



MAG (Medical Advice Generator)

Kátia M. Barros, Juan Brugada, and Andrew Rosenswie

Mentored by Sina Rampe and Markus Hinsche

05.07.2023



github.com/ARosenswie/DSR_Final_Project



Overview

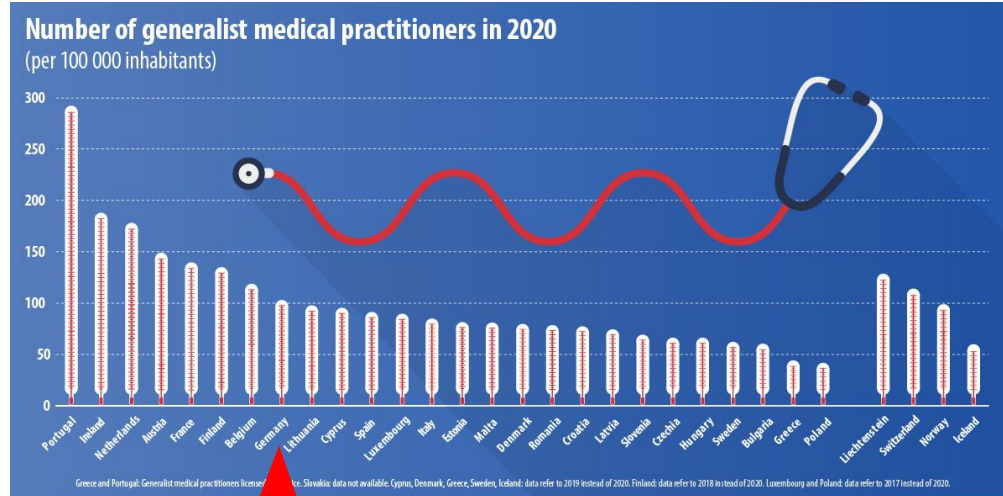
- Motivation
- Problem Definition
- Our Solution
- Model Concept
- *MAG*
- Evaluation
- Conclusion



"Doctors, teachers and tech workers are all in short supply in Germany. So the government is now looking to recruit new talent and make the country more attractive to foreigners with the needed qualifications. That means changing the laws around immigration to **bring in more people from outside the EU.**"

