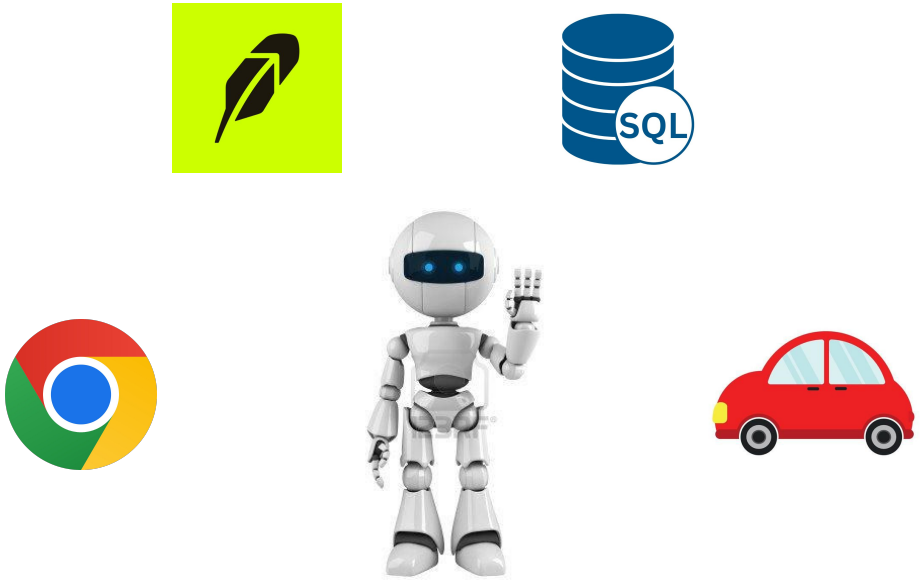
AI Agents with Langchain

What are AI Agents

- AI Agents are tool-assisted LLMs/AI Models.
- The “tools” they have access to are various APIs, databases, and/or hardware.
- Agentic AI is the process of multiple AI agents working together to achieve a goal.



What is Langchain (and Ollama)

- Langchain is an open source Python API for building and running AI Agents.

- Ollama is an open source software and platform for accessing local llms.

Goals of the workshop

We Will:

- Build and demo stand alone agents that have varying roles.
- Review best practices.
- Build an agentic stock analyzer pipeline!!!

We will not:

- Demo any RAG powered agents. (we will review what rag is though)
- Build any agents powered by closed-source llms.

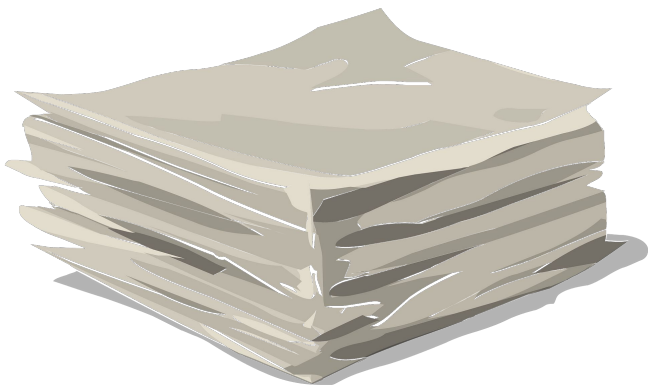
Right now:

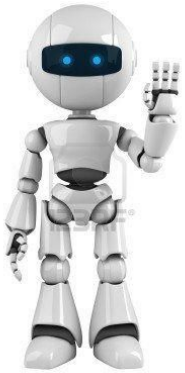
- Download Python, Langchain, and Ollama (along with an llm of your choice, I’m using Gemma2). Also download the python scripts via git.

- This will take a while if you don’t already have Ollama and a chosen llm ready, but don’t feel bad since the scripts are all written.

