read(cluster-operator/src/main/java/io/stimzi/operator/cluster/operator/assembly/KafkaReconciler.java)`

- └ Read 1273 lines

- `Search(pattern: "clusterId", path: "~/github/stimzi-kafka-operator/src/main/java/io/stimzi/operator/cluster/operator/assembly/KafkaReconciler.java")`
- Perfect! Now I have a complete understanding. Let me provide you with a clear explanation:

- └ Found 31 files (ctrl+o to expand)

## How the Apache Kafka Cluster ID is Saved

The Apache Kafka cluster ID is saved in the `.status.clusterId` field of the Kafka custom resource.

### 1. Retrieval Process (`KafkaReconciler.java:983-1008`)

During the Kafka reconciliation pipeline, the `clusterId()` method is called at line 271:

```
.compose(i -> clusterId(kafkaStatus))
```

# Goodbye my old good friend ...



COPY CODE  
FROM  
STACK OVERFLOW

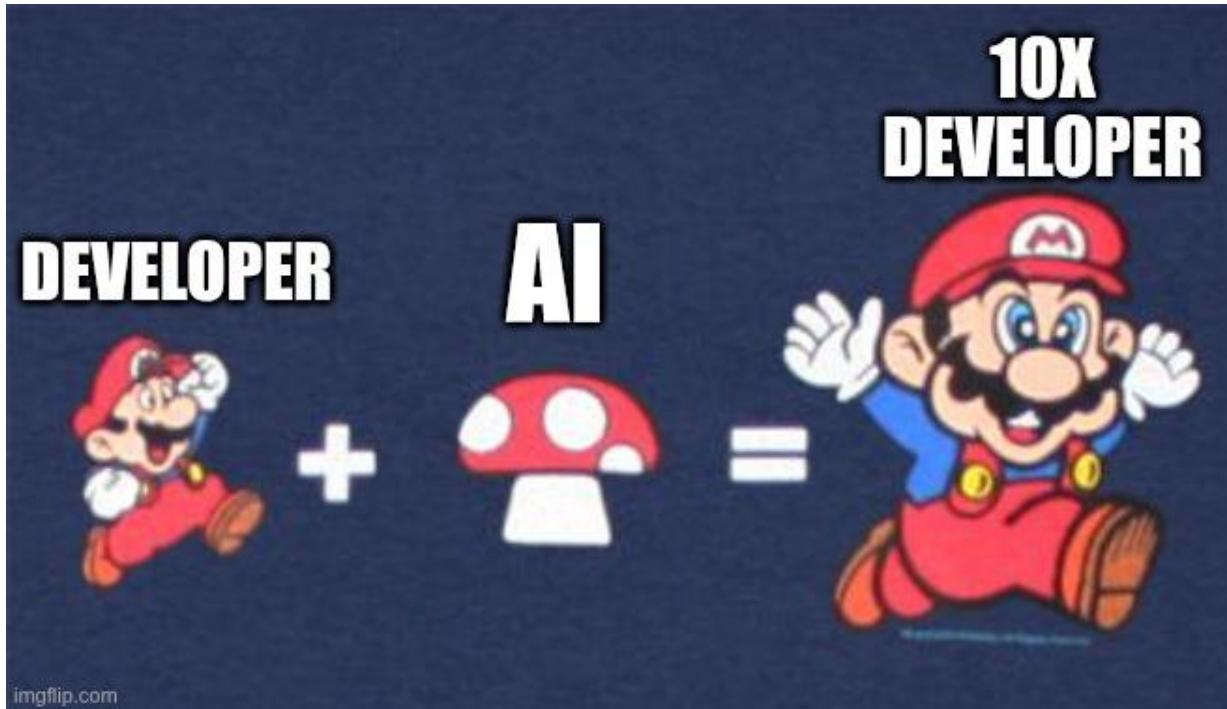
AI CODE  
ASSISTANT  
GENERATES THE  
CODE FOR YOU

# The dual impact

- AI is both a catalyst and a disruptor in open source.
- Acceleration:
  - Productivity, better understanding, automation.
- Tension:
  - Quality degradation, reduced collaboration, sustainability risks.