Tatjana Schweizer, 06/23/2023

Motivation





Problem Definition

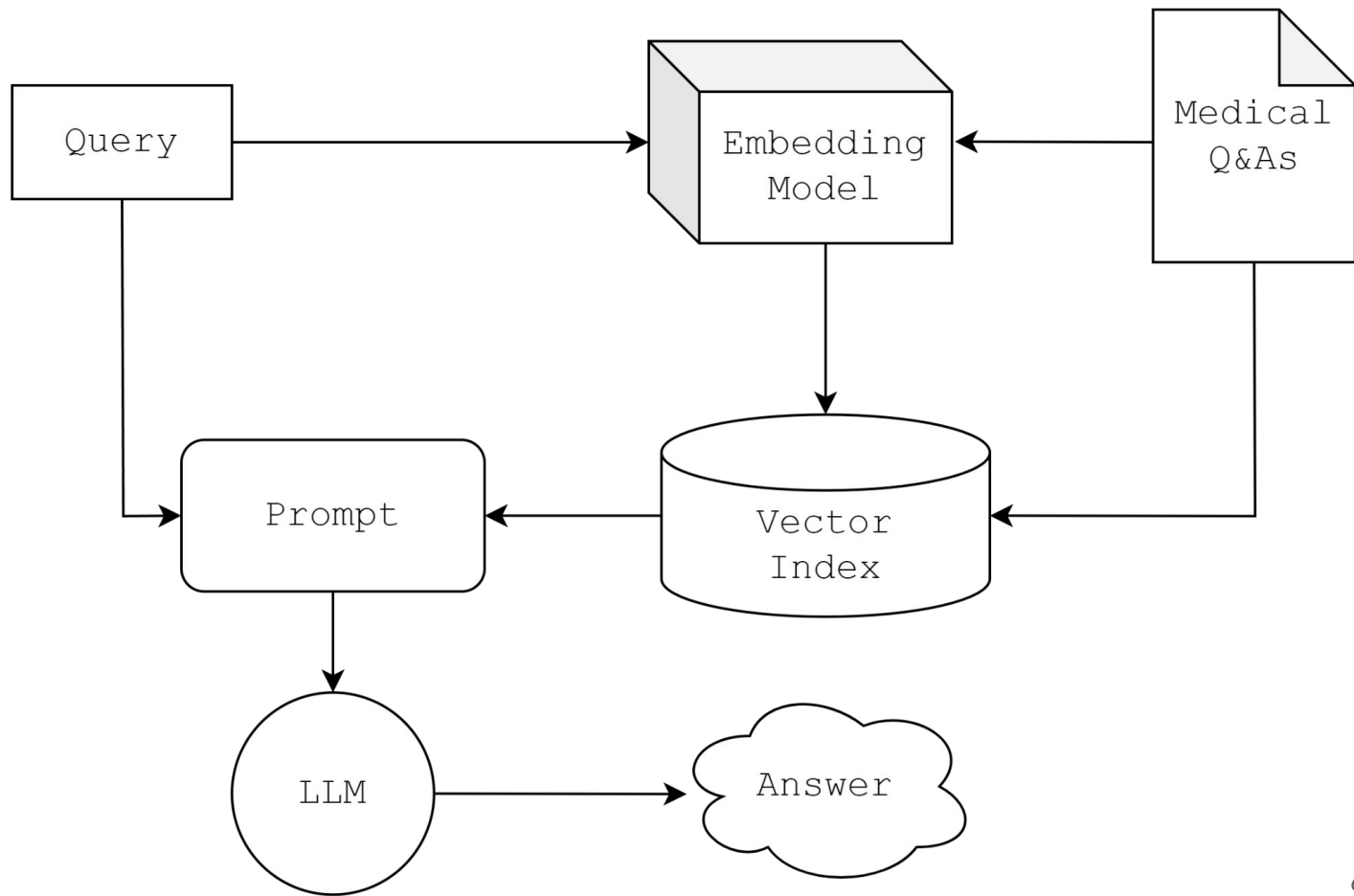
- Kenntnisprüfung
 - Examination for non-EU doctors to practice medicine within Germany
 - Patient screening and oral examinations
 - Costs:
 - € 266 - € 1100 (German state dependent)
 - Three attempts, two retakes
 - If failed, the international medical professional cannot practice medicine in Germany



Our Solution: ***Chatbot***

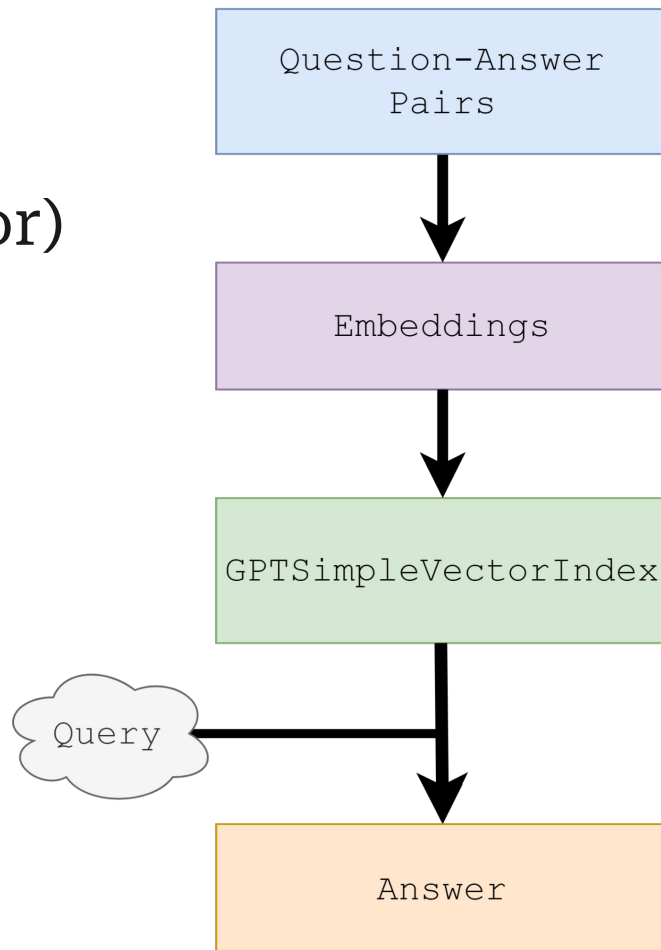
- Create an interactive Large Language Model (LLM) bot, that can aid medical professionals who are studying for the Kenntnisprüfung
- Collect Question-Answer (Q&A) pairings
 - PubMedQA^[6]
 - MedQA^[7]
 - MMLU^[4]
 - MedMCQA^[12]
 - **Total Q&A pairs: 570 712**

