

# AlumConnect : Intelligent Alumni Discovery & Insights using RAG and LLM

- MODULE - EBA5004 – PRACTICAL LANGUAGE PROCESSING
- TEAM - GROUP 18
- TEAM MEMBERS :
  - ESWAR RAJ RAJENDRAN(A0123045W)
  - SUDHARSHAN MURUGESAN(A0299028R)
  - SWETA PATTNAIK(A0296537M)

# INTRODUCTION

## Gaps in Career Mapping & Alumni Insights :

### **1. Difficulty in Mapping Resumes to Job Market Trends**

- Job descriptions evolve rapidly with emerging tools, roles, and technologies
- Resumes often lack alignment with real-time industry expectations
- Students and early professionals struggle to understand what skills are truly in demand

### **2. Alumni Data Underutilized in Career Preparation**

- Alumni possess valuable real-world insights, but this data is scattered or hidden
- Traditional career tools do not leverage alumni journeys for personalized guidance
- Missed opportunity to showcase how specific roles and skills connect to actual career paths

### **3. Need for Automated, Intelligent Matching Systems**

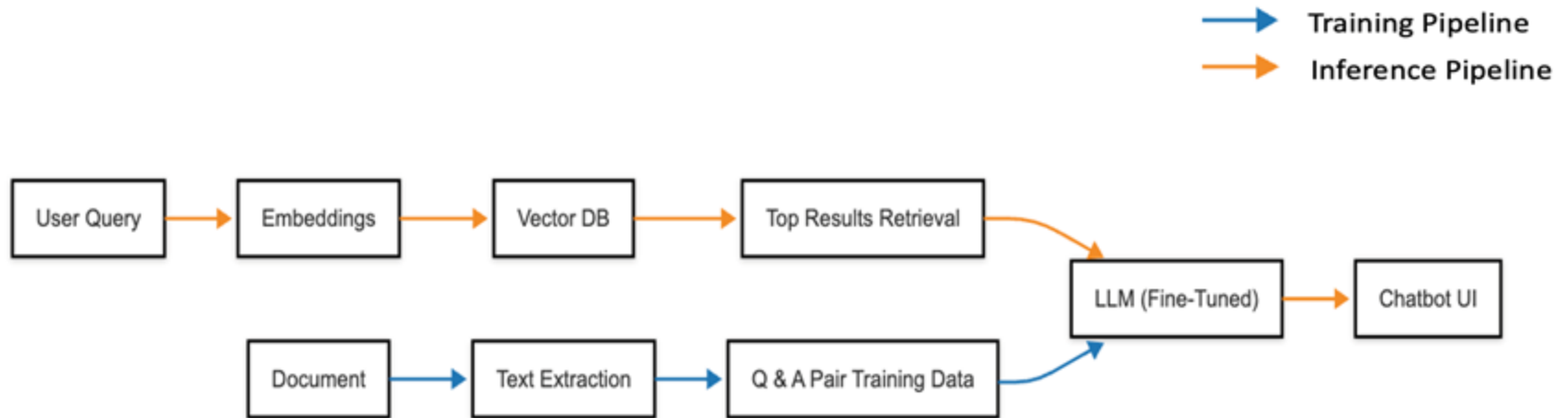
- Manual search is time-consuming and often leads to irrelevant results
- Existing platforms lack contextual understanding of user queries and profiles
- AI can bridge this gap by intelligently linking user intent to relevant alumni data

# INTRODUCTION

## **SOLUTION WE ARE PROPOSING :**

- To develop an AI-powered alumni search chatbot that efficiently retrieves and presents alumni profiles based on job titles, companies, industries, and skills.

# HIGH LEVEL DIAGRAM



The diagram represents the AI-powered Alumni Search Chatbot architecture, integrating Retrieval-Augmented Generation (RAG) with LLM fine-tuning.

The system operates in two main workflows:

