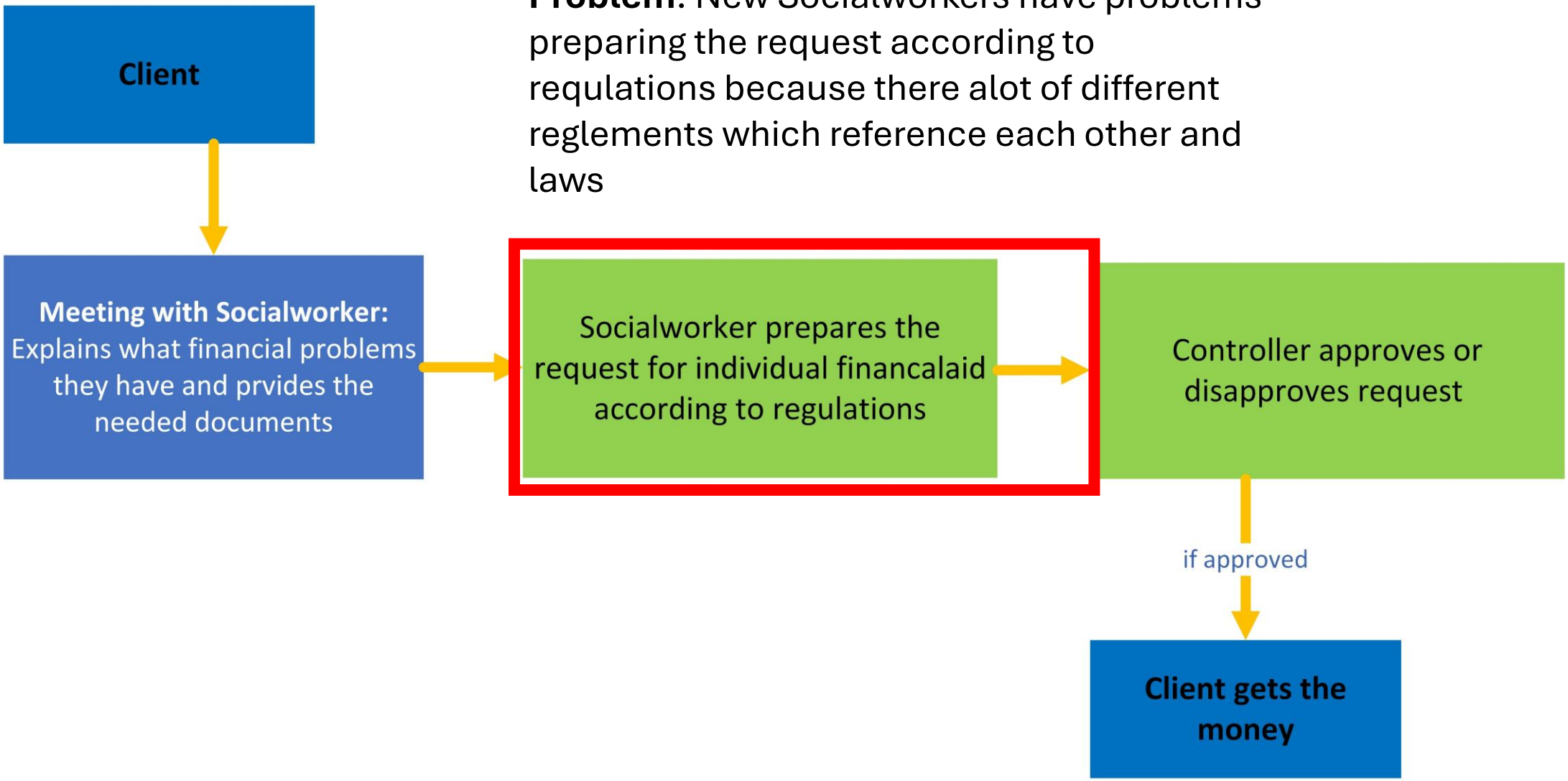
AI assisted searchengine in reglements for new socialworkers