Model Concept



MAG (Medical Advice Generator)

- Essential llama index parameters:
 - Required: **Personal OpenAI API Key**
 - Prompt Helper:
 - max input size = 4096
 - num outputs = 2000
 - max chunk overlap = 20
 - chunk size limit = 600
 - Generative Pre-trained Transformer (GPT)
 - model = *text-davinci-003*
 - 175B parameters
 - 45 TB text-data
 - inference temperature = 0.1

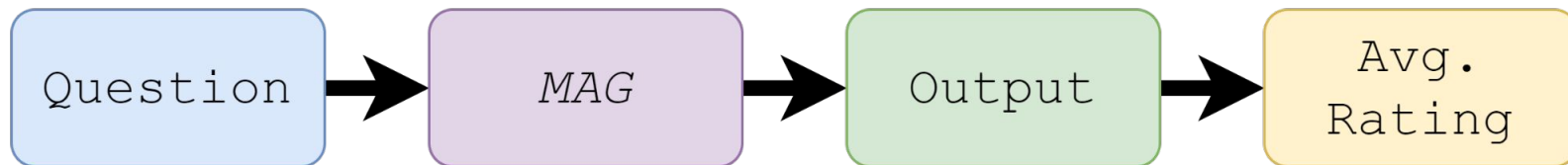




Live Demonstration with Audience Input

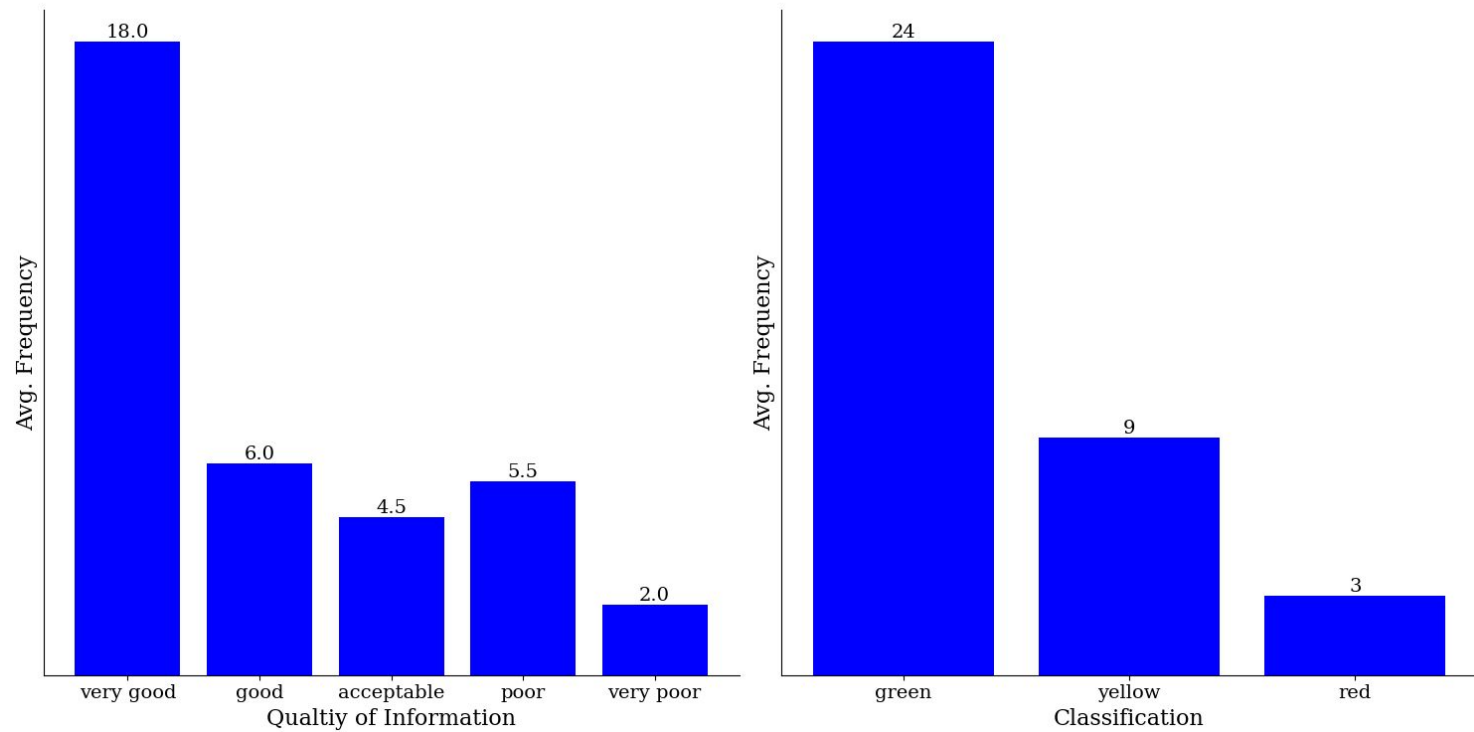
Evaluation

1. Medical expert evaluation of 36 questions from various medical sources^[2, 3, 5, 10, 11] answered by *MAG*
 - a. Quality of Information (very good, good, acceptable, poor, and very poor)
 - b. Classification (green, yellow, and red)
 - c. Dr. Tijana Mancic and Sina Rampe
2. Metric calculations



Evaluation

Average Expert Evaluation of *MAG*





MAG vs. ChatGPT

Metric	Score Range	MAG_{Score}	GPT-3.5 _{Score}	Source
BLEU	0-1	0.049	0.012	[9, 13]
MAUVE	0-1	0.544	0.004	[14]
METEOR	0-1	0.239	0.207	[1]
ROUGE-1	0-1	0.223	0.116	[8]
ROUGE-2	0-1	0.068	0.030	[8]
ROUGE-L	0-1	0.165	0.080	[8]
ROUGE-L _{sum}	0-1	0.163	0.080	[8]
F1-score _{avg}	0-100	20.402	10.968	[15]

How does *MAG* compare with ChatGPT?



Textbook Question:

How does omeprazole suppress *Helicobacter pylori*?

Textbook Answer:

In vitro, omeprazole inhibits the growth of *H. pylori* below pH 7. Clinically, it is thought that omeprazole enhances the local immune response by increasing intragastric pH. It also reduces the washout of antibiotics from the mucosa and lowers the inhibitory concentrations of pH sensitive antibiotics

***MAG* Answer:**

Omeprazole suppresses *H. pylori* by inhibiting the enzyme H⁺/K⁺-ATPase, which is responsible for acid production in the stomach. This reduces the acidity of the stomach, making it less hospitable for *H. pylori* to survive. In addition, the inhibition of H⁺/K⁺-ATPase may also prevent the interaction between coactivators p300 and PCAF and the foamy viral trans-activator, which could further reduce the survival of *H. pylori*.

