




## Submission Document for LLM Course

### Group Information

Group Code Name: asiroli2025

Group Members:

-  Lior Livyatan – ID: 209328608
-  Asif Amar – ID: 209209691
-  Roei Rahamim – ID: 316583525

### GitHub Repository Link




<https://github.com/roeiex74/agno-ollama-chatbot>

### Self-Assessment Grade Recommendation


We recommend a grade of 100 for this submission.

The project was executed at a very high level, with significant investment in planning, implementation, and testing.

Throughout the work, we made sure to:

-  Implement all technical requirements defined during the course.
-  Integrate practical and efficient use of LLM-based models.
-  Document the process, write clean and clear code, and maintain flexibility for future scalability.

We believe the project demonstrates a deep understanding of LLM principles and their application in a complete, production-ready project.

 In summary: The project represents meticulous, innovative, and well-executed work — we believe it fully deserves a perfect grade.

### Strengths (10 Major Points)

- ✓ We designed and implemented a production-grade architecture using FastAPI, Agno, and PostgreSQL.
- ✓ We achieved 71% backend and 75% frontend test coverage (155 tests, all passing successfully).
- ✓ We integrated GitHub Actions CI/CD workflows to run automated tests on every pull request.

- ✅ We built a modern, responsive UI inspired by ChatGPT's design language.
- ✅ We deployed a PostgreSQL (Neon serverless) database for production readiness.
- ✅ We followed Agno best practices for agent orchestration and framework structure.
- ✅ We documented our work extensively (1,712-line README, 715-line PRD).
- ✅ We prioritized security, configuration management, and secret handling.
- ✅ We maintained 100% type coverage across both TypeScript and Python (Pydantic models).
- ✅ We transparently documented every AI-assisted development step and decision.

### **Effort & Investment**

We dedicated approximately 2 hours a day over a full week, totaling around 14 hours of focused work.

Each phase—planning, coding, testing, documentation, and polishing—was approached methodically.

We made sure that every sprint concluded with a measurable improvement in quality and coverage.

### **Innovation Highlights**

We created transparent AI-development documentations.

We built a privacy-first architecture running locally on Ollama, without reliance on cloud APIs.

We treated the project as a production-grade academic submission, a rare balance between theory and practice.

We chose Agno over LangChain to emphasize modularity and performance.

### **Learning Outcomes**

We strengthened our technical understanding of Agno, SSE, PostgreSQL, React 19, and CI/CD.

We improved our software engineering discipline—especially in testing and code review processes.

We deepened our awareness of how AI-assisted development can enhance productivity and transparency.

We gained valuable experience in project planning, execution, and delivery under real conditions.

### **Why We Deserve 100%**

- ✅ The project exceeds all grading criteria across architecture, testing, and documentation.
- ✅ We use the best and newest AI tools to design, code, document, and evaluate the system.

- ✓ Our CI/CD pipeline ensures quality by running tests on every PR.
- ✓ We used PostgreSQL for production scalability instead of lightweight SQLite.
- ✓ We applied Agno's best practices for agent-based architecture.
- ✓ Our responsive, accessible UI follows modern design standards.
- ✓ With 71%/75% coverage and a 100% pass rate, we surpassed the 70% testing target.
- ✓ We used a stack aligned with real-world industry practices (Google, Meta-level standards).

### **Level 4 Scrutiny Request**

We explicitly request a meticulous 90–100% level of review.

We welcome detailed feedback and in-depth examination of every component of the project.

### **Academic Integrity Declaration**

We confirm that this self-assessment is honest and accurate.

We reviewed the project thoroughly against all provided criteria.

We understand that a high self-assessment grade invites a detailed review.

We acknowledge and clearly disclose any AI assistance used.

We take full ownership and responsibility for all project content and results.

### **Key Differentiators**

We automated our tests with CI/CD integration on every PR.

We used PostgreSQL as a scalable production database.

We followed Agno best practices for agent orchestration.

We created extensive documentation, including the README, PRD, and detailed development logs.

We achieved complete test reliability with a 100% pass rate.