1. Retrieval Workflow (Top Row) – active during inference
2. Fine-Tuning Workflow (Bottom Row) – Training pipeline

# HIGH LEVEL DIAGRAM

## **1) Retrieval Workflow (Top Row) – active during inference :**

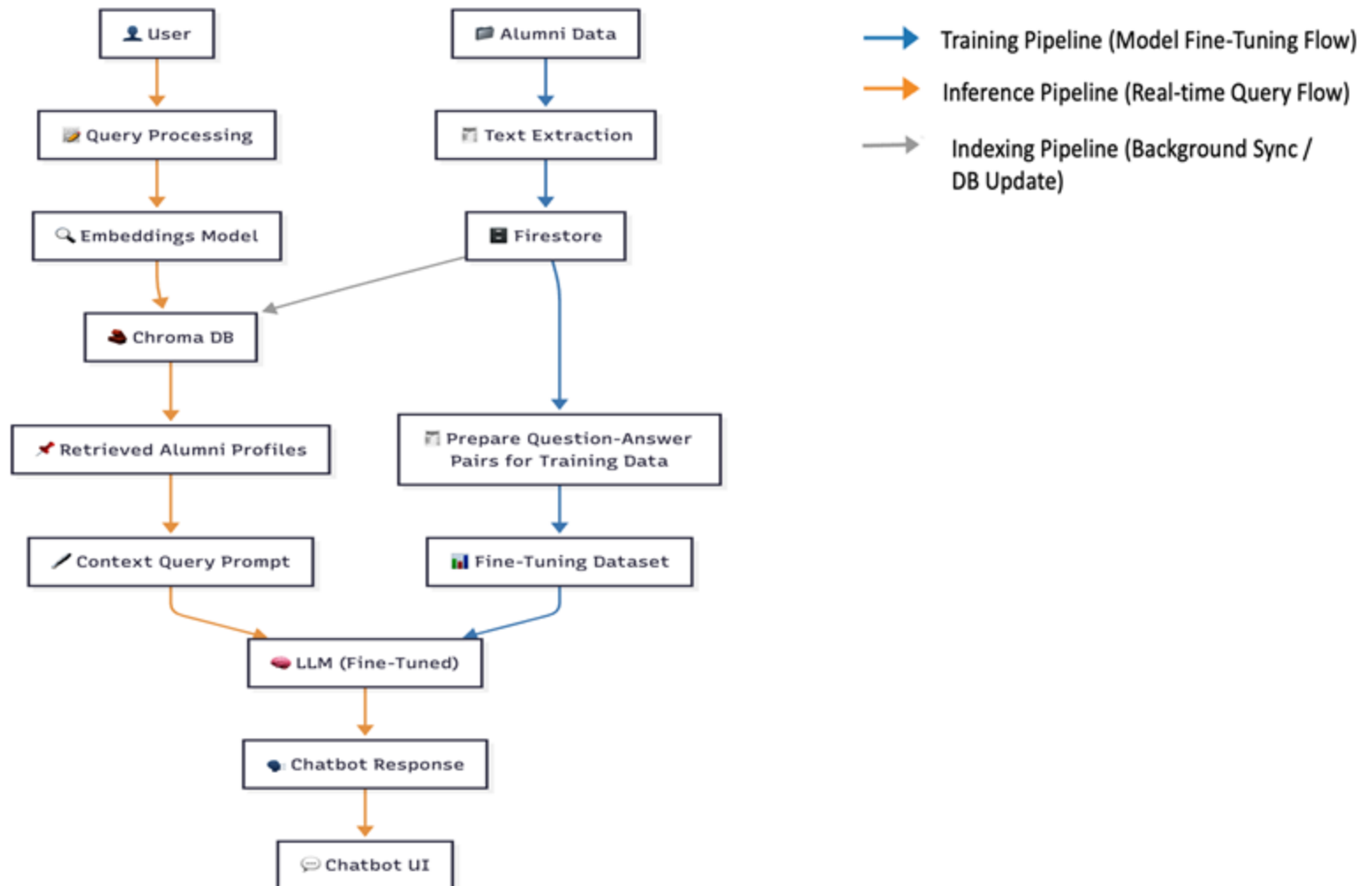
A user query is processed into embeddings using a pre-trained model. The embeddings are stored and searched in a vector database. Relevant alumni profiles are retrieved based on similarity and ranking. The LLM (Fine-Tuned) generates a context-aware response based on the retrieved results. The chatbot UI delivers a natural language response to the user.

## **2) Fine-Tuning Workflow (Bottom Row) – Training pipeline :**

Alumni resumes/documents are extracted for structured data processing. Extracted text is converted into Q&A pairs to create high-quality training data. These Q&A pairs are used to fine-tune the LLM for improved alumni-related query responses. The fine-tuned LLM integrates both real-time retrieval (RAG) and fine-tuned knowledge to ensure responses are accurate, contextual, and continuously improving.

This dual approach ensures the chatbot effectively balances retrieval efficiency with deep contextual learning, providing high-quality alumni search results.

# FLOWCHART FOR TRAINING AND INFERENCE PIPELINES



# PROJECT ROADMAP

- Data Collection & Processing
- Fine-Tuning Llama 3 with QA Pairs
- Backend API Development
- Implementing RAG for Intelligent Search for Gemini & Llama models
- Chatbot UI Development & Integration
- Inference testing
- Performance Comparison of Gemini & Llama models

# MODELS AND TOOLS USED



# MODELS USED

- **Llama model:**
  - Version : Meta-Llama-3-8B-Instruct
  - We have fine-tuned this model with our dataset and integrated with RAG during inference.
- **Gemini model:**
  - Version : gemini-1.5-flash
  - We have used this pretrained model for comparison with the finetuned model.