How does *MAG* compare with ChatGPT?



Textbook Question:

How does omeprazole suppress *Helicobacter pylori*?

Textbook Answer:

In vitro, omeprazole inhibits the growth of *H. pylori* below pH 7. Clinically, it is thought that omeprazole enhances the local immune response by increasing intragastric pH. It also reduces the washout of antibiotics from the mucosa and lowers the inhibitory concentrations of pH sensitive antibiotics

CHATGPT Answer:

Omeprazole is a medication known as a proton pump inhibitor (PPI) that is commonly used to reduce the production of stomach acid. While omeprazole itself does not directly suppress *Helicobacter pylori* (*H. pylori*) bacteria, it plays an important role in the treatment of *H. pylori* infections.

H. pylori is a bacterium that can infect the lining of the stomach and upper small intestine, leading to various gastrointestinal conditions such as gastritis, peptic ulcers, and even stomach cancer in some cases. One of the factors that contribute to the development and persistence of *H. pylori* infections is the acidic environment of the stomach. The bacteria are adapted to survive and thrive in this acidic environment.

Here's how omeprazole helps in the suppression of *H. pylori*:

1. Acid Reduction: Omeprazole inhibits the action of the proton pumps in the stomach lining, which are responsible for producing stomach acid. By reducing acid production, it raises the pH level in the stomach, making it less acidic. This change in the stomach environment makes it less favorable for the growth of *H. pylori* bacteria.
2. Enhanced Antibiotic Effectiveness: In the treatment of *H. pylori* infections, a combination of antibiotics is usually prescribed along with a PPI like omeprazole. By reducing the acidity in the stomach, omeprazole creates an environment that enhances the effectiveness of antibiotics against *H. pylori*. The reduced acidity allows the antibiotics to work more efficiently and reach the bacteria more effectively.
3. Healing of Ulcers: *H. pylori* infections are associated with peptic ulcers. Omeprazole, by reducing stomach acid production, helps in the healing of these ulcers and provides relief from the associated symptoms.

It's important to note that omeprazole alone is not sufficient to eradicate *H. pylori* infection. Combination therapy involving multiple antibiotics, usually along with a PPI like omeprazole, is the standard treatment approach. The antibiotics directly target and kill the bacteria, while omeprazole creates a more favorable environment for their effectiveness. Therefore, it's crucial to follow the prescribed treatment regimen and complete the full course of antibiotics to effectively suppress and eliminate *H. pylori*.



Conclusion

- **MAG, the chatbot that aids non-EU medical professionals to become licensed doctors in Germany surpassed ChatGPT**
- Future model selection(s)
 - BERT, GPT-4, LaMDA, opt-impl-max-30b, ...
- Human vetting is needed for validity
 - More medical professional evaluators
- Limited knowledge based from Q&A pairings
 - Advanced topics -> advanced Q&A
- Other information:
 - AMBOSS: education and clinical references
 - <https://www.amboss.com/de>
 - Non-interactive
- Other future releases:
 - Galen: The AI Chatbot for Medicine
 - <https://www.galenai.org/>



References:

