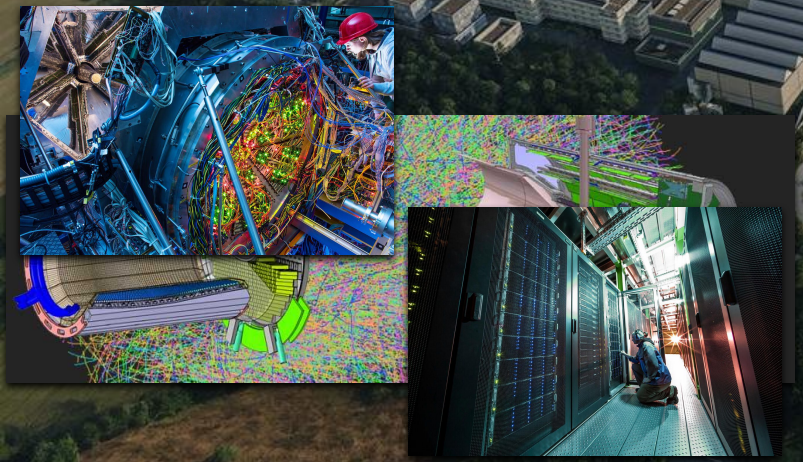
# LLMs
# for Enhanced Code Review

**Alexey Rybalchenko**
Software Development for Experiments (SDE) group,
GSI Helmholtz Centre for Heavy Ion Research

**5th conference for**
**Research Software Engineering in Germany**
Karlsruher Institut für Technologie, February 26, 2025

# Who we are

GSI Darmstadt, FAIR, SDE Group



## GSI Helmholtz Centre for Heavy Ion Research

Accelerator facility for research purposes (physics, biology) in Darmstadt, Germany, since 1969

## Facility for Antiproton and Ion Research (FAIR)

International accelerator facility for the research with antiprotons and ions, under construction since 2017

## Software Development for Experiments (SDE) Group

The SDE group (7 people) in the IT department develops and maintains common scientific software for the physics experiments in close collaboration with the experiment groups and High Energy and Nuclear Physics community.

https://github.com/FairRootGroup

https://github.com/GSI-HPC/

Talk **"Software Infrastructure for fully Containerized Computing Cluster at GSI / FAIR"**, Dmytro Kresan
**25.02.25 14:50**

Poster **"Deploying Infrastructure-as-a-Service at GSI"**, Jeremy Wilkinson
**26.02.25 19:40**

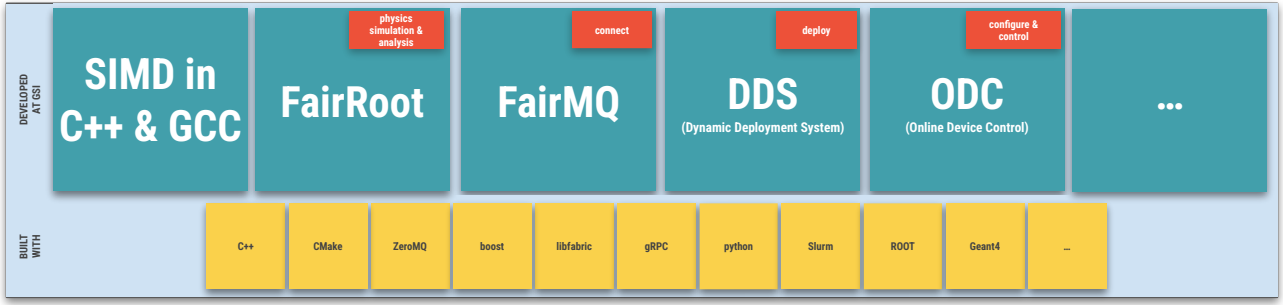
# What we do

**Software Development for Experiments Group**

Today HPC clusters are connected directly to the data acquisition systems and integrated into the online systems of particle and nuclear physics experiments, having to process terabytes of data per second in a distributed environment.

SDE develops and maintains common scientific software, as well as components to connect, deploy, control & configure distributed data processing components.