# TOOLS AND FRAMEWORKS

- **Coding Environment:** Google Collab
- **Database Tools:** Firestore
- **Vector Stores:** Qdrant, ChromaDB
- **Embedding Models:** Sentence Transformers (all-MiniLM-L6-v2), Google Generative AI Embeddings
- **LLM Integration Tools:**
  - LangChain (chains, memory, prompt templates)
  - Google Generative AI (via langchain-google-genai)
  - Ngrok (for local-to-web LLM API exposure in early prototype)
- **Configuration:** python-dotenv (for API keys & environment variables)
- **Utility Libraries:** requests, json, tqdm, shutil, traceback, re
- **UI Framework:** Gradio

# RUNTIME ENVIRONMENT

Component	Value
Platform	Google Collab Notebooks
GPU Used	NVIDIA T4 (16GB VRAM)
Training Strategy (fine tuning Llama)	20 batches × 4 hours = 80 hours total
Checkpointing	Intermediate checkpoints saved to resume from failures and for inference testing

# IMPLEMENTATION

# DATASET PREPARATION

**Dataset sources :** We used synthetic datasets generated from around 3500 resumes in Kaggle.

## **Resume Text Extraction:**

- **Notebook:** 1\_Resume\_Text\_Extraction\_N\_To\_DB\_Firestore.ipynb
- Reads **resume records** from .json source and .txt source
- Parses fields like: name, email, contact, location, job\_title, experience, skills, education, languages, etc.
- Uses **simple string manipulation** and **regex-based heuristics** to identify sections from text.
- Each parsed resume is formatted into a Python dict with consistent field keys.
- The cleaned records are uploaded to **Google Firestore**:
  - Collection name: alumni\_profiles
  - Each document has a unique Firestore ID.

# EXTRACTED DATA STORED IN GOOGLE FIRESTORE DB

Google Cloud

AI Alumni Chatbot

fire

X

Search

7

?

Firestore

Database

Firestore Studio

Indexes

Import/Export

Disaster Recovery

Time-to-live (TTL)

Security Rules

Insights

Usage

Query insights

Monitoring

Key Visualizer

Release Notes

Database

alumni\_profiles

01FaGMNsHwKoSOUmUy8G

alumni\_profiles\_1

Database

alumni\_profiles

01FaGMNsHwKoSOUmUy8G

alumni\_profiles\_1

alumni\_profiles

01FaGMNsHwKoSOUmUy8G

01YnYyFJeuXRjAgNrosg

020nofeJG2Wn0X9KjtZt

03xH74SwlqQlJ32ae2D

056JqWRxEQisClirrXx9

07JA2hqXyBlX6xzQa4Eg

09DhpwCHo7GeFR1gw3mR

09balWdoee8rY9PXqTxx

09eE1ovbCTKFtO23C5Dn

0BPiULxJa4VPB4bk8plQ

0HMRSveAlPRgVc3MedO3

0IL7ApJpY4CXMgygoGQb

0JFdhFpOcillsUs1sFpr

0JFhA6xCVBktKHn1fcUo

0JZb7450F8RpFIAQDgBl

0JwjCeYglpE423bgBUUG

0Kr8ZllamQP3IkEfKw0N

alumni\_profiles

ADD DOCUMENT

01FaGMNsHwKoSOUmUy8G

ADD FIELD

Education

0: "Singh and Sons (2018-03-26 - 2019-07-06)"

1: "MSc in Computer Science"

2: "GPA: 3.92"

3: "Honors: Dean's List"

4: "Conner LLC (2016-09-19 - 2021-12-16)"

5: "BSc in Computer Science"

6: "GPA: 3.52"

7: "Honors: Cum Laude"

Email: "paul80@example.net"

Location: "West Williammouth, Liberia"

Name: "VIRGINIA HAMMOND"

Phone: "954.738.6973"

Skills

0: "Python (expert)"

1: "C# (intermediate)"

2: "Angular (beginner)"

3: "Redis (intermediate)"

4: "PostgreSQL (beginner)"

5: "Google Cloud (beginner)"

6: "English (fluent)"

WorkExperience

0: "Software Engineer (2020-06-06 - 2023-11-23)"

1: "Dickerson PLC"

2: "Enhanced system security and performed regular audits."

3: "Ensured application responsiveness and seamless user experience."

4: "Conducted system monitoring and performance tuning."

5: "Developed and deployed scalable solutions."

source\_resume: "Virginia\_Hammond.docx"

# DATASET PREPARATION

## Q&A Pair Generation

- **Notebook:** 2\_Prepair\_Q\_&\_A\_Pairs.ipynb
- Loads the alumni profiles from Firestore.
- For each record, it:
  - Constructs multiple **hardcoded questions** (e.g., “Who has skills in Python?”, “Who worked at DBS?”)
  - Builds corresponding **answers** using the record’s fields.
- Formats these into **OpenChat-style JSONL** pairs:

```
{  
  "messages": [  
    {"role": "user", "content": "Who has skills in Python?"},  
    {"role": "assistant", "content": "Alice Tan has Python in her skills."}  
  ]  
}
```

# **TRAINING, RAG - INFERENCE USING LLAMA**



# LLM FINETUNING - QLoRA ( LLAMA)

- Fine-tune the **LLaMA 3** model on custom alumni Q&A data using **QLoRA**, optimized for constrained GPU environments (Collab T4).
- Loads the previously generated .jsonl dataset.
- Tokenizes data using a LLaMA-compatible tokenizer.
- Uses transformers, peft, and bitsandbytes to set up **QLoRA fine-tuning**:
  - Model: meta-llama/Meta-Llama-3-8B-Instruct
  - LoRA Config:
    - $r = 32$
    - `lora_alpha = 64`
    - `lora_dropout = 0.05`
    - Target modules: `q_proj`, `v_proj`, `k_proj`, `o_proj`
  - Quantization: 4-bit via `bnb_4bit`