1. Banerjee, S., & Lavie, A. 2005, in Proceedings of the ACL Workshop on Intrinsic and Extrinsic Evaluation Measures for Machine Translation and/or Summarization (Ann Arbor, Michigan: Association for Computational Linguistics), 65–72.
<https://www.aclweb.org/anthology/W05-0909>
2. Differential diagnosis and treatment in primary care. Collins D. 6th Edition. Wolters Kluwer, 2018
3. Differential diagnosis in medical care. Rasul, N., Syed, M Wiley-Blackwell, 1st Edition, 2009
4. Hendrycks, D., Burns, C., Basart, S., et al. 2021, Proceedings of the International Conference on Learning Representations (ICLR)
5. Internal Medicine, just the facts. Schmitz P., Martin, K., Miller, D. McGraw Hill Medical, 2008
6. Jin, Q., Dhingra, B., Liu, Z., Cohen, W., & Lu, X. 2019, in Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP), 2567–2577
7. Jin, D., Pan, E., Oufattole, N., et al. 2020, arXiv preprint arXiv:2009.13081
8. Lin, C.-Y. 2004, in Text Summarization Branches Out (Barcelona, Spain: Association for Computational Linguistics), 74–81.
<https://www.aclweb.org/anthology/W04-1013>
9. Lin, C.-Y., & Och, F. J. 2004, in COLING 2004: Proceedings of the 20th International Conference on Computational Linguistics (Geneva, Switzerland: COLING), 501–507. <https://www.aclweb.org/anthology/C04-1072>
10. One thousand questions and answers in Clinical Medicine, Kumar, P, Clark M. 2nd Edition, Elsevier, 2011
11. Oxford Handbook of Clinical Medicine. Longmore M., Wilkinson I., Baldwin, A., Wallinl E. , Oxford University Press. 9th edition, 2014
12. Pal, A., Umapathi, L. K., & Sankarasubbu, M. 2022, in Proceedings of Machine Learning Research, Vol. 174, Proceedings of the Conference on Health, Inference, and Learning, ed. G. Flores, G. H. Chen, T. Pollard, J. C. Ho, & T. Naumann (PMLR), 248–260.
<https://proceedings.mlr.press/v174/pal22a.html>
13. Papineni, K., Roukos, S., Ward, T., & jing Zhu, W. 2002, in , 311–318
14. Pillutla, K., Swayamdipta, S., Zellers, R., et al. 2021, in NeurIPS
15. Rajpurkar, P., Zhang, J., Lopyrev, K., & Liang, P. 2016, in EMNLP



Thank You!



Contact:

<https://www.linkedin.com/in/katia-mendes-de-barros/>

<https://www.linkedin.com/in/juan-brugada/>

<https://www.linkedin.com/in/andrew-rosenswie/>



Appendix

- Inability to answer non-medical related questions:
 - Question: *Who is the best soccer player in the world?*
 - Answer: It is not possible to answer this question with the given context information.
 - Question: *What can cause allergies?*
 - Answer: Pollen, organisms, chemicals, plants, dead skin particles from dogs and cats, insect stings, and some food can all cause allergies. Margaret Chase Smith was instrumental in making Father's Day a national commemorative day in the United States.



Datasets: topics included in each

1. PubMedQA: Jin et al., 2019 (273,387 Q&A)
 - a. Biomedical texts with such topics: adolescent, adult, aged, child, cohort studies, cross-sectional studies, female, male, follow-up studies, human, infant, middle aged, pregnancy, preschool, predictive value of tests, prognosis, prospective studies, retrospective studies, risk factors, sensitivity and specificity, surveys and questionnaires, time factors, treatment outcome, young adult, 80 and over
2. MedQA: Jin et al., 2020 (12,721 Q&A)
 - a. Anatomy, biochemistry, cellular biology, first aid, gynecology, histology, immunology, internal medicine, neurology, obstetrics, pathology, pediatrics, pharmacology, physiology, psychiatry, surgery.
3. MMLU: Hendrycks et al., 2021 (100,338 Q&A)
 - a. Anatomy, clinical knowledge, college medicine, human aging, medical genetics, nutrition, professional medicine, virology
4. MedMCQA: Pal et al., 2022 (187,005 Q&A)
 - a. Anesthesia, anatomy, biochemistry, dental, ENT, forensic medicine (FM), obstetrics and gynecology (O&G), medicine, microbiology, ophthalmology, orthopedics, pathology, pediatrics, pharmacology, physiology, psychiatry, radiology, skin, preventive & social medicine (PSM), surgery



What is langchain and llama-index?