**Intro**

**Review proposal**

**Quick demo of software and documentation**

**Review GPT-3 Playground and gpt-3/PDF VSCode projects**

**Discuss project challenges**

* Not being able to efficiently determine accuracy/loss other than potential yes/no responses (since we can indicate 1 token to for a yes or no response, but can’t grade responses at scale.
* Having client side settings, which to use, and having to manually retest previously working responses after making changes to correct bad responses.
* Getting the correct text prepended to the GPT-3 query without providing to much that it gets confused
* To many potential questions and documentation covering entire software suite for a short project, maybe just focus on the web app intended for end users? (excluding windows client and management tools used by admins)

We will just do web

**Discuss next steps/tasks**

1. What type of webcontrol, page etc to display chatbot discussion (whilehiding prepended text) and thumbs up/down buttons, and what exact technology will the python scripting that retrieves the PDF text and submits the GPT-3 restapi query? I think labs covered Gradio/ngunk as an easy method or amazon aws lamda serverless?
2. What label annotation software to flag responses GPT-3 shouldn’t handle and what type of decision tree or neural network classifier to use? Or maybe just come up with keywords like “software licensing” to avoid needing a ton of manually created training data to train?
3. Create more real question and answer examples
4. Create Yes/No question and answers that can automatically determine accuracy? (if this makes sense)
5. What are the list of word replacements for common alternatives such as “IRM”->”iManage Records Manager”? Also will need to ensure all questions submitted to GPT-3 contain the phrase “imanage records manager” (and does case matter?)
6. If multiple people are accessing GPT-3 we should use separate API keys and $18 of credits will be given once registered with an email and cellphone.
7. How do we determine exactly what words in the user’s question are searched for pdf text, all words that aren’t super common like “the” “question”, etc or specifically search for records

management software terminology and pick out those? And how much surrounding text, previous and next sentence or entire paragraph?

Jack to try creating several orange section prompts and see if 3-5 pages of text is to much,

1. In order to ensure demo w/ FSDL goes smoothly let’s try to find 5 high quality questions and answers early on so we can be confident it will at least answer those questions correctly at Temperature = 0 (to ensure reproducibility).