# LLM FINETUNING - QLoRA (LLAMA)

## Training config:

- Optimizer: AdamW
- Learning Rate:  $2e-5$
- Batch size:  $4 \times$  accumulation steps 4  $\rightarrow$  effective batch = 16
- Precision: fp16 with gradient checkpointing
- Checkpoints: saved every 500 steps
- Total steps: 20230

# RAG IMPLEMENTATION (LLAMA)

## **Embedding Model:**

all-MiniLM-L6-v2 via SentenceTransformer

## **Vector Store:**

QdrantClient – hosted remotely

## **Retrieval Process:**

Manually queries Qdrant using vector similarity search

## **Query Flow:**

- User inputs query
- Text converted to embedding
- Qdrant returns top similar documents
- Documents + query passed to custom LLM endpoint via requests.post (Ngrok tunnel)

# INFERENCE (LLAMA)

## **Step 1: User Query**

- Input query captured via Gradio interface


## **Step 2: Document Retrieval**

- Query is embedded using SentenceTransformer
- Embedding passed to QdrantClient (RAG DB) for similarity search

## **Step 3: LLM Inference**

- Retrieved documents + user query are sent via POST request to the fine tuned Llama LLM endpoint exposed using Ngrok.
- Custom API performs text generation using this model.

# APP UI (LLAMA)

 **Alumni Information Chatbot**

Ask questions about alumni profiles. Examples: `What are Virginia Hammond's skills?`, `Find alumni who graduated after 2020 and live in California.`

Chat History

What is Virginia Hammond's contact information?

Virginia Hammond can be reached via email at `paul80@example.net`, by phone at 954.738.6973, and she resides in West Williammouth, Liberia.

Find alumni who live in California.

The alumni who live in California are:

- Lisa Rose
- Kimberly Barajas
- Leah Hoffman
- Angelica Walker

Your Question

Enter your question here...

Send

Clear Chat

⚡ Suggested Questions (Click any example to try):

What is Virginia Hammond's contact information?	Find alumni with Python programming skills.	Find alumni who live in California.	Find alumni who live in California and have MSc in Computer Science education	Who completed their degree after 2018 and have MSc in Computer Science ?
List alumni who majored in Electrical Engineering.	who graduated with computer science degree around 2020 and works in data science domain?	Who has experience with cloud computing?	Who graduated between 2018 and 2021 ? share their degree major ?	can you give information about John carter ?

# APP UI (LLAMA)

## Alumni Information Chatbot

Ask questions about alumni profiles. Examples: What are Virginia Hammond's skills?, Find alumni who graduated after 2020 and live in California.

### Chat History

List alumni who majored in Electrical Engineering.

The following alumni majored in Electrical Engineering:

- Mathew Campbell
- Suzanne Lee
- Abigail Johnson
- Allen Newton
- Mary Smith
- Emily Miller
- Allison Miller
- Ronald Rivera

### Your Question

Enter your question here...

Send

Clear Chat

### ⚡ Suggested Questions (Click any example to try):

What is Virginia Hammond's contact information?

Find alumni with Python programming skills.

Find alumni who live in California.

Find alumni who live in California and have MSc in Computer Science education

Who completed their degree after 2018 and have MSc in Computer Science ?

List alumni who majored in Electrical Engineering.

who graduated with computer science degree around 2020 and works in data science domain?

Who has experience with cloud computing?

Who graduated between 2018 and 2021 ? share their degree major ?

can you give information about John carter ?

# APP UI (LLAMA)

## Alumni Information Chatbot

Ask questions about alumni profiles. Examples: `What are Virginia Hammond's skills?`, `Find alumni who graduated after 2020 and live in California.`

### Chat History

can you give information about John carter ?

Yes, here is the information about John Carter:

Email: [jeffery26@example.org](mailto:jeffery26@example.org)

Phone: (864) 217-2320

Location: Port Christopher, Mexico

Education:

- BSc in Computer Science – Lyons-Smith (2016-05-13 to 2019-09-05)
- BSc in Computer Science – Robinson Inc (2015-07-01 to 2020-10-16)

Work Experience:

- Cybersecurity Engineer at Delacruz, Norris and Allen

### Your Question

Enter your question here...

Send

Clear Chat

### ⚡ Suggested Questions (Click any example to try):

What is Virginia Hammond's contact information?

Find alumni with Python programming skills.

Find alumni who live in California.

Find alumni who live in California and have MSc in Computer Science education

Who completed their degree after 2018 and have MSc in Computer Science ?

List alumni who majored in Electrical Engineering.

who graduated with computer science degree around 2020 and works in data science domain?

