Tool-Augmented Large Language Models

Raphael Schumann
Sommersemester 2024
PS/HS

Department of Computational Linguistics

Course Format

- mostly online (heiCONF), announced if in-person (INF 325 / SR 24)
- English
- 6 LP (Bachelor), 8 LP (Master)

Contact:

- rschuman@cl.uni-heidelberg.de

How to get points?

- Annotated Paper Presentation (no slides)
  - 8LP: 30-35 min
  - 6LP: 20-25 min
  - 10-15 min discussion

- Preparation
  - read papers to be presented (at least one)
  - two questions or comments per session
  - via Email to me the before the session
- Active Participation
  - participate in discussion (ask your prepared questions, comments about presentation…)

How to get points?

- Implementation Project
  - (Re-)Implementation of relevant approach, compare/evaluate/extend existing models
  - technical report (paper) of 4-6 pages.
  - details will be discussed in a later session
  - Deadline: Start of next semester

OR

- Second Paper Presentation (with slides)
  - 8LP: 30-35 min
  - 6LP: 20-25 min
  - 10-15 min discussion

How to get points?

- Grading
    - Preparation: 30%
    - Presentation: 40%
    - Implementation Project / Second Presentation: 30%

General Advice

- look for additional material
    - appendix (often not in the conference pdf)
    - supplementary Material (often on conference website)
    - project/author website (videos, demos, context)
    - github (README, code details)
    - conference talks
    - reviews (openreview.net)
- read cited papers for background information
- read citing papers to learn about drawbacks and impact

Paper Selection

- Send an Email with your Top 3 paper picks
  - and if you want to do a second presentation instead of the implementation project
- Deadline: Sunday, April 28th
  - rschuman@cl.uni-heidelberg.de

Next Week: Reading Group

"TALM: Tool Augmented Language Models", Parisi et al., 2022

- you don't need to read the paper or send questions

Motivation - Tool-Augmented LLMs

- LLMs are bad at:
  - factual lookup
  - up-to-date information
  - math/arithmetic

Who is the current President of the United States?

I don't have real-time data or internet access.

What day of the week is it today?

I do not have the capability to access the current time or date.

What is the result of 3435 * 235 / 9?

This is approximately 89,937.22.   True answer: 89691.66

Motivation - Tool-Augmented LLMs

- External Tool/API calling can help

Out of 1400 participants, 400 (or [Calculator(400 / 1400) → 0.29] 29%) passed the test.

The New England Journal of Medicine is a registered trademark of [QA("Who is the publisher of The New England Journal of Medicine?") → Massachusetts Medical Society] the MMS.

Challenges - Tool-Augmented LLMs

- Tool/API Selection
- Training data for tool calling is rare
- Human annotations are expensive
- LLM dependent of when tool/API call is useful
- Tool usage should work task independent
- Nested Tool calls

See you next week (online)