Abstract

TODO: at the end

Introduction (2/3 pages)

> context, problem, state-of-the-art, specific problem of the thesis, objective/objectives, expected results, structure

context, problem, state-of-the-art, specific problem of the thesis, objective/objectives

context: AI, Generative systems problem: AI system are capable of planning and reaching a goal. Are LLM capable? state-of-the-art: LLM applied to similar problems specific problem of the thesis: objective/objectives: show the effectiveness of an LLM of a typical AI problem (logistics) expected results: a generative system without any reasoning capabilities is able to solve a problem of this type? And to what extent?

Everyone says LLM are incredible for everything. My thesis: let's try to apply them to the delivery problem. A very simple problem, a classic problem for AI, that is composed by:

- a planing problem
- a reasoning problem
- ...

This can be faced in a classic way with PDDL (time intensive), RL. We want to address this with LLMs.

structure (to be done at the end of the thesis)

Chapter 1 - Background

> state-of-the-art, background

Background

- LLMs
  - Focus on Attention mechanism
  - Few shot learners
- Uncertainty in LLMs
  - KnowNo framework
    - Spain?
  - Other approaches
- BDI Agents
- SotA: PDDL
  - pros and cons
- SotA: Reinforcement Learning
  - pros and cons

Chapter 2 - Experiment setting

- Problem definition: This is a generative approach, there is no reasoning, no planning, no pathfinding
- Deliveroo.js
- GPT models
    - why GPT
    - why not LLAMA or other open source/open weight models
    - OpenAI API
- Prompts

Chapter 3

- First approach: parsing server info + helping parameters -> going in the wrong direction
- Second approach: stateless full raw + stateful
- Final agent: new prompt, new data, still bad results

Chapter 4

- Stateless & Stateful, why stateful works better
- Attention visualizer in BERT
- encoding the map (base64) to reduce the number of characters
- Better prompt creation accordingly to papers (give a role, where to put the goal, knowno, few-shot working means the prompt is correct)
- Heatmap creation
- Last chance, find the closest cell

Results discussion

- Stateful: path evaluation
- Stateless: heatmaps
- Stateless: closest cell

Conclusions (2 pages)

thesis objectives, achieved results, limitations, future developments

limitation: token for context limit in stateful