Local Agents with Python and Ollama

June 2025

Agenda

- Why should I run LLMs locally?
- A practical example scenario
- Coding session
- Conclusion
- Q&A

Why should I run LLMs locally?

- Control over **data management**
- Control over **service dependencies**
- **No API costs**, good for experimentation
- (Truly) **deterministic** executions
- It’s a nice way to **learn**

aitho

The test subject

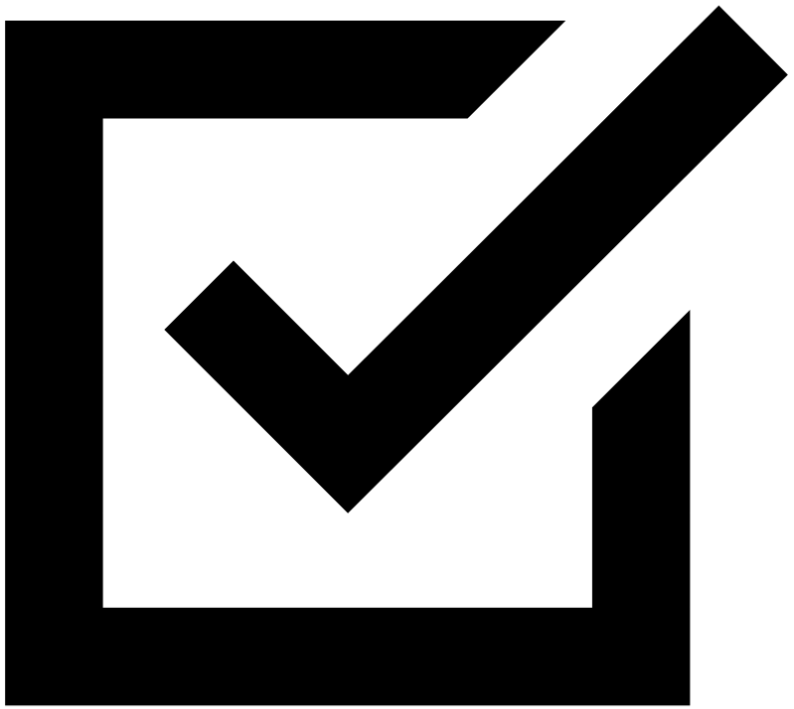
- Arch user
- Laptop with AMD stack and no dedicated GPU
- Not really into Python coding

The test scenario

- All players but one (the spy) are given a card with the name of a secret location.
- Players ask each other questions to identify the spy.
- The spy tries to blend in and guess the location.

aitho

What did we do?



- Defined the protocol of a system based on human interaction
- Implemented local LLMs able to act in the system
- Added human-in-the-loop

aitho

What did we learn?

- Running local instances of open source models
- Writing code that interacts with such models
- Even a laptop with Arch can run local assistants

aitho

What did we earn?

A virtually free-of-charge LLM system that can be used for:

- Testing (e.g. finding flaws in a modified version of Spyfall)
- Training (leveraging human-in-the-loop)
- Having fun (with real friends)

aitho

aitho
welcome home developers