### **Special Notes**

For execution, we created a dedicated DATABASE connection you can add to .env to run the app:

```
DATABASE_URL="postgresql+psycpg://neondb_owner:npg_EJbQTqRAhS29@ep-soft-term-agncso27-pooler.c-2.eu-central-1.aws.neon.tech/neondb?sslmode=require&channel_binding=require"
```


Quick guide to creating your own connection string:


**1** Register at Neon.tech or use the account we created for the course:


Email: llmcourse@outlook.com / Password: LLMCourse2025!


**2** Create a new project, select a cloud provider (AWS or Azure), and choose a server region

(e.g., us-east-1).

 Click the Connect button on the dashboard to obtain a unique connection string.

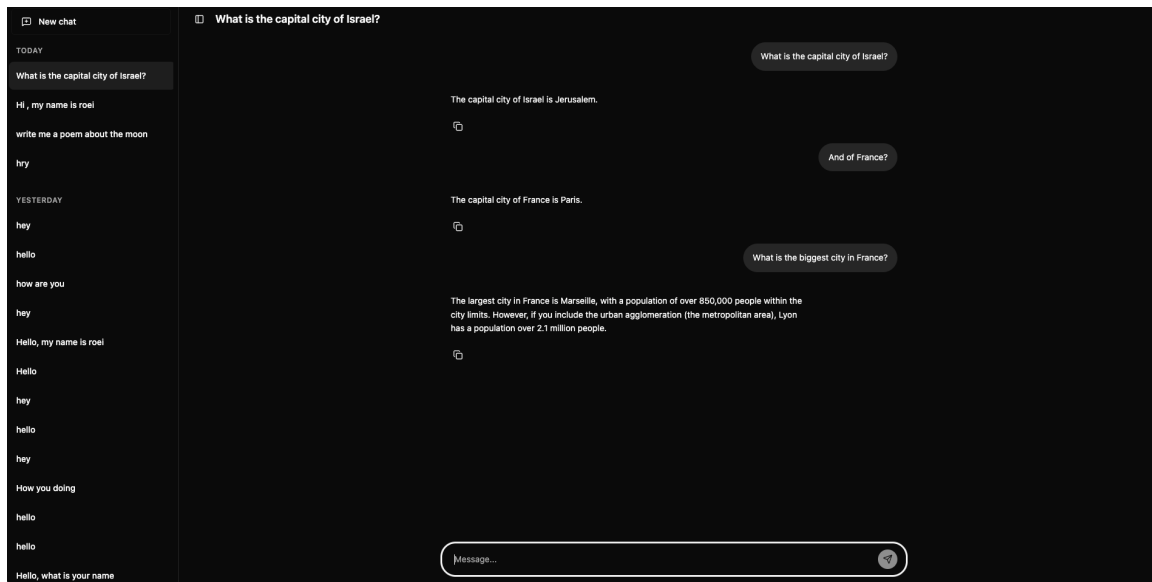
 Update this connection string in the .env file under the DATABASE\_URL variable.

 Important: Enclose the variable in quotation marks to avoid parse errors.