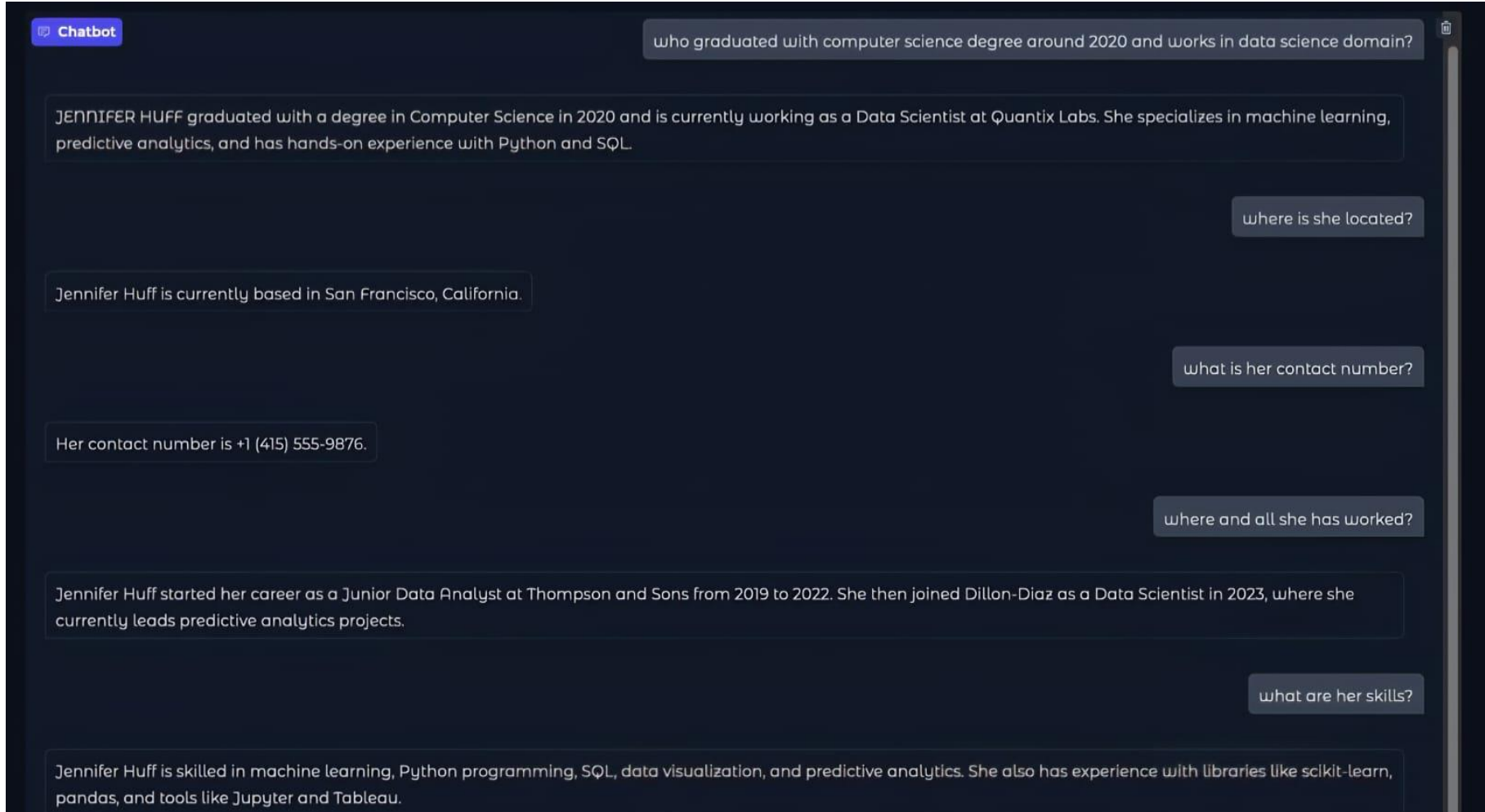
Who has experience with cloud computing?

Who graduated between 2018 and 2021 ? share their degree major ?

can you give information about John carter ?



# APP UI (LLAMA)





# **RAG & INFERENCE USING GEMINI**

# RAG IMPLEMENTATION (GEMINI)

## **Data Source:**

Extracted resume data is imported from Google Firestore DB.

## **Embedding Model:**

GoogleGenerativeAIEmbeddings via langchain-google-genai embedding-001

## **Vector Store:**

ChromaDB – embedded, fast local storage

## **Retriever:**

.as\_retriever() – wraps vector store into a retriever interface

## **Query Flow:**

- User query passed to ConversationalRetrievalChain
- Retriever pulls relevant chunks from ChromaDB
- ChatGoogleGenerativeAI responds using prompt template + context

# INFERENCE (GEMINI)

## Step 1: User Query

- Handled through Gradio chat interface

## Step 2: Context Retrieval

- Embedding generated using GoogleGenerativeAIEmbeddings
- Chroma vector store used to fetch relevant chunks

## Step 3: LLM Inference with Memory

- Query and retrieved documents passed to LangChain's ConversationalRetrievalChain
- Uses ChatGoogleGenerativeAI to generate responses with context from ConversationBufferMemory (temperature = 0.2)
- `retrieved_docs = retrieve_with_reranking(query, metadata_filter, k=kwargs.get('k', 10))`
- Prompt templates provide structured guidance