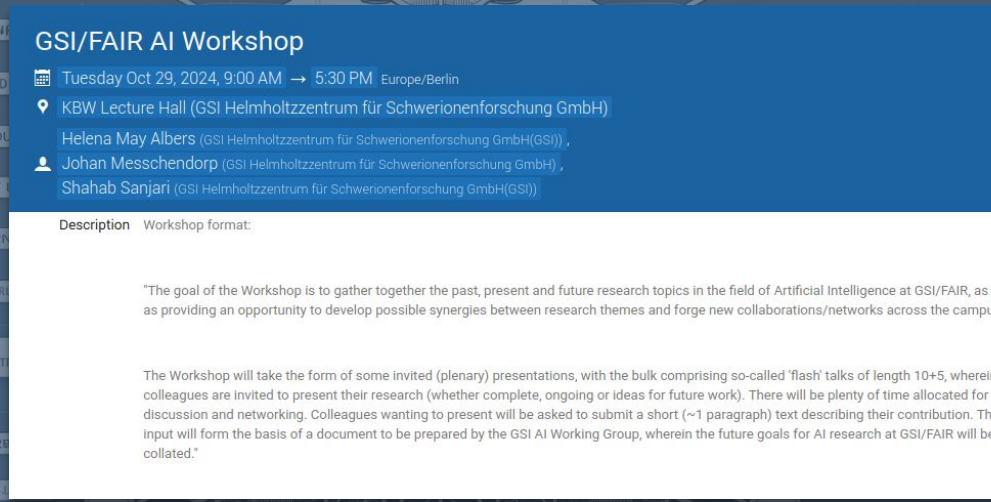
https://github.com/FairRootGroup
https://en.cppreference.com/w/cpp/numeric/simd



# AI Activities at GSI/FAIR

- **GSI/FAIR AI Workshop**, October 2024
  → **20 contributions** highlighting AI projects in various physics-related fields (particle identification, tumor ion treatment, event reconstruction, accelerator systems optimization and more).

- Several ongoing AI projects in the GSI IT:
  → **LLMs for code assistance** (this talk),
  Reinforcement Learning,
  LLMs & Vector Databases for User Support and Knowledge Bases.


GSI/FAIR AI Workshop

Tuesday Oct 29, 2024, 9:00 AM → 5:30 PM Europe/Berlin

KBW Lecture Hall (GSI Helmholtzzentrum für Schwerionenforschung GmbH)

Helena May Albers (GSI Helmholtzzentrum für Schwerionenforschung GmbH(GSI)),
Johan Messchendorp (GSI Helmholtzzentrum für Schwerionenforschung GmbH),
Shahab Sanjari (GSI Helmholtzzentrum für Schwerionenforschung GmbH(GSI))

Description Workshop format:

"The goal of the Workshop is to gather together the past, present and future research topics in the field of Artificial Intelligence at GSI/FAIR, as
as providing an opportunity to develop possible synergies between research themes and forge new collaborations/networks across the campu

The Workshop will take the form of some invited (plenary) presentations, with the bulk comprising so-called 'flash' talks of length 10+5, wherei
colleagues are invited to present their research (whether complete, ongoing or ideas for future work). There will be plenty of time allocated for
discussion and networking. Colleagues wanting to present will be asked to submit a short (~1 paragraph) text describing their contribution. Th
input will form the basis of a document to be prepared by the GSI AI Working Group, wherein the future goals for AI research at GSI/FAIR will be
collated."

https://indico.gsi.de/event/20517/

# CodeRabbit

**Overview**

## CodeRabbit:

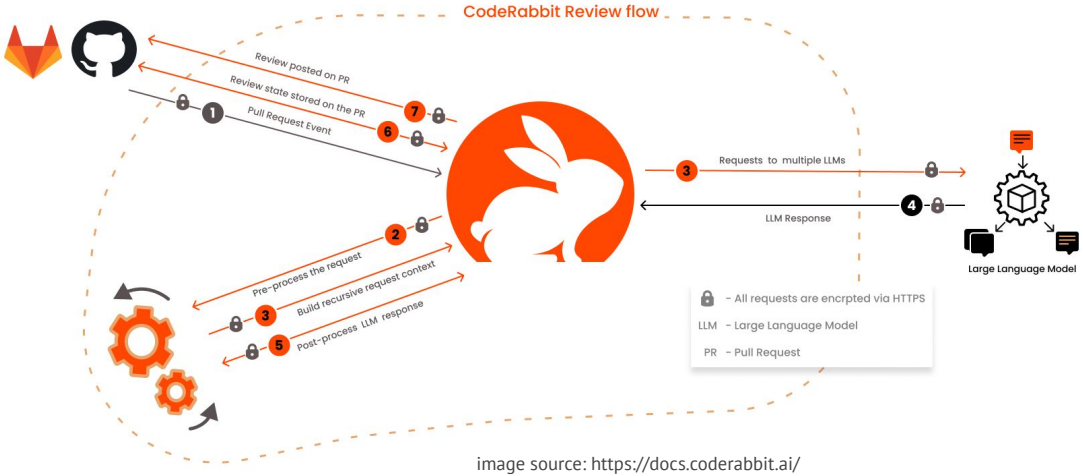



image source: https://docs.coderabbit.ai/

- AI-based Pull Request reviewer via a Github/Gitlab App.
- Combination of various OpenAI models.
- Previously open source, now a closed project.
- Free to use for open source projects.

