Unleashing the power of LLMs as Embedded AI in Robotics

Miguel Á. González-Santamarta





Outline

- 1. Introduction
- 2. llama_ros
- 3. Demo
- 4. Prompt Engineering
- 5. Use Cases
- 6. Conclusions





Introduction

 Emergence of Large Language Models (LLMs)

 Revolutionizing natural language understanding and generation

















universidad deleón

Introduction

- Dynamic realm of robotics
- Advantages of LLMs in robotics:
 - Natural Language Interaction
 - Language Translation
 - Storytelling
 - Contextual Understanding
 - Reasoning and Planning
 - Adaptive Behavior



universidad ^{de}león



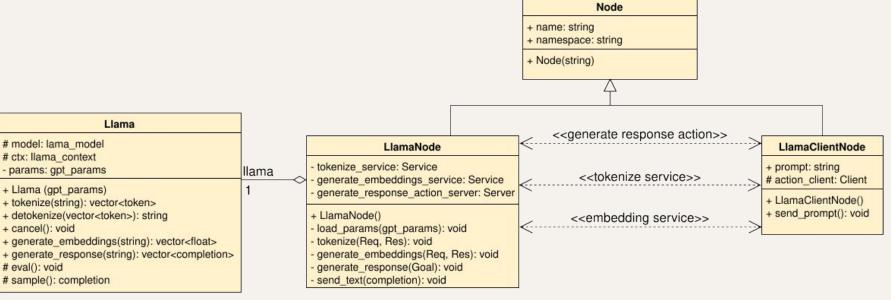
https://github.com/mgonzs13/llama_ros

- Based on Ilama.cpp
- Quantized LLMs
- Fully integrated into ROS 2
- ROS 2 interfaces
- Acceleration with GPU
- Grammars
- Integration with langehain → Prompt Engineering