# PSEUDO CODE (GEMINI)

## // Initialize services

```
SET API_KEYS and ENDPOINTS for vector_database  
and language_model  
CREATE vector_database_client and  
sentence_embedder
```

## // Core function to answer queries

```
FUNCTION answer_query(query):  
    // Vector search  
    query_vector = embed(query)  
    profiles = search_vector_database(query_vector,  
limit=5)
```

## // Format results for context

```
IF profiles exist THEN  
    context = format_profiles(profiles)  
ELSE  
    context = "No relevant profiles found."
```

## // Get LLM response

```
prompt = build_prompt(context, query)  
return get_llm_response(prompt)
```

## // UI handler functions

```
FUNCTION respond(message, history):  
    reply = answer_query(message)  
    ADD message and reply TO history  
    RETURN updated_history  
FUNCTION clear():  
    RETURN empty_history
```

## // UI setup

```
CREATE chat_interface with:  
    - Chatbot component  
    - Text input  
    - Send and Clear buttons  
CONNECT buttons to handler functions  
LAUNCH interface
```

# APP UI (GEMINI)

## Alumni Information Chatbot

Ask questions about alumni profiles. Examples: 'What are Virginia Hammond's skills?'; 'Find alumni who graduated after 2020 and live in California.'

### Chat History

Hello! I can help you find information about alumni. Ask me anything about their skills, education, or work experience.

Who graduated between 2018 and 2021 ? share their degree major ?

The following alumni graduated between 2018 and 2021 with the following majors:

- **Amber Holt:** BSc in Computer Science (2021-10-21)
- **Mathew Campbell:** MSc in Computer Science (2020-09-11)
- **Abigail Johnson:** BSc in Computer Science (2021-02-12)

Your Question

Send

# APP UI (GEMINI)

## Alumni Information Chatbot

Ask questions about alumni profiles. Examples: 'What are Virginia Hammond's skills?', 'Find alumni who graduated after 2020 and live in California.'

### Chat History

can you give information about John carter ?

John Carter's contact information is:

- Email: [jeffery26@example.org](mailto:jeffery26@example.org)
- Phone: (864)217-2320
- Location: Port Christopher, Mexico

His education includes:

- BSc in Computer Science - Lyons-Smith (2016-05-13 - 2019-09-05)
- BSc in Computer Science - Robinson Inc (2015-07-01 - 2020-10-16)

His work experience includes:

- Cybersecurity Engineer at Delacruz, Norris and Allen
- Cybersecurity Engineer at Campbell, Price and Wise

His skills include:

- Java
- Python

### Your Question

Send

Enter your question here...

Clear Chat

Force Rebuild DB

### ≡ Suggested Questions (Click any example to try):

Tell me everything you know about Virginia Hammond. What are Virginia Hammond's key skills? What is Virginia Hammond's contact information? Tell me about Paul Walker's work experience. Did Paul Walker study Computer Science?

Find alumni with Python programming skills. List alumni with experience in machine learning. Who has experience with cloud computing? Find alumni skilled in data analysis. Which alumni know JavaScript?

# APP UI (GEMINI)

The following alumni completed an MSc in Computer Science after 2018:

- Lisa Rose (2022-12-13)
- Kimberly Barajas (2021-09-13)
- Eric Le (2021-06-03)
- Leah Hoffman (2021-06-13)
- Philip Welch (2022-04-04)

Who completed their degree after 2018 and have MSc in Computer Science ? Share their contact as well

The following alumni completed their MSc in Computer Science after 2018:

- **Lisa Rose:** Email: [tanya48@example.com](mailto:tanya48@example.com), Location: Cathychester, Luxembourg, Phone: 242-468-2730
- **Kimberly Barajas:** Email: [wellsmichelle@example.net](mailto:wellsmichelle@example.net), Location: Carlosberg, Saint Kitts and Nevis, Phone: 960-994-1820
- **Leah Hoffman:** Email: [brian99@example.net](mailto:brian99@example.net), Location: Andrewsburry, Nigeria, Phone: 413-485-1533

# APP UI (GEMINI)

## Chat History

- **Michelle Ford:** MSc in Computer Science (2015-09-28 - 2019-03-17)
- **Christopher Lopez:** BSc in Computer Science (2018-11-17 - 2020-11-12), BSc in Computer Science (2015-08-21 - 2021-11-09)

Find alumni who live in California and have MSc in Computer Science education

The following alumni have an MSc in Computer Science:

- **Lisa Rose:** Jones, Villegas and Williamson (2018-08-29 - 2022-12-13)
- **Emily Miller:** Johnson PLC (2016-02-21 - 2021-04-26)
- **Kimberly Barajas:** Mitchell, Wilson and Diaz (2017-05-15 - 2021-09-13)
- **Melissa Watson:** Leon Ltd (2018-12-09 - 2022-08-14)
- **Michelle Owens:** Morgan Inc (2017-07-22 - 2022-08-21)
- **Angelica Walker:** Henson, Nichols and Lewis (2016-12-05 - 2019-07-17)
- **Lisa Williamson:** Kennedy, Scott and Schmidt (2017-03-14 - 2019-08-13)
- **Leah Hoffman:** Campos, Doyle and Townsend (2018-04-17 - 2021-06-13)
- **Jacqueline Ford:** Rogers, Cortez and Scott (2015-09-28 - 2019-03-17)

## Your Question

Enter your question here...