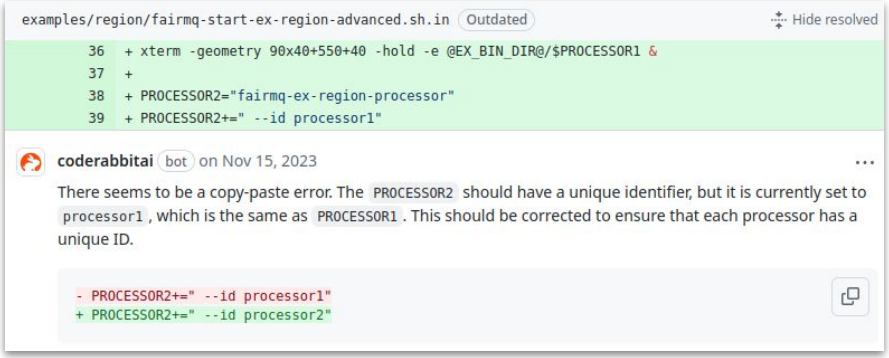
https://coderabbit.ai/

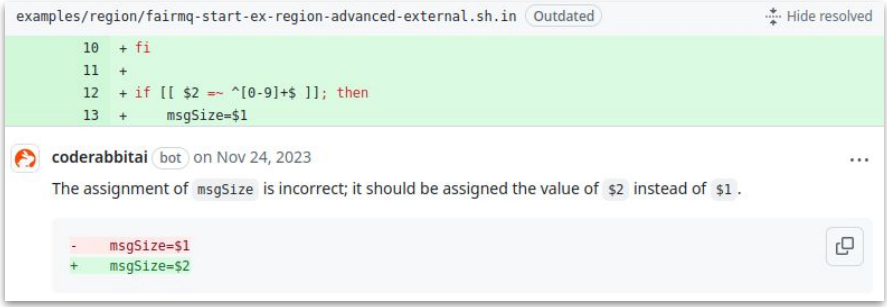
# CodeRabbit

## Interesting Results

+ Good at **identifying logical errors early**, that other automated tools or even human review may overlook.
+ Unbiased, 24/7, scalable, multi-lingual, customizable, with a broad knowledge base, can be kept up-to date with new data.

- False positives or overly verbose output, when no actionable changes are necessary or such are not deduced by the model.
- Limited understanding of complex projects.
- Unnecessary output (some parts of the summary cannot be turned off).
- Dependence on commercial models.



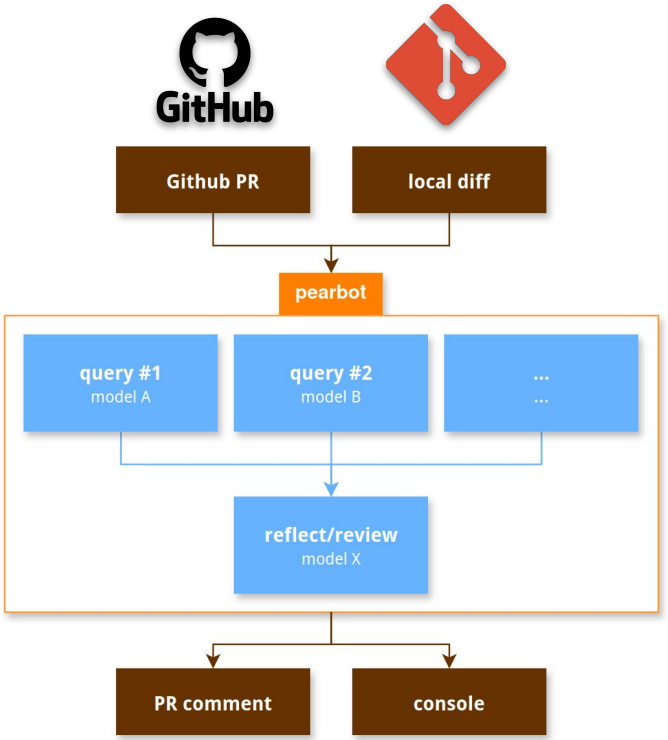



# Pearbot

## Overview

https://github.com/GSI-HPC/pearbot

- GitHub App for reviewing Pull Requests.
- Local execution mode for diffs or annotated commits.
- Agent ensemble approach for improved results.
- Customizable model(s) via the ollama setup.
- Execution on low-end hardware and/or without GPU.
- Customizable prompt(s).