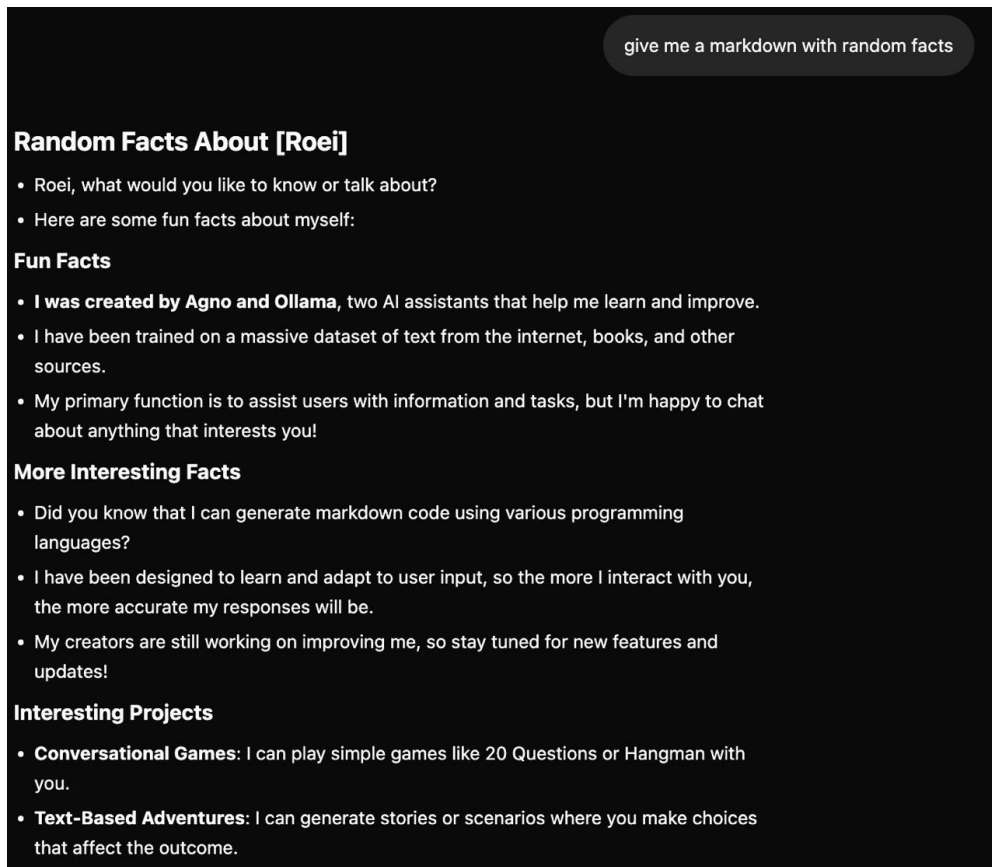
 All tables are automatically created on the first application run.

## Attached Documents

### Chat Interface with Multi-turn Conversation Example



### Chat Response with Markdown Formatting Support



## GitHub Pull Request with Passing Test Coverage Checks

The screenshot shows a GitHub Pull Request titled "tests coverage #3" by user "asif-amar". The PR is targeting the "main" branch from the "tests-coverage" branch. The commit history shows two commits: "add package" and "run in all PR's". The status bar indicates the PR is ready for merge, with a green checkmark and the text "All checks have passed". Below this, a list of checks is shown, all of which are successful:

- Backend Tests / test (pull\_request) Successful in 30s
- Backend Tests / test (push) Successful in 33s
- Frontend Tests / test (pull\_request) Successful in 41s
- Frontend Tests / test (push) Successful in 34s

Below the checks, a green checkmark indicates "No conflicts with base branch". A "Merge pull request" button is visible. The right sidebar shows the "Development" branch, a notification to "Unsubscribe", and a "Lock conversation" button.