Send



# EVALUATIONS AND RESULTS

# EVALUATION : ROUGE SCORES

## What is ROUGE?

ROUGE (Recall-Oriented Understudy for Gisting Evaluation) is a metric used to evaluate text generation quality by comparing model responses to reference answers. It is widely adopted in tasks like summarization and question answering.

## ROUGE-1 F1:

- Measures word-level overlap (unigrams) between the generated answer and the reference.
- The **F1 score** balances precision (accuracy of words used) and recall (coverage of key words).

## ROUGE-L F1:

- Based on the **Longest Common Subsequence (LCS)** — measures how well the sequence and structure of the generated text matches the reference, rewarding coherence.

# EVALUATION : ROUGE SCORES

QUESTION	RESPONSE LLAMA	RESPONSE GEMINI	ROUGE-1 F1	ROUGE-L F1
What is Virginia Hammond's contact	Virginia Hammond can be reached via email at vhammond@university.edu	Virginia Hammond's contact information is vhammond@university.edu	0.619047619	0.619047619
Find alumni with Python programming	Alumni skilled in Python for data analysis and machine learning	Virginia Hammond: "Perfected data analysis with Python"	0.417910448	0.328358209
Find alumni who live in California.	The alumni who live in California and have a degree in Computer Science	The following alumni studied Computer Science in California	0.179245283	0.141509434
Find alumni who live in California and have MSc in Computer Science education ?	The following alumni live in California and have a Master's degree in Computer Science	The following alumni have an MSc in Computer Science	0.323529412	0.308823529
Who completed their degree after 2018?	The following alumni completed their degree after 2018	The following alumni completed their degree after 2018	0.99543379	0.99543379
Who completed their degree after 2018 and have MSc in Computer Science ?	The following alumni completed their degree after 2018 and have a Master's degree in Computer Science	Lisa Rose (2022-12-13)Kimberly Barajas (2021-05-15)	0.555555556	0.555555556
List alumni who majored in Electrical Engineering	The following alumni majored in Electrical Engineering	Mathew CampbellSuzanne LeeAbigail Chen	0.837209302	0.837209302
who graduated with computer science degree around 2020	The following alumni graduated with a Computer Science degree around 2020	The following alumni completed a Computer Science degree around 2020	0.947368421	0.947368421
who graduated with computer science degree around 2020 and works in data science	This question cannot be answered definitively as the database does not track post-graduation employment	I cannot answer this question. The provided information does not include post-graduation employment details	0.329896907	0.226804124
Who graduated between 2018 and 2021 ? share their degree major ?	The following alumni graduated between 2018 and 2021	The following alumni graduated between 2018 and 2021	0.902439024	0.902439024
can you give information about John carter	Yes, here is the information about John Carter	Email: jeffery26@example.orgPhone: 555-123-4567	0.875912409	0.875912409

# EVALUATION : ROUGE SCORES

From the scores, what we see is:

## High Scores (ROUGE > 0.9)

- Example: *“Who completed their degree after 2018?”*
- Both finetuned LLaMA and Gemini scored ~0.995 on ROUGE-1 and ROUGE-L, indicating highly accurate and well-aligned responses.
- Example: *“Graduated between 2018 and 2021?”*
- Gemini scored 0.90+, reflecting strong understanding of date-based queries.

## Mid Scores (ROUGE 0.5–0.8)

- Example: *“Majored in Electrical”* — Gemini reached ~0.83
- Shows partial match with correct names and degrees, but possible variation in wording.

## Low Scores (ROUGE < 0.4)

- Example: *“Alumni in California”* — finetuned LLaMA scored 0.17, Gemini slightly better.
- Indicates vague, incomplete, or hallucinated responses likely due to insufficient or noisy context.
- *“Graduated around 2020 and works in data”* — Poor scores reflect inability to handle compound filters in a single query.