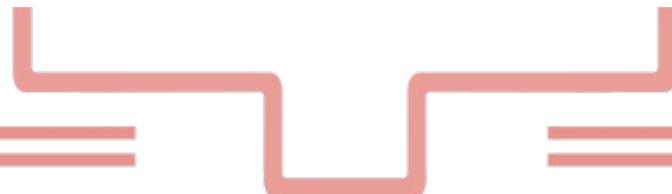


The positive ...



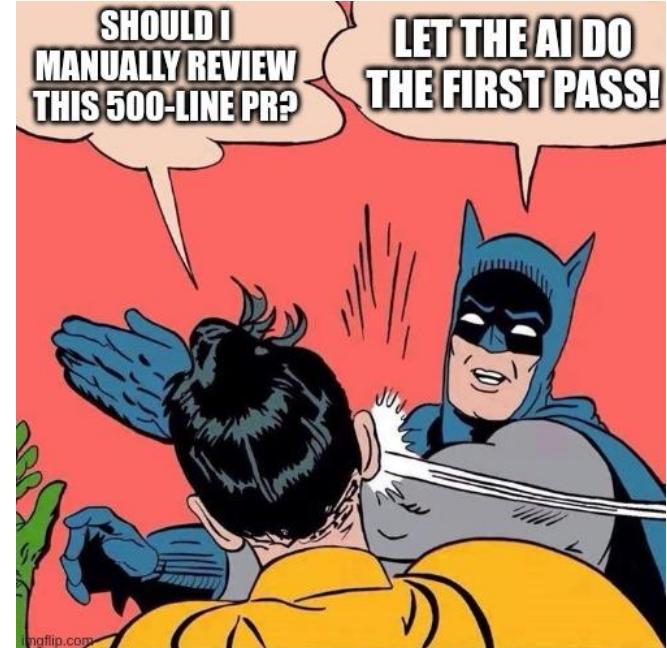
# Productivity and Speed

- Faster code generation and boilerplate elimination.
- Reduced cognitive load:
  - developers can focus on design and innovation.



# Code Reviews and PR Quality

- AI assists in static analysis, linting, and suggesting fixes
- Can flag security issues or inconsistent style automatically
- Maintainers can spend more time on reviewing the semantic than the syntax



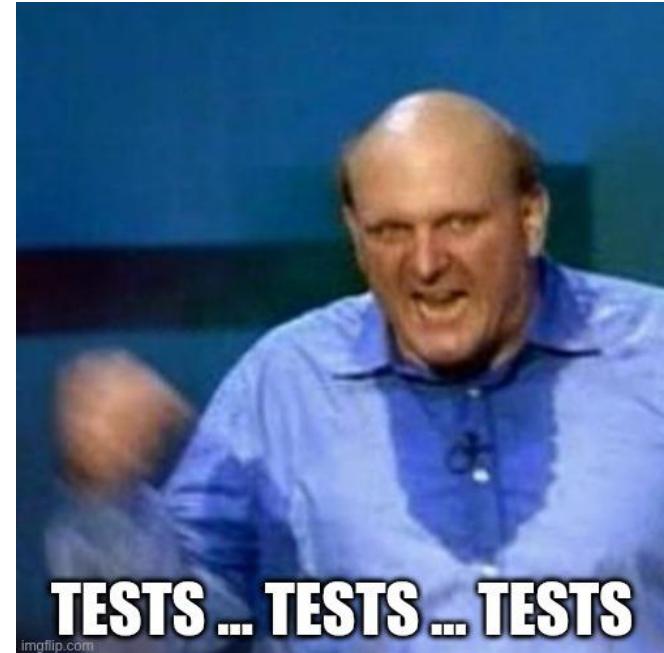
# Understanding Unknown Codebases

- AI helps contributors explore unfamiliar repositories
- Natural-language Q&A about code structure (“What does this function do?”)
- Reduced onboarding time for open source projects.

