ITEA Topical roadshow Large Language Models

Practical examples

26 March 2024 | Online

Abdelkrim Boujraf (ALT-F1)

Agenda

1. ALT-F1 & iRARP presentations
2. RAG LLM Architecture
3. Case Study 1 – Query urology best practices document
4. Case Study 2 – Support the Service Desk with ChatBot
5. Case Study 3 – GTIN Barcodes
6. Conclusions and Discussion

ALT-F1 presentation

https://www.alt-f1.be

- Brussels-based software company incorporated in Oct-2010
- We design, build, deploy and support complex software in any industry
- We have worked for clients in:
  - Public Administration, Broadcasting, Telecom, Retail, Aviation, Defence, Automotive, Transportation, Banking, Finance, Insurance, Underwriting, Logistics And Health

iRARP presentation

Intelligent Robotic Assisted Radical Prostatectomy

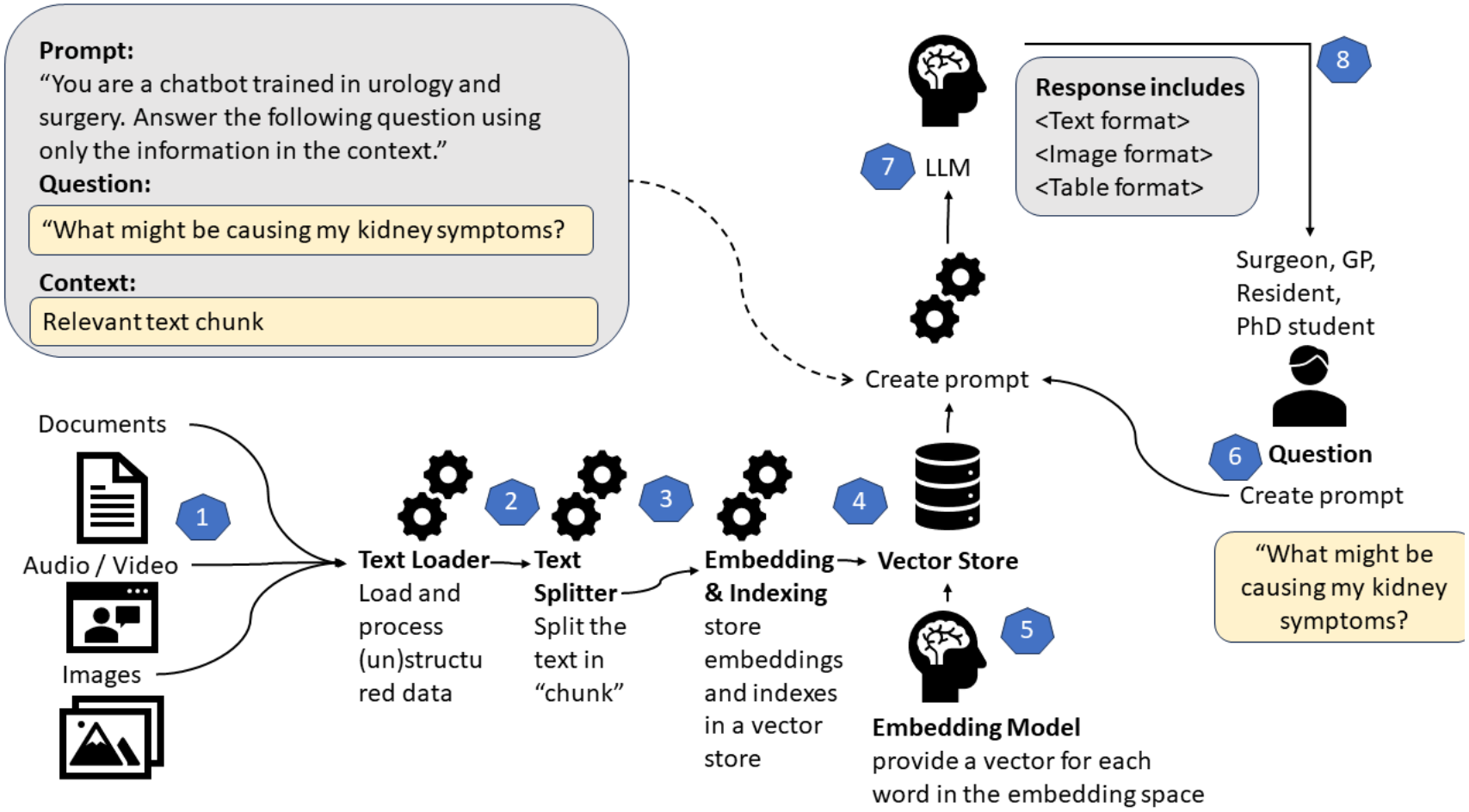
https://www.irarp.com - https://orsi-online.com