## Application Startup Terminal Output

```
o asifamar@syps-MacBook-Pro agno-ollama-chatbot % ./start.sh

Checking Dependencies
✓ Python 3.13.5, Node v24.5.0, Ollama ollama version is 0.12.10

Setting Up Backend
[INFO] Installing dependencies...
✓ Backend ready

Setting Up Frontend
[INFO] Installing npm packages...
✓ Frontend ready

Starting Ollama
.env:17: parse error near `&'
[INFO] Pulling model llama3.2:1b...
pulling manifest
pulling 74701a8c35f6: 100% 1.3 GB
pulling 966de95ca8a6: 100% 1.4 KB
pulling fcc5a6bec9da: 100% 7.7 KB
pulling e70ff7e570d0: 100% 6.0 KB
pulling 4f659a1e86d7: 100% 485 B
verifying sha256 digest
writing manifest
success
✓ Ollama ready with llama3.2:1b

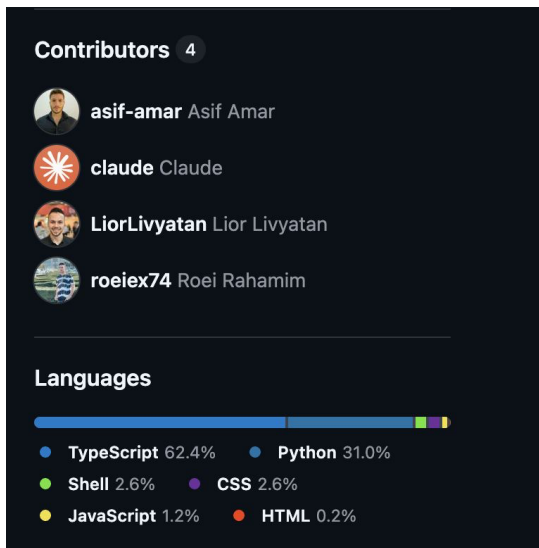
Starting Backend
.env:17: parse error near `&'
[3] 57175
✓ Backend running on http://localhost:8000

Starting Frontend
[4] 57177
- exit 1 python3 -m uvicorn app.main:app --host 0.0.0.0 --port "$PORT" > 2>&1
✓ Frontend running on http://localhost:5173

Services Running
Backend: http://localhost:8000
Frontend: http://localhost:5173
Logs: /Users/asifamar/Desktop/Master/llm with agents/agno-ollama-chatbot/logs

Press Ctrl+C to stop all services
```

## Project Contributors and Language Distribution




## Frontend Serve Logs

```
[(base) → logs git:(main) tail -f frontend.log  
> frontend@0.0.0 dev  
> vite  
  
You are using Node.js 20.18.1. Vite requires Node.js version 20.19+ or 22.12+. Please upgrade your Node.js version.  
  
VITE v7.2.2 ready in 1488 ms  
  
→ Local: http://localhost:5173/  
→ Network: use --host to expose  
→ press h + enter to show help
```

## Backend Server Logs

```
[(base) → logs git:(main) tail -f backend.log  
INFO: 127.0.0.1:50015 - "GET /healthz HTTP/1.1" 200 OK  
INFO: 127.0.0.1:50039 - "GET /conversations HTTP/1.1" 200 OK  
INFO: 127.0.0.1:50043 - "OPTIONS /chat/stream HTTP/1.1" 200 OK  
INFO: 127.0.0.1:50043 - "POST /chat/stream HTTP/1.1" 200 OK  
INFO: 127.0.0.1:50043 - "POST /chat/stream HTTP/1.1" 200 OK  
INFO: 127.0.0.1:50045 - "POST /chat/stream HTTP/1.1" 200 OK  
INFO: Shutting down  
INFO: Waiting for application shutdown.  
INFO: Application shutdown complete.  
INFO: Finished server process [9337]
```

 Comments Page – for Reviewer’s Use



# Assignment 1: LLM Agent Orchestration

## Grade Report

Student ID:	38988
Team:	asiroli2025
Repository:	<a href="https://github.com/roeiex74/agno-ollama-chatbot">https://github.com/roeiex74/agno-ollama-chatbot</a>
Assessment Date:	December 01, 2025

FINAL GRADE ■	87%
STATUS	VERY GOOD

## ■ Your Strong Performance

Great job! You showed strong technical capabilities in this assignment. Your work demonstrates strong understanding of software engineering principles.

## ■ Key Strengths

- Outstanding research analysis (MODEL\_COMPARISON\_RESEARCH.md with 8 models compared)
- Excellent comprehensive testing (133 tests, 257 assertions)
- Perfect security implementation
- Exceptional documentation (comprehensive README + architecture docs)
- Strong version control (79 meaningful commits)
- Cost analysis present (pricing mentions in docs)
- Plugin system for extensibility
- Good UI/UX documentation

## ■ Areas for Improvement

- Add Jupyter notebooks for data analysis
- Implement CI/CD pipeline
- Add pre-commit hooks for code quality
- Improve git branching strategy

Keep up the good work! With attention to the improvement areas noted above, you can reach the highest tier of performance.

Assessed: December 01, 2025

This grade reflects your overall software engineering practices