What did I get done today?

- Read some literature and asked Poe.com assistant bot questions on the topics of RAG and dialogue systems.
- Planned how to express the ideas of RAG systems to non technical stakeholders.

How well did it go?

Not a great day today either. Quite inefficient with a lot of time just staring at the pages without reading. I sum it up to my insomnia catching up to me. Unemployment has not been great for my sleep cycle. But no more excuses. I will write down some planning for the design document here and then start on that tomorrow. If I can't sleep tonight either I might work some more tonight and make another commit to the git later tonight.

After reading about RAG systems and refreshing my memory on them. (Never bothered memorizing any of the math before now.) I now have a pretty good idea on how to explain them to non technical stakeholders. A possible line for the design doc could be.

> *"What differentiates RAG systems from normal chatbots is that one can metaphorically view the RAG-system as having a compass that shows the direction of the topic the user provided. And then by comparing the user's topic with the directions of all topics it has viewed before, the system knows which topic path to take by adding the closest path according to the direction into the chatbot's memory."*

Reading my idea now I think I need to think some more about it. Test it on somebody non technical to see if they get the metaphor. But it should convey the idea of how the knowledge retrieval work if anybody asks. Which is always important to prepare for if you want other people's confidence.

The idea comes from the fact that RAG-systems use a vectorized explicit knowledge base (data base) to compare the vectorized input to the transformer network and select the most similar vector data if the data is in some radius from the input. I.e. the most similar direction dictates the chosen knowledge. The magnitude of the vectors matters as well if one uses the dot product similarity instead of simply the cosine similarity. But I do think that idea is somewhat relayed with the "path" metaphor.

Lastly I do have now some more idea of how I want to make the interactive prototype. Doing the vector-database would be overkill just for an interactive demo. So I will replace it with some JSON formatted data and a regex search function. Will suffice for the demo purpose.

What will I do and be done with tomorrow?

Well I will probably do some more stuff tonight and make a second commit later.

As I did not get anything done today on the design doc I need to get going with that. No more reading. I will try to get done all the images and visual representations of the app and holistic system overview. Partially so that I have something to strive towards for the prototype. And I find it easier to write about an image then the other way around so I will start with the figures.

Time wise I think I will be okay so long that I start on the prototype Monday morning. The demo prototype won’t be a looker with only 3 work days time. But I think I can get it to communicate the idea of the app.

This gives me 2 more work days left to get the design document done if I do not want to work over the weekend. I will try to make the design document double as a resource overview as well and define how much the expected cost of the app will be. This will probably take the most amount of time.

So if I get the images done tomorrow and then the text the day after I think I will still be on track. I am a bit worried if I will have time for the cost estimation though. I have no idea how much Preem pays for IT people. But I will check the Union web page and use that as an estimation.