Workshop

“Where engineers herd LLMs and wrangle prompts”

Og Maciel - Chief Llama Whisperer

Welcome to Ollama Drama

Workshop Goals:

- Run and customize LLMs locally
- Use CLI and Python to chat with models
- Have Fun!?!?
- Submit solutions via GitHub PRs

OLLAMA
DRAMA

Why Ollama and Why Local?

- Local serving of LLMs made easy
- Works on macOS, Windows (WSL), Linux
- Benefits: Privacy, Speed, No API limits

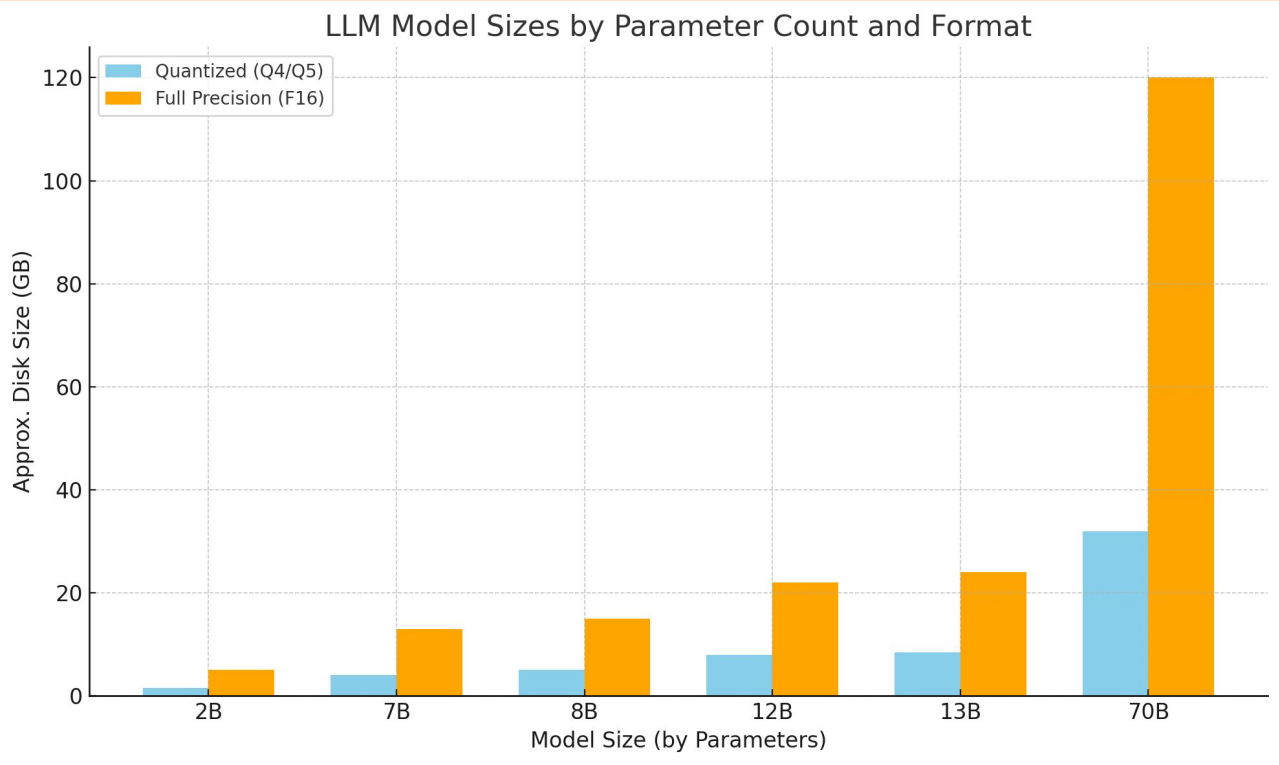
OLLAMA
DRAMA

Installing Ollama

https://ollama.com/download

OLLAMA
DRAMA

Which LLM?