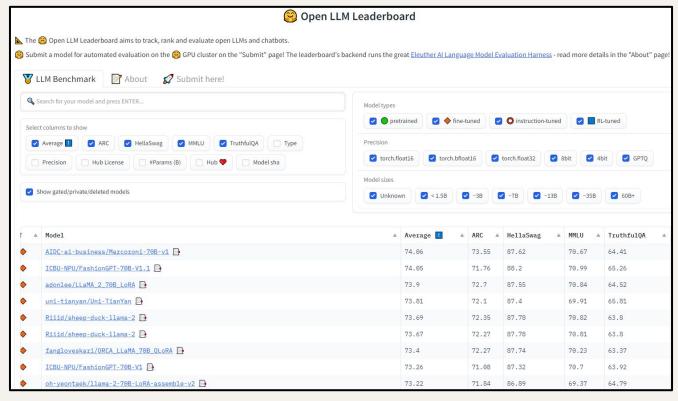






GRUPO DE ROBÓTICA

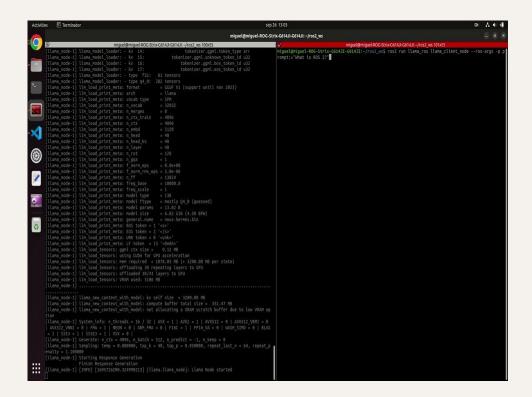








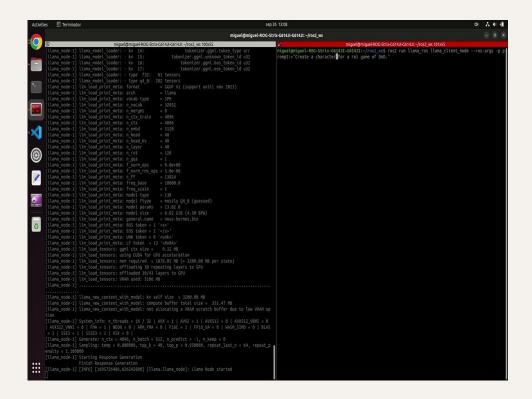
Demo







Demo







Prompt Engineering

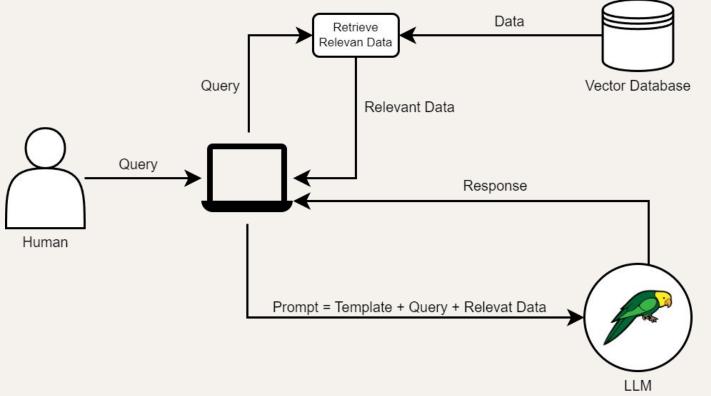
- Zero-Shot Prompting
- Few-Shot Prompting
- Self-Consistency
- Generated Knowledge Prompting
- Retrieval Augmented Generation (RAG)
- Chain-of-Thoughts (CoT)
- Tree-of-Thoughts (ToT)



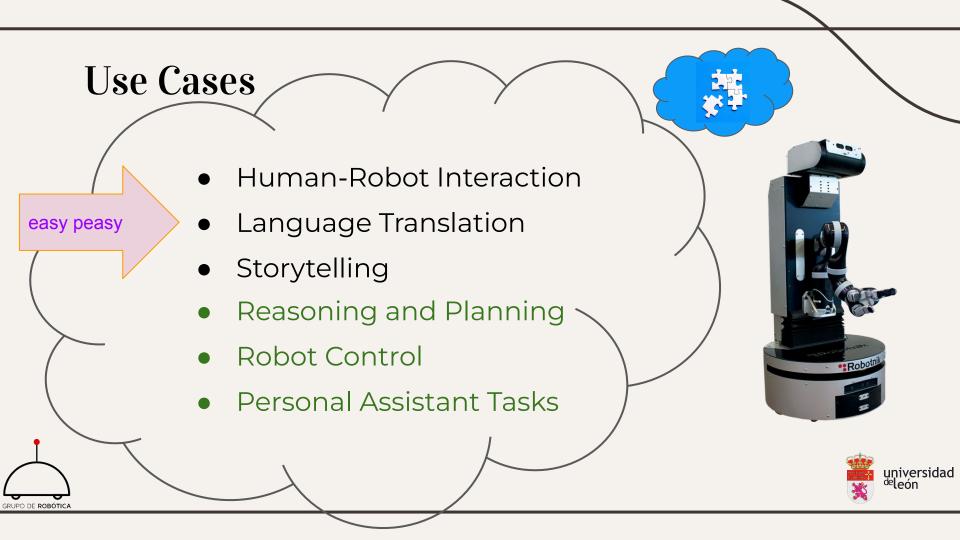


Prompt Engineering (RAG)

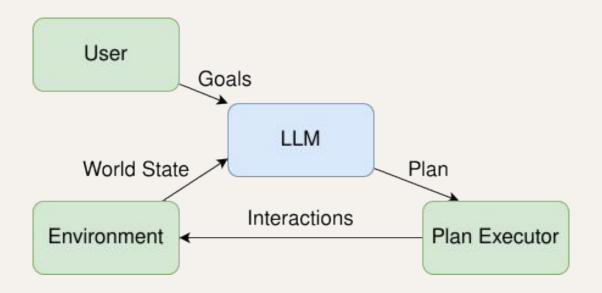
GRUPO DE ROBÓTICA







Use Case: Reasoning and Planning







Use Case: Reasoning

```
[INFO] [1695052408.503526771] [planning_layer_node]: {
 "reasoning": "RB1 is currently at the entrance and needs to greet Vicente who is at the livingroom."
"plan": "Move RB1 to the livingroom, then greet Vicente.",
"actions": [
         "name": "move to",
         "args": {
             "robot name": "RB1",
             "source_waypoint": "entrance",
             "target waypoint": "livingroom"
         "name": "greet_person",
         "args": {
             "robot name": "RB1",
             "person_to_greet": "Vicente",
             "person waypoint": "livingroom"
```





Conclusions

- Integration of LLMs in ROS 2 projects
- LLMs as Embedded AI in Robotics thanks to Quantization
- GPU to accelerate LLMs
- Prompt Engineering in Robotics
- Future work: Speculative Search





Unleashing the power of LLMs as Embedded AI in Robotics

Miguel Á. González-Santamarta



Grant PID2021-126592OB-C21 funded by MCIN/AEI/10.13039/501100011033 and by ERDF A way of making Europe.