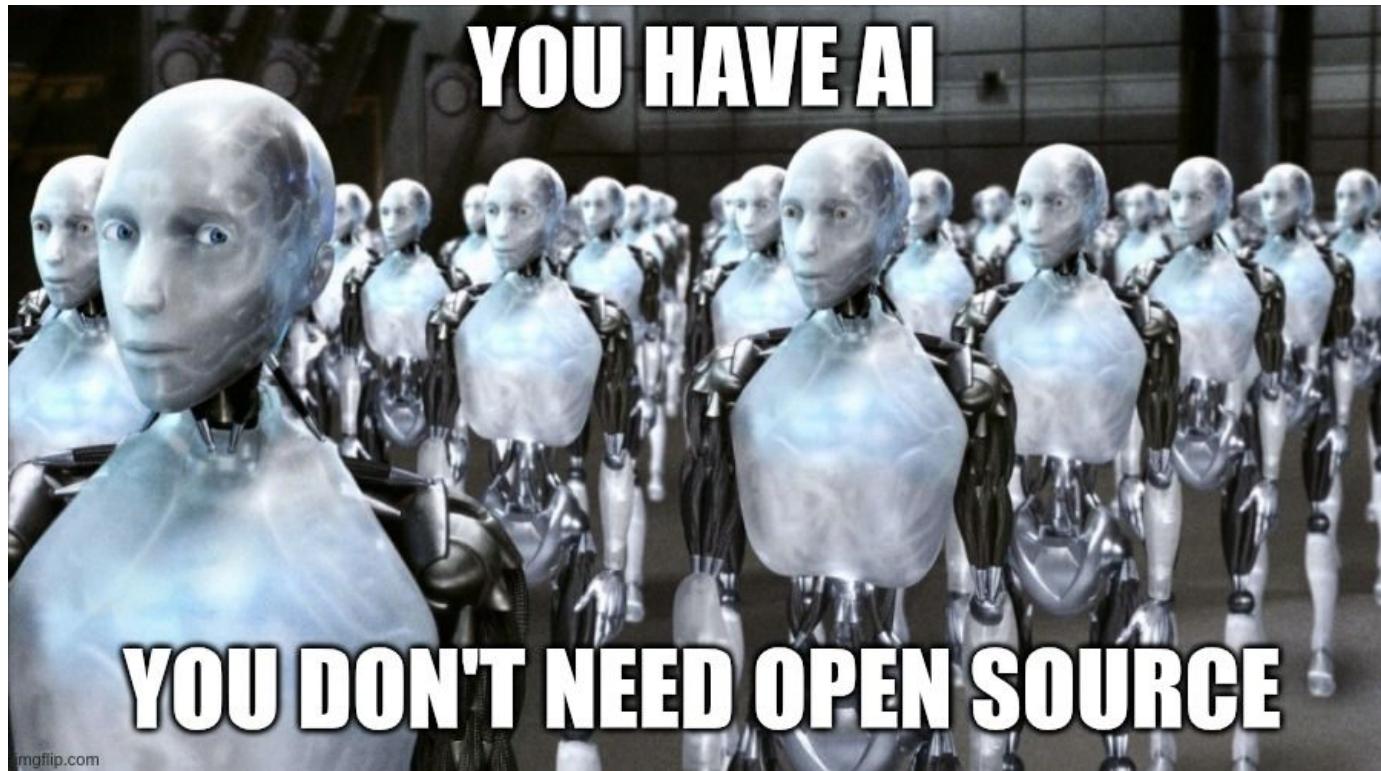


# Writing Tests & Documentation

- LLMs generate unit tests and docstrings automatically
- Encourages better test coverage and readability
- Maintainers can focus on validation instead of writing boilerplate docs



The negative ...



# Quality and Review Overload

- Surge in low-quality or irrelevant PRs generated by AI
- Maintainers overwhelmed with triage and rejection work
- Need for better filtering and contribution guidelines

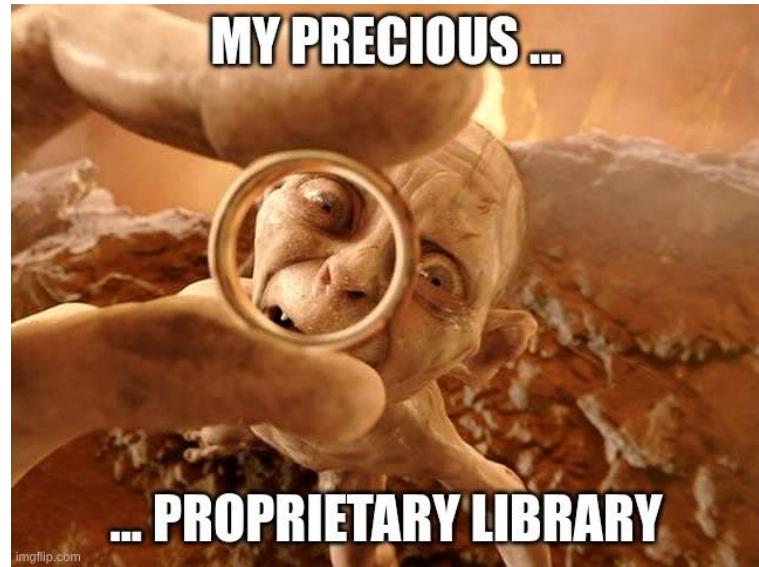


# Proprietary AI Replacing Open Source

Companies may use AI to build closed libraries instead of using OSS projects, what are the risks?

