LLM-based Puzzle Solver

Karolina Klimek, Weronika Wronka

1 Project Description

The goal of this project is to create an agent based on a large language model (LLM) that can solve logical and mathematical puzzles using advanced natural language processing strategies. Example tasks include: *"John is older than Mary, Mary is older than Tom. Who is the oldest?"*

The agent's functionality will include:

- Interpreting tasks and breaking them down into logical steps (chain-of-thought prompting),
- Generating and comparing multiple possible solutions,
- Selecting the best solution based on accuracy and consistency.

The project allows for comparing various problem-solving strategies:

- Zero-shot prompting (no examples),
- Few-shot prompting (with examples),
- Chain-of-Thought (CoT) (step-by-step reasoning).

An analysis of the effectiveness of each strategy on a selected set of logic puzzles is also planned.

1.1 Technologies

- **Ollama** – local deployment of LLMs (e.g., Mistral, Mixtral),
- **Python** – prompt management, agent logic, evaluation,
- **(Optional) GUI** – a simple graphical interface for testing the agent from the user perspective.

2 Work Schedule

Stage	**Scope of Work**	**Deadline**
1. Topic analysis and planning	Collecting example puzzles, developing the agent's strategy	22.04 – 28.04
2. Agent prototype (baseline)	Implementing a basic agent version in Python with prompt support	29.04 – 12.05
3. Extension with CoT and few-shot strategies	Testing different prompting approaches, solution comparison logic	13.05 – 19.05
4. Effectiveness evaluation	Running tests on various types of puzzles, result analysis	20.05 – 02.06
5. Documentation and presentation	Final write-up, corrections, preparing a demo	03.06 – 17.06