Organisation: Pro Senectue Aargau

Project Team: Dragan, Anand, Mateo, Elena

Current state

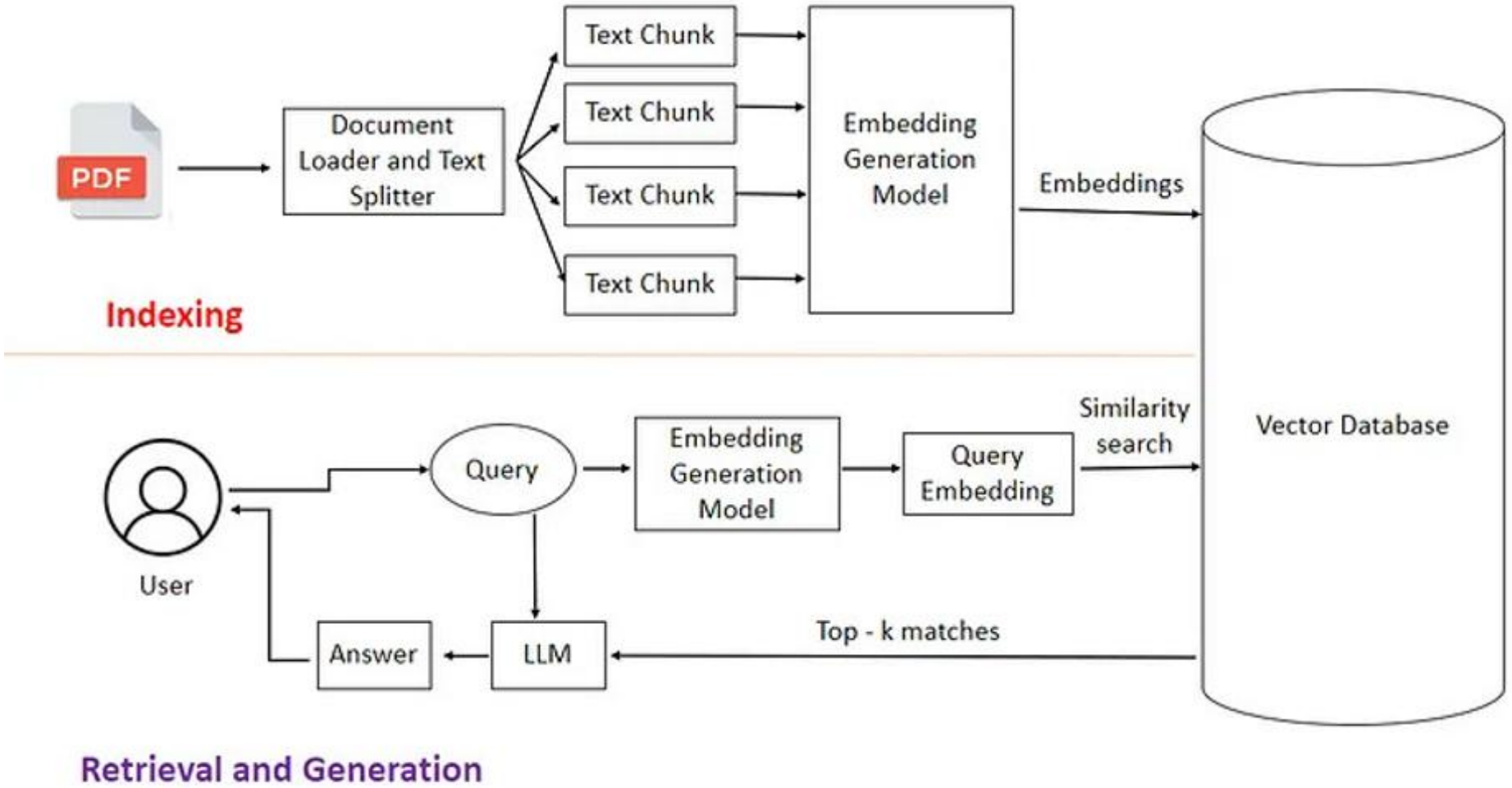
Problem: New Socialworkers have problems preparing the request according to requlations because there alot of different reglements which reference each other and laws



Goal

- Create a AI-assisted Searchengine in the provided documents
- Answers must be given from the provided documents
- Answers must contains a source (within the document)

The prototype



To-do list

- (ok) Make gradio app usable from the web for testing purposes, with ngrok done.
- (ok) Make sure that all documents in folder are being loaded and used for processing
- (ok) create a button for clearing the input
- (ok) alles auf Deutsch
- (ok) Instruct RAG to reference to the document, page and text section that it refers to.
- (ok) showing examples for questions
- (ok) create an option to chose between most relevant clues for the question (k=3-5, temp=0.1, different prompt) or most of clues (k=20, temp=0.6, openly formulated propmt)
- (ok) Define a template for the prompt: like "I am a new SW at Pro Senectute help me answer this question... (Mateo)

•(ok) Test the example questions and save the output (Mateo/Anand)

•(ok) Define a file for prompt_template for development (Dragan)

•(ok) output the used chunks in a file for validation

•(ok) make a presentation of our challenge (Mateo)

•make a version running on a linux server

•(ok) test bigger models 100B+ for inference (with larger GPUs)

•analyze caseexamples for usable questions to ask

•(ok) log every session in a log_file

Nice to have

- create a feature to be able to save the queries/prompts for further use
- create a feature to load additional documents beside the available ones
- link all the related documents that were used for asissting
- create a docker enviroment for the roll out
- create a login feature for SW based on their AD credentials
- extend the feature of giving feedback for the output with thumbs up or down and collect the it in log and make a simple frequency statistics on the feedback

Github QR-Code