OLLAMA DRAMA

Qwen3:0.6B

523**MB**

Llama3.1:8B

4.9**GB**

OLLAMA DRAMA

https://ollama.com/library

Basic Ollama CLI

```
$ ollama pull qwen3:0.6b
$ ollama list
$ ollama run qwen3:0.6b "What is AI?"
```

OLLAMA DRAMA

More Ollama CLI

```
$ ollama ps
$ ollama show qwen3:0.6b
$ ollama cp qwen3:0.6b ollama-drama
$ ollama stop qwen3:0.6b
$ ollama rm
```

OLLAMA DRAMA

Useful Patterns

```
$ ollama run qwen3:0.6b "Summarize this file: $(cat README.md)"
```

OLLAMA DRAMA

REST API: Generate a Response

```
curl
http://localhost:11434/api/generate
-d '{
  "model": "qwen3:0.6b",
  "prompt":"Why is the sky blue?",
  "stream": false
}' | jq -r '.response'
```

OLLAMA
DRAMA

REST API: Chat with a model

```
curl http://localhost:11434/api/chat -d '{
  "model": "qwen3:0.6b",
  "stream": false,
  "messages": [
    { "role": "user", "content": "why is the
sky blue?" }
  ]
}'  | jq -r '.message.content'
```

OLLAMA DRAMA

Ollama Drama Code

https://github.com/omaciel/ollama-drama

chatbot.py Example

```
$ python3 -m venv .venv
$ source ./venv/bin/activate
$ pip install -r requirements.txt
$ ollama pull qwen3:0.6b
$ python chatbot.py
```

OLLAMA DRAMA

Challenge 1 – Customize the LLM's System Prompt

- Edit the code in ***problems/problem1/student_code.py***
- Test
  - `$ pytest -k test_problem1.py`

OLLAMA DRAMA

Submit Your PR

- Push your solution to GitHub
- Send me a Pull Request for certification

Challenge 2 – Create Your Own Customized LLM

- Edit the system prompt:
  - My name is…
  - Favorite food…
  - Favorite color…
- Build (locally) your LLM:
  - `$ ollama create <username>/mario-model -f problems/problem2/Modelfile`
- Edit the code in ***problems/problem2/student_code.py***
- Test
  - `$ pytest -k test_problem2.py`

OLLAMA DRAMA

Submit Your PR

- Push your solution to GitHub
- Send me a Pull Request for certification

BONUS – Publish Your Customized LLM

- Create an account on Ollama.com
- Add your ~/.ollama/id_ed25519.pub to https://ollama.com/settings/keys
- Build (locally) the LLM:
  - `$ ollama create <username>/mario-model -f problems/problem2/Modelfile`
- Publish your LLM:
  - `$ ollama push <username>/mario-model`
- Edit the code in ***tests/test_bonus.py***
- Test
  - `$ pytest -k test_bonus.py`

OLLAMA DRAMA

Review and Takeaways

- Run LLMs locally via Ollama
- Chat via CLI and Python
- Customize system prompts
- Add SMART goal for your manager
- Submit your PRs and get your Certificate of Completion

OLLAMA DRAMA

S.M.A.R.T. Goal

“Complete the “Ollama Drama” hands-on workshop on May 9th and earn a Certificate of Completion by learning how to run, customize, and interact with local LLMs using CLI and Python. This will contribute toward my Development Plan goal of building AI-related technical expertise.”

OLLAMA
DRAMA

Certificate of Completion

Ollama Drama Workshop

You've herded the llamas, now enjoy the glory.

OLLAMA DRAMA

Participant Name

Og Maciel

Workshop Instructor

Date: May 9th, 2025

Serial #: **000000**

What’s Next?

COMING SOON

OLLAMA
DRAMA

Thank You

- Explore embeddings and RAG
  - Upcoming “**From RAGs to Riches**” course?
- Try more models: ollama.com/library

OLLAMA DRAMA