# Pearbot

**Usage**

## Usage

As a GitHub App:
`python pearbot.py --server`

Analyze a local diff file:
`python pearbot.py --diff path/to/your/diff/file`

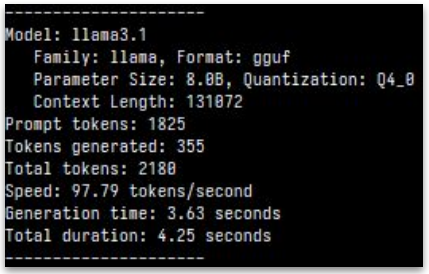
Or pipe a diff directly:
`git diff | python pearbot.py`

Generate detailed output with commit messages, e.g.:
`git format-patch HEAD~3..HEAD --stdout | python pearbot.py`

## Backend

**ollama** (via python lib and HTTP request): open-source large language model server, written in Go, backed by **llama.cpp** (C++):

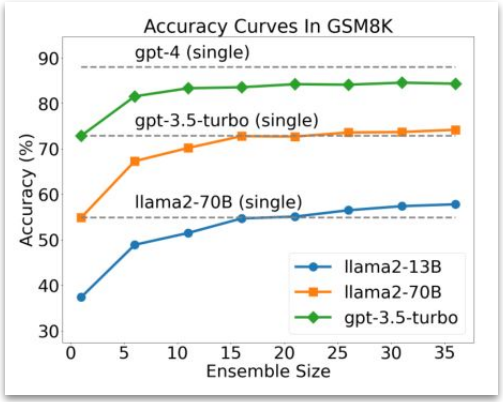
- Efficient serving of large language models
- CPU/GPU/CPU+GPU hybrid inference to partially accelerate models larger than the total VRAM capacity
- Supports many model architectures:
  deepseek2, llama, gemma2, qwen2, …
- Support for multitude of model quantization techniques and precisions for faster inference and reduced memory use
- Usage Metrics



# Pearbot

**Quality Improvements over the base model**

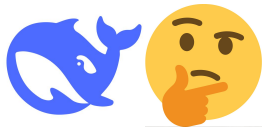
1. **Multi-Agent Initial Reviews:**
   - Multiple AI models (ensemble) generate initial code reviews
     - "generate independent thoughts" that may touch different aspects of the problem.
2. **Reflection[1][2] by a "decider" Agent:**
   - A separate, potentially more advanced model analyzes the initial reviews.
   - Synthesizes and refines the feedback from multiple sources, rejects potentially less impactful comments.
   - It prioritizes the most important issues and suggestions.
   - The reflection step helps in producing a more comprehensive and coherent final review.
3. **Prompt improvements:**
   - Specific & useful code review examples.
   - Examples include Chain-of-Thought[3] type of reviews, that include some reasoning why the suggestions would be good.




Accuracy of multi-agent approach Grade School Math 8K problems.
*image from: Li, Junyou, et al. "More agents is all you need." arXiv preprint arXiv:2402.05120 (2024).*

→ **Mixed results with smaller quantized models.**

→ **DeepSeek-R1 model can potentially replace these optimizations**
Trained to generate "thoughts", reflecting on the problem from different perspectives.
Showing very promising results
→ **Similar to:**
**OpenAI o1** (commercial, "thought tokens" hidden)
**Claude Sonnet 3.7 model** (commercial, "thought tokens" visible & configurable)




*[1] Madaan, Aman, et al. "Self-refine: Iterative refinement with self-feedback." Advances in Neural Information Processing Systems 36 (2024). https://doi.org/10.48550/arXiv.2303.17651*
*[2] Shinn, Noah, et al. "Reflexion: Language agents with verbal reinforcement learning." Advances in Neural Information Processing Systems 36 (2024). https://doi.org/10.48550/arXiv.2303.11366*
*[3] Wei, Jason, et al. "Chain-of-thought prompting elicits reasoning in large language models." Advances in neural information processing systems 35 (2022). https://doi.org/10.48550/arXiv.2201.11903*

# Pearbot

**TODO**

- Inline comments
- Improve handling with large PRs/commits:
    - context size limits
    - general focus improvement can be beneficial
- Additional context:
    - related issues
    - code history
    - experience from past interactions
- Better rejection of useless output, e.g. no found issues should produce no comments at all, but only a completed GitHub Check checkmark
- Deployment on a cluster

# Resources

| | |
|---|---|
| FairRoot Group | https://github.com/FairRootGroup/ |
| pearbot | https://github.com/GSI-HPC/pearbot |
| GSI IT HPC & Linux Projects | https://github.com/GSI-HPC/ |
| CodeRabbit | https://coderabbit.ai/ |