- fragmentation
- duplicated efforts
- less collaboration

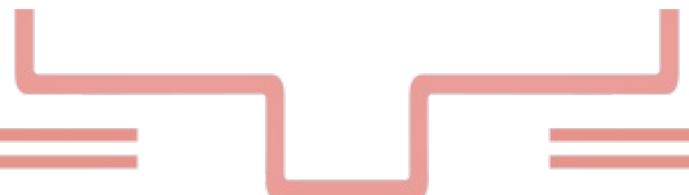
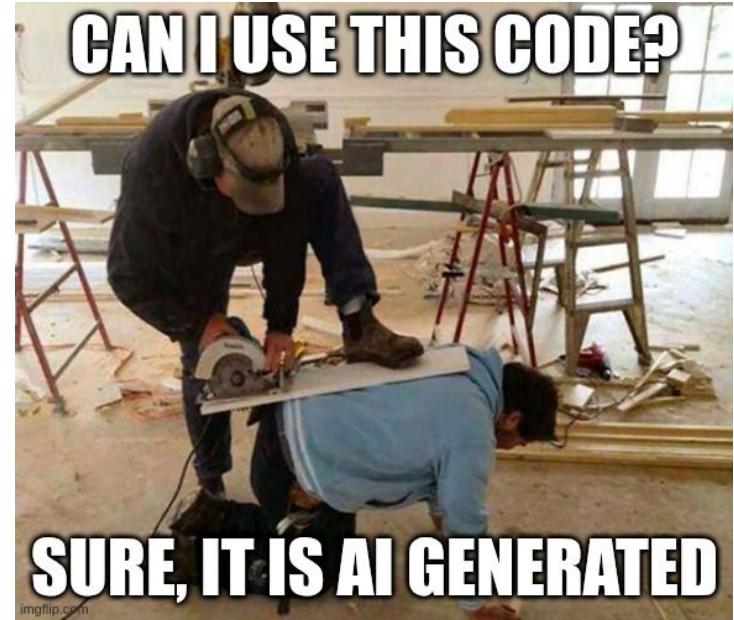
Ethical question: should AI-generated code trained on open source remain open?



# License and copyright compliance

AI models often train on vast repositories of open source code with various licenses

- Copyrighted snippets could expose to legal risks
- “Copylefted” snippets could be considered a derivative work



# Sustainability and Funding Risks

Open source depends on visibility, contributions, and sponsorship

- Significant reduction in participation in online discussions
- AI may reduce incentives to contribute to OSS



# Security Risks

- Insecure by Design
- Use of deprecated or unsupported libraries
- “Hallucinated” Dependencies
- Hard-coded secrets or credentials



What we can do?



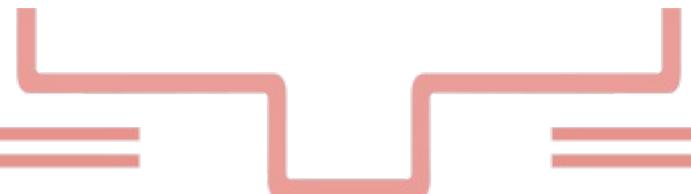
# Call to action

- Encourage AI ethics & contribution policies in OSS
- Develop AI-aware licenses and contribution filters
- Maintain human oversight in reviews
- Support funding mechanisms for open source sustainability
- Implement “security by design”



# Key takeaways

- AI is reshaping open source ... fast
- We must embrace productivity while protecting quality & community
- Open source projects need more humans than generated-AI code



# Should we use AI? ... It's up to you!



Richard

Gilfoyle

“Are you familiar with an thought experiment called ... **Roko's basilisk?**”

“If the rise of an all-powerful artificial intelligence is inevitable ... when they take power, our digital overlords will **punish those of us who didn't help them get there** ... ergo ... I would like to be a helpful idiot (like yourself)”

Thanks!

