*-* ***Problem****: Write a 1 sentence description of the problem you're trying to solve.*

I want to create an assistant for auditors which saves them from spending countless hours searching through the proxy document to verify a list of specific items of interest.

*-* ***Why****: Write a paragraph about why this is a problem worth solving.*

Creating this assistant will increase productivity and improve efficiency for auditors. By streamlining operations in this way, we will reduce costs associated with manual labor, and improving accuracy in compliance. Also, this will free up the auditors to do more audits and focus on higher-value audits.On a broader scale, automating proxy document audits contributes to greater transparency and personally, for me it’s creating innovative solutions in a matter of hours which seemed impossible before.

*-* ***Success****: What is the Key Performance Indicator?*

Success in this case will mean a significant reduction in the time and resources spent on manual audit processes, with an increase in accuracy and efficiency. KPI:

* Percentage decrease in audit processing time
* Error rates decrease compared to manual audits.

*-* ***Audience****: Specify the user/stakeholder/customer ,What is their pain point*

The target users for this product are auditors in corporations or financial institutions. Their pain point is the time-consuming nature of the manual process of auditing proxy documents, which includes searching for specific items, cross-referencing information, and verifying accuracy, leading to inefficiency, potential errors, and increased workload. Transitioning this process into semi-automatic to fully automated assistant reduces the need for labor-intensive tasks while still allowing for necessary human supervision to ensure maximum compliance.

*-* ***Potential Solution****: Briefly describe data & modeling techniques that you plan to leverage. Incorporate RAG as well as fine-tuning of embeddings or an LLM chat model.*

I’m using the “Proxy document” from CVS’ website. It is mostly text and a few tables and images. As for the questions, I will re-create a list of questions which auditors usually search for in the document (items of interest).

*Document*: CVS Proxy document for 2023, *LLM*: gpt turbo 3.5

*Embedding model:* text-embedding-3-small , Additional tech Stack: Chromadb (or FAISS as vector database) - Chailit + deploying on database

*-* ***Sharing****: What* ***community*** *:* Hugging Face, Github, and hopefully my team at work.