- iRARP is a Brussels-based software and robotic company that helps urologists prepare for a Robot-Assisted Radical Prostatectomy (RARP) intervention by providing a device-independent software.

- iRARP is a spin-off from Orsi Academy, a training & innovation center in minimal invasive & robotic surgery

RAG LLM Architecture

RAG architecture used in the case studies



Architecture design based on "All you need to know to build your first LLM app", Dominik Polzer, Published in Towards Data Science, Jun 22, 2023 https://towardsdatascience.com/all-you-need-to-know-to-build-your-first-llm-app-eb982c78ffac

Case Study 1 – Query urology best practices document

Using LLMs to query urology best practices document

Objective

- Create CustomGPT to query a 160-page long PDF
- 1st task - identify the authors of the document (Authors are mentioned on the 1st page)

Outcome - (spoiler) unsuccessful

- The context was too long and had to be split
- ChatGPT-4 **could NOT** find the authors
- LLaMa2 7B parameter model **could NOT** find the authors
- Claude AI **DID** find the authors

Case Study 2 – Support the Service Desk with ChatBot

Using LLMs to query social security administrative instructions

Objective

- Social Security' collaborators query the administrative instructions based on the model built by OpenAI API, using a ChatBot like ChatGPT
- Context used was a > 600-pages unstructured PDF
- Fixed test set of 20 question / answer pairs prepared (supervised testing)

Outcome using Open-Source Software & OpenAI ChatGPT

- Accuracy: 17/20 (validated by the Subject Expert Matters)

Outcome using Microsoft Azure & OpenAI ChatGPT-3.5

- ~30% **correct** answers
- ~30% **incorrectly** answers
- ~40% **hallucinations** or **incomplete** answers

Case Study 3 – GTIN Barcodes

Check validity of new GS1 barcodes

Source : GS1

Context

- GS1 created the GTIN, a new kind of barcode

Objective

- Ask ChatGPT-4 to write the Python class to “check the validity of the GTIN code”
- We used the strict definition of the GTIN available on the GS1 website (to avoid questions concerning the quality of the requirements)

Outcome

- The code raises an “index out of bounds” because the generated code did not consider all GTIN structures (They can be 8, 12, 13, or 14 digits long)

Conclusion

- Code generation of strict business requirements is NOT reliable using ChatGPT-4

Conclusions and Discussion

- LLMs are not ready to take the jobs of IT people or Service Desk
  - Too many hallucinations
  - Not suited to provide precise answers
  - Not able to compute

- There are a wealth of false and misleading claims published regarding outcomes achieved using LLMs

- Most of the time the results of the RAG are sufficient for the Service Desk, but the LLM produces false or incorrect statements!
  (This requires a thorough study)

Contact Details

- Abdelkrim Boujraf (CEO)
  abo@alt-f1.be
  +32 497 480 970

- Seamus Carey (Business Engineer)
  sca@alt-f1.be
  +353 857 482 351

https://itea4.org

ITEA is the Eureka Cluster on software innovation

https://eurekanetwork.org

Thank you