

# AIM Midterm

## Task 1: Define your Problem and Audience

1. At pre-seed and seed-stage investing, team is everything – but finding and evaluating talent is a massive challenge.
2. Seed-stage investors face a fundamental challenge in finding and evaluating founders because, at this early stage, there is often little to no traction, revenue, or product-market fit to assess. Unlike later-stage investing, where financial metrics and customer adoption provide signals of success, seed investing relies heavily on qualitative factors—such as the founder’s vision, grit, and ability to execute under uncertainty.

## Task 2: Propose a Solution

1. These characteristics (vision, grit, execution ability) are difficult to diligence outside-in; rather investors use proxies for founder-market fit: professional and academic experiences. The more a founders’ experience matches to the problem they’re trying to solve, the faster pattern matching the VC will apply to decide if they will add them to the top of their prospect list. Solution: Natural Language Searchable founder database that allows you to define a startup domain and search for the best talent match with enriched data searched and scraped from the web. Starting with prompts like: which founder is best suited for infrastructure startups, the app will query against the contexts of our vector db for the closest matched candidates, then enrich those profiles with data scraped from the web. Finally, that combined context will be converted into quick tear sheets for top-of-funnel evaluation.
2. Tooling:
  - a. LLM: GPT4o-mini: solid all-around model balancing cost and speed for both orchestration and function-calling
  - b. Embedding Model: OAI text-embedding-3-small; good enough performance for the cost and size of corpus
  - c. Orchestration: LangGraph - easy to manage states for internal rag retrieval and external search using tavily
  - d. Vector DB: QDrant in-memory; easy for prototyping, would consider hosting via cloud or persistent on-disk as to avoid having to reembed each time the app runs
  - e. Monitoring: Langsmith: langsmith monitoring to understand traces and cost for each run
  - f. Evaluation: RAGAS; using SDG abstractions out of the gate and having better baked-in metrics for evaluating RAG system performance
  - g. User Interface: Chainlit; would eventually move to Lovable or V0 for better interfacing but Chainlit is good enough for hacking together prototypes
3. Agents: Retrieval:
  - a. Agentic RAG (is agentic RAG just RAG, after all)

- b. External search for getting additional information about founders using web-search to enrich the profiles of each founder

### Task 3: Dealing with the data

1. RAG Data: Scraped LinkedIn Data including everything from their page as well as additional information about the companies they are currently working for and their previous work and education history
2. The chunking strategy treats each profile as a single chunk but also captures metadata (each field can be filtered) to use hybrid search approaches. This allows both “SQL” like filtering as well as natural language search.
3. I will need data from web searching to enrich the report generation for the write-up of each founder’s background

### Task 4: End-to-End Prototype

- [HF Link](#)

### Task 5: Golden Test Data Set

1.

#### Results of RAGAS Eval Prior to Fine-tuning:

{'context\_recall': 0.3333, 'faithfulness': 0.5122, 'factual\_correctness': 0.4292, 'answer\_relevancy': 0.7233, 'context\_entity\_recall': 0.4103, 'noise\_sensitivity\_relevant': 0.1962}