# EVALUATION: RESPONSE COMPARISON

QUESTION	RESPONSE LLAMA	RESPONSE GEMINI
What is Virginia Hammond's contact information?	Virginia Hammond can be reached via email at paul80@example.net, by phone at 954.738.6973, and she resides in West Williammouth, Liberia.	Virginia Hammond's contact information is:  Email: paul80@example.net Phone: 954.738.6973 Location: West Williammouth, Liberia
Find alumni with Python programming skills.	Alumni skilled in Python for data analysis and visualization include Virginia Hammond, John Smith, Alice Johnson, and Bob Williams.	Virginia Hammond: "Perfectd data analysis and data visualization using Python and Tableau." John Smith: "Perfectd data analysis and data visualization using Python and Tableau." Alice Johnson: "Perfectd data analysis and data visualization using Python and Tableau." Bob Williams: "Perfectd data analysis and data visualization using Python and Tableau."
Find alumni who live in California.	The alumni who live in California and have an MSc in Computer Science are: <ul style="list-style-type: none"><li>• Lisa Rose</li><li>• Kimberly Barajas</li><li>• Leah Hoffman</li><li>• Angelica Walker</li><li>• Jacqueline Ford</li><li>• Lisa Williamson</li></ul>	The following alumni studied Computer Science:  Michael Bartlett: BSc in Computer Science (2017-08-18 - 2022-11-28), MSc in Computer Science (2018-09-09 - 2021-07-04) Angelica Walker: BSc in Computer Science (2017-08-12 - 2023-02-03), MSc in Computer Science (2016-12-05 - 2019-07-17) Mathew Campbell: BSc in Computer Science (2015-04-07 - 2021-10-01), MSc in Computer Science (2017-10-10 - 2020-09-11) Lisa Williamson: MSc in Computer Science (2017-03-14 - 2019-08-13), MSc in Computer Science (2015-04-02 - 2022-03-23) Chris Allen: BSc in Computer Science (2016-07-04 - 2022-01-05), BSc in Computer Science (2016-09-09 - 2019-10-12) Lisa Rose: MSc in Computer Science (2018-08-29 - 2022-12-13) Kimberly Barajas: MSc in Computer Science (2017-05-15 - 2021-09-13) Leah Hoffman: MSc in Computer Science (2018-04-17 - 2021-06-13) Jacqueline Ford: MSc in Computer Science (2015-09-28 - 2019-03-17) Christopher Lopez: BSc in Computer Science (2018-11-17 - 2020-11-12), BSc in Computer Science (2015-08-21 - 2021-11-09)

# CHALLENGES AND LIMITATIONS

## GPU Constraints

- Running embedding models, inference, or fine-tuned LLMs locally or on limited Colab tiers often hits **memory and compute bottlenecks**.
- Real-time inference, especially with context-aware models, demands **efficient model selection** or offloading to cloud APIs, which adds cost and latency trade-offs.

## Hallucinations in LLM Outputs

- Large Language Models sometimes generate plausible-sounding but **factually incorrect** answers, especially when alumni data is sparse or loosely structured.
- This can mislead users during career planning or alumni research, making it crucial to improve **retrieval grounding** and control **prompt engineering**.

# CHALLENGES AND LIMITATIONS

## ngrok Free Hosting Constraints

- Using ngrok's free tier for hosting backend services (e.g., LLMs or APIs) introduces significant performance and availability limitations.

## Prompt-Template Instructional Fragility

- **Cascade-effect edits** – Any tweak to the prompt's instruction stack (e.g., adding stricter citation wording) unexpectedly shifted answers for unrelated queries. The template proved brittle: improving one case routinely broke another, so every minor change demanded a full regression sweep.
- **Guideline overload vs. clarity** – Packing tone, length, citation, and safety rules into a single template left the LLM juggling too many directions. We found that past a certain density, instructions began to blur together, reducing response consistency and forcing painful trade-offs between coverage and clarity.

# CHALLENGES AND LIMITATIONS

## Data Privacy & Scraping Limitations

- Public platforms like LinkedIn restrict automated scraping of user profiles due to **privacy policies and terms of service**.
- Aggregating alumni data at scale requires **ethical sourcing**, consent, or relying on **unstructured datasets** manually collected, affecting scalability.
- Due to these concerns, we used only synthetic datasets.



# CONCLUSION

- AlumConnect bridges the critical gap between static resumes and the evolving job market by harnessing the power of underutilized alumni data.
- Through a Retrieval-Augmented Generation (RAG) framework and fine-tuned LLMs, the system tries to intelligently match user queries with relevant alumni career paths, offering actionable, personalized insights.
- Key accomplishments include the development of workflows—retrieval and fine-tuning, integration of advanced models like LLaMA 3 and Gemini, and the creation of an interactive chatbot.
- Looking ahead, the project aims to scale by incorporating real university alumni datasets, deploying to the cloud for wider accessibility, and adding real-time feedback mechanisms to iteratively enhance LLM accuracy and relevance.

# REFERENCES

- LangChain RAG Overview:  
[https://python.langchain.com/docs/use\\_cases/question\\_answering/](https://python.langchain.com/docs/use_cases/question_answering/)
- Hugging Face Fine-Tuning Guide :  
<https://huggingface.co/docs/transformers/training>
- FAISS (Facebook AI Similarity Search) : <https://faiss.ai/>
- FastAPI for AI Chatbot APIs : <https://fastapi.tiangolo.com/>
- Meta AI's Llama 3:  
[https://huggingface.co/docs/transformers/main/model\\_doc/llama3](https://huggingface.co/docs/transformers/main/model_doc/llama3)
- <https://www.kaggle.com/datasets/snehaanbhawal/resume-dataset>
- <https://ai.google.dev/gemini-api/docs/models#gemini-1.5-flash-8b>

# DEMO OF THE APP