User Profile

* Authentication (handled using AWS Cognito)
* /user/create: add user to database
* /user/get: retrieve user details
* /user/update: update specific user profile details

User Workspaces [user authenticated]

* Each user can create a MAX\_LIMIT number of workspaces
* Each workspace can have MAX\_LIMIT assistants
  + /workspace/create: create a workspace and attach it with user
  + /workspace/get: fetch list of workspaces
  + /workspace/{id}: fetch specific workspace details
  + /workspace/{id}/assistants: fetch specific workspace list of assistants
  + /workspace/{id}/assistants/{assist\_id}: fetch specific assistant details
  + /workspace/{id}/invite: invite user to this workspace
  + /workspace/delete/{id}: deletes all assistants of specific workspace and removes workspace

Datasets

* A dataset can have one more from files, raw text, or website
* One or more dataset can be attached to an assistant
  + /dataset/create: create an empty dataset, and an index on pinecone
  + /dataset/{id}/add: add one of the above mentioned files. Generate vector embeddings and store it under specific dataset index
  + /dataset/{id}/get: returns list of available files in dataset
  + /dataset/{id}/get/{doc\_id}: returns embeddings of document

External Dataset

Assistant

Data

Prompt

Tools

How to store this RAG model? How will we finetune and keep track of the updated model?

Handling of API Keys

Pricing using Stripe

Entities

User

Workspace

Assistant

Dataset