	user_input	retrieved_contexts	reference_contexts	response	reference	synthesizer_name
0	Whit is the role of SignalRank Corporation?	[Name: Keith Teare\nLocation: Palo Alto, Calif...	[Experience]; "Consumer Internet"; Digital Mark...	The role of SignalRank Corporation, as describ...	SignalRank Corporation is a company where the ...	single_hop_specific_query_synthesizer
1	What JPRS do in Japan?	[Name: coinstats.app\nLocation: 42629\nAbout: ...	[about the IP Communications space, and partic...	The context provided does not contain any spec...	In Japan, JPRS was involved in deals for names...	single_hop_specific_query_synthesizer
2	What services does Airbnb provide?	[Name: Allen Narcisse\nLocation: Los Angeles, ...	[Experience]; ["Company Name"; "Start-up-Chris...	The provided context does not mention Airbnb o...	Airbnb, Inc. operates an online marketplace fo...	single_hop_specific_query_synthesizer
3	What role does Airbnb play in Christopher Ober...	[Name: Christopher Obereder\nLocation: Atherto...	[Name: Christopher Obereder Location: Atherton...	Airbnb plays a significant role in Christopher...	Airbnb is one of the industry giants included ...	single_hop_specific_query_synthesizer
4	How does Poornima Vijayashanker's experience i...	[Name: Poornima Vijayashanker\nLocation: Palo ...	[<1-hop>\n\nName: Poornima Vijayashanker Locat...	Poornima Vijayashanker's experience in cross-f...	Poornima Vijayashanker's role at Apple involve...	multi_hop_abstract_query_synthesizer
5	How does B. Pagels-Minor's advocacy for mutual...	[Name: B. Pagels-Minor\nLocation: Los Angeles,...	[<1-hop>\n\nName: B. Pagels-Minor Location: Lo...	B. Pagels-Minor's advocacy for mutual accounta...	B. Pagels-Minor's advocacy for mutual accounta...	multi_hop_abstract_query_synthesizer
6	How has Mark Goldenson's experience in artific...	[Name: Mark Goldenson\nLocation: Mountain View...	[<1-hop>\n\nEducation"; "HR and Recruiting"; "He...	Mark Goldenson's extensive experience in artif...	Mark Goldenson's extensive experience in artif...	multi_hop_abstract_query_synthesizer
7	How has Greg Badros contributed to brand trans...	[Name: Greg Badros\nLocation: Los Altos, Calif...	[<1-hop>\n\nExperience: ["Company Name"; "Glur...	Greg Badros has made significant contributions...	Greg Badros has significantly contributed to b...	multi_hop_abstract_query_synthesizer
8	How does Greg Badros' experience and expertise...	[Name: Greg Badros\nLocation: Los Altos, Calif...	[<1-hop>\n\nName: Greg Badros Location: Los AL...	Greg Badros' experience and expertise signific...	Greg Badros, with his extensive experience in ...	multi_hop_specific_query_synthesizer
9	How has Jeff Pressman's experience with Operam...	[Name: Jeff Pressman\nLocation: Los Angeles, C...	[<1-hop>\n\nName: Jeff Pressman Location: Los ...	Jeff Pressman's experience with Operam and OPE...	Jeff Pressman's experience with Operam, where ...	multi_hop_specific_query_synthesizer
10	How did Oliver Walsh's role at Aritzia contrib...	[Name: Oliver Walsh\nLocation: Los Angeles, Ca...	[<1-hop>\n\nExperience: ["Company Name"; "Invi...	Oliver Walsh's role at Aritzia significantly c...	Oliver Walsh served as the CMO & Board Directo...	multi_hop_specific_query_synthesizer
11	How did Poornima Vijayashanker's experience wi...	[Name: Poornima Vijayashanker\nLocation: Palo ...	[<1-hop>\n\nSkills: ["Software Development"; "P...	Poornima Vijayashanker's experience with OPEAR...	Poornima Vijayashanker's experience with OPEAR...	multi_hop_specific_query_synthesizer

2. The SDG questions generated from the profiles were mixed in their efficacy; while some got the “vibe” that the use case of this data would be linking founders to their backgrounds, some were more general about what companies. As such the metrics weren’t as performant as hoped out of the box.

### Task 6: Fine-tuning OS Embeddings

- [Link to HF Hosting](#)

## Task 7: Assessing Performance

### 1. Fine-tuned embedding vs original:

```
... {'context_recall': 1.0000, 'faithfulness': 0.7812, 'factual_correctness': 0.4720, 'answer_relevancy': 0.9383, 'context_entity_recall': 0.4722, 'noise_sensitivity_relevant': 0.4856}
```

- a.
  - b. The performance was much better using the fine-tuned embeddings – recall was 100%! (this probably means something is off with the methodology; perhaps over-fitting). Interesting, the multi-hop questions were still related to questions about one-profile; a better set of questions would be related to general questions that require searching across multiple chunks to get an answer. For example: “which founder is best for starting AI companies?”
- ### 2. Changes expected to be made:
- a. Overhaul of the chunking strategy: use parent-child embeddings to have the entire profile as well as more fine-grained chunks for larger context fields like “about sections” and work experience descriptions.
  - b. Reconfigure app to have better filtering like Text to SQL to get a better set of chunks to rank for similarity
  - c. Build more distinct states for the agentic flow with more customized tools for each agent (research, scoring, writing, editing, etc.)
  - d. Better front-end with better user updates to know where the agents are in sequence