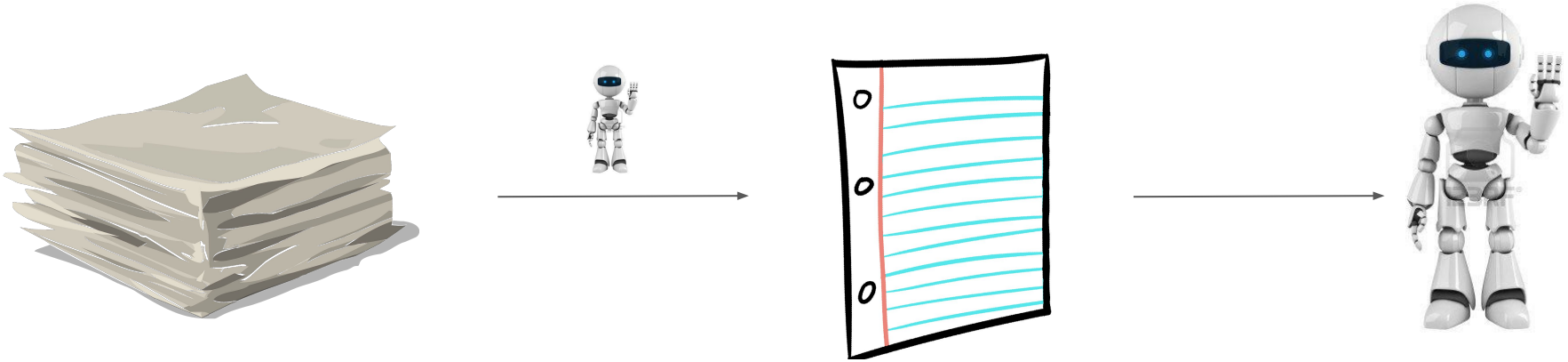
Code Inspection #1 - Basic Stateless Agent

Code Inspection #2 - Stateful Agent (Buffer)

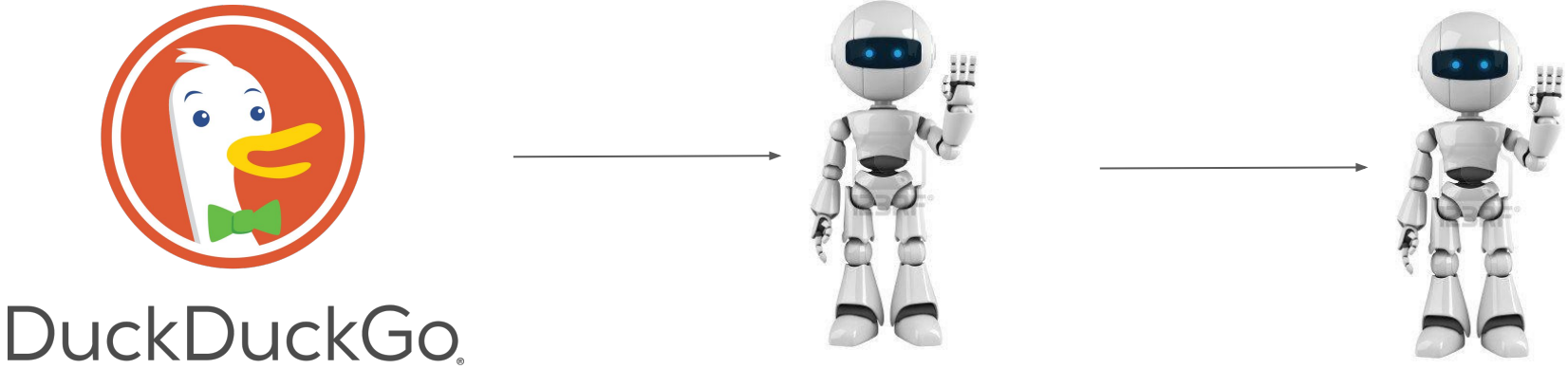




Code Inspection #3 - Stateful Agent (Summary)



Code Inspection #4 - Agentic Stock Analyzer



Brief coverage of RAG… (i dont know how to visualize ts)

- As we have seen, LLMs alone do not hold memory, so we'd have to stuff prior messages into the prompt window to "simulate" memory (of the user only).

- But what if we want to allow our agents to be smarter and up-to date about real world events not shared directly by the user? -> why RAG is used.

- Various documents and external data is "RAG'd" so the agent can look it up within a database in relation to a new prompt before returning its response. (why ChatGPT remembers conversations very well)

- RAG can also be used on data such as files and existing conversations to make smarter responses to the user.

Issues with RAG

- RAG is not foolproof! The main disadvantage is Data Storage.

- Some challenges with mass data storage:
  - The method to store data and incurred costs (cloud, hardware, etc.)
  - Maintaining data quality and scalability
  - Data collection compliance (does it follow the law of most jurisdictions?)
  - How efficient is data retrieval?

- Personal prediction of mine is that with the rise AI, we will definitely see more Data Science/Engineer related roles, possibly even more than software roles.