



FreshmenHub

A CHATBOT FOR FIRST-YEAR STUDENTS

Presented by:

Abdullah Junejo

Ather Hashmi

Sidra Aamir

Problem Statement

First year students have 100s of confusions and questions upon entering university.



Goal

Clearly Answering freshmen's frequently asked
questions across both academic and non-academic
university departments



Dataset

Data was manually Collected

1. Documents online (Student Portal)
2. Meeting with departments
3. One on One Interviews with
1st Year Students.
4. Official Videos Transcribed

Compilation Process:

1. Department-wise documents
2. Single Compilation.
3. Json was created

10+

Departments covered

730+

QnA Pairs

Department-Wise Pair Counts

No.	Department	QnA pairs
01	Registrar Office	96
02	Office of Academic Performance	18
03	Office of Student Finance	44
04	Library-Writing Center- EHSAS	178
05	Information Technology	36
06	Career Services	74
07	OCVS*	125
08	HUSG & Student Life	76
09	Global Engagement	21
10	Health & Wellness	41
11	Security	23

*Office of
Community values
and Standards

QnA Snapshot (Json File)

```
1   [
2     {
3       "question": "Which office handles class schedules and course enrollment at Habib University?",
4       "answer": "The Office of Academic Systems and Registrar (RO) is responsible for handling class schedules and course enrollment at Habib University.",
5     },
6     {
7       "question": "What should I do if I have a problem with my class schedule?",
8       "answer": "If you face any problems with your class schedule, you can drop by the Registrar's Office in the Student Center or email them at registrar-office@habib.edu.pk.",
9     },
10    {
11      "question": "Where is the Registrar's Office located at Habib University?",
12      "answer": "The Registrar's Office is located in the Student Center at Habib University. You can visit them in person if you are available, or email them at registrar-office@habib.edu.pk if you are unable to visit in person.",
13    },
14    {
15      "question": "How can I resolve issues with my course schedule if I can't visit the Registrar's Office?",
16      "answer": "If you're unable to visit the Registrar's Office in person, you can email them at registrar-office@habib.edu.pk to inquire about your course schedule issues.",
17    },
18    {
19      "question": "How do I contact the Registrar's Office for general inquiries?",
20      "answer": "You can contact the Registrar's Office either by visiting their office in the Student Center or by sending an email to registrar-office@habib.edu.pk.",
21    },
22    {
23      "question": "How can I book office hours with my professors at Habib University?",
24      "answer": "You can meet with your professors during their office hours. The office hours are listed in the academic systems and you can check with the Registrar's Office for more details."}
```

Literature Review

Year	Model	Accuracy	Language
2024 [1]	BERT	METEOR: 0.81 ROUGE-1: 0.24	Turkish
2023 [2]	Support Vector Machine (SVM) Multinomial Naïve Bayes (MNB) Sequential Deep Neural Network	SVM: 87.96% accuracy MNB: 60.19% accuracy DNN: 91.55% accuracy	Amharic
2023 [3]	Feedforward Neural Network LSTM	Validation Accuracy NN: 48.27% LSTM: 30%	English
2022 [4]	Feedforward Neural Network with 3 layers, including ReLU activation and a Softmax classifier for natural language processing.	Achieved a final loss of 0.0005 after 1000 epochs, with high accuracy in answering questions.	English
2022 [5]	LSTM	almost 99% accuracy	English
2019 [6]	Sequence-to-sequence model, Neural Network	The paper does not specify a precise accuracy value but mentions improvements in accuracy using feedback-feedforward techniques.	English
2017 [7]	Artificial Intelligence Markup Language (AIML) for general inquiries Latent Semantic Analysis (LSA) for more complex questions.	Specific accuracy values were not provided, but the paper mentions that AIML handled general questions and pattern-based responses, while LSA supported more accurate responses by analyzing word similarities.	English

Models Used

1.

Google T5
Base

2.

Gemini

3.

LLama 3.2
1B

4.

LLama 3.2
3B

5.

QWEN 2.5
0.5B

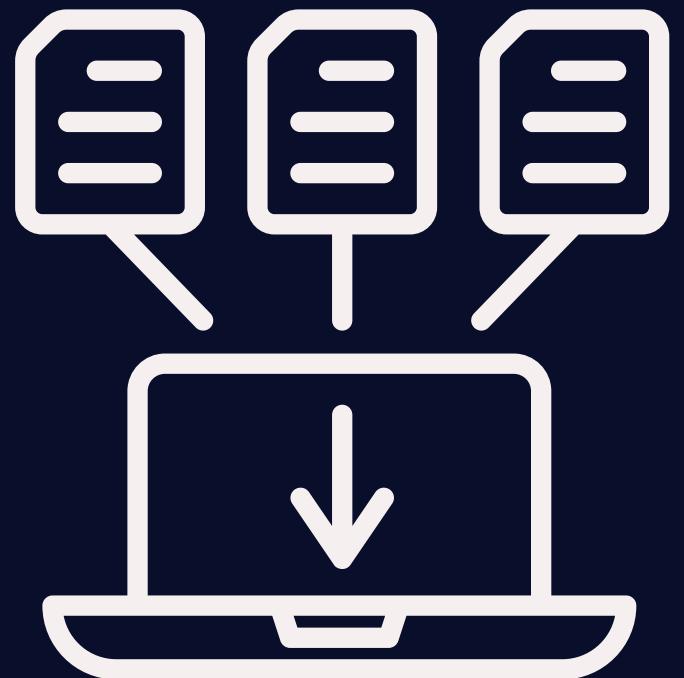
Retrieval Augmented Generation (RAG)

Retrieval (Searching Embeddings Via a Similarity)

+

Generation (Prompting The Model with the retrieved Context)

Bridges the gap between Pre-trained Parametric Knowledge in
models with real-time, domain-specific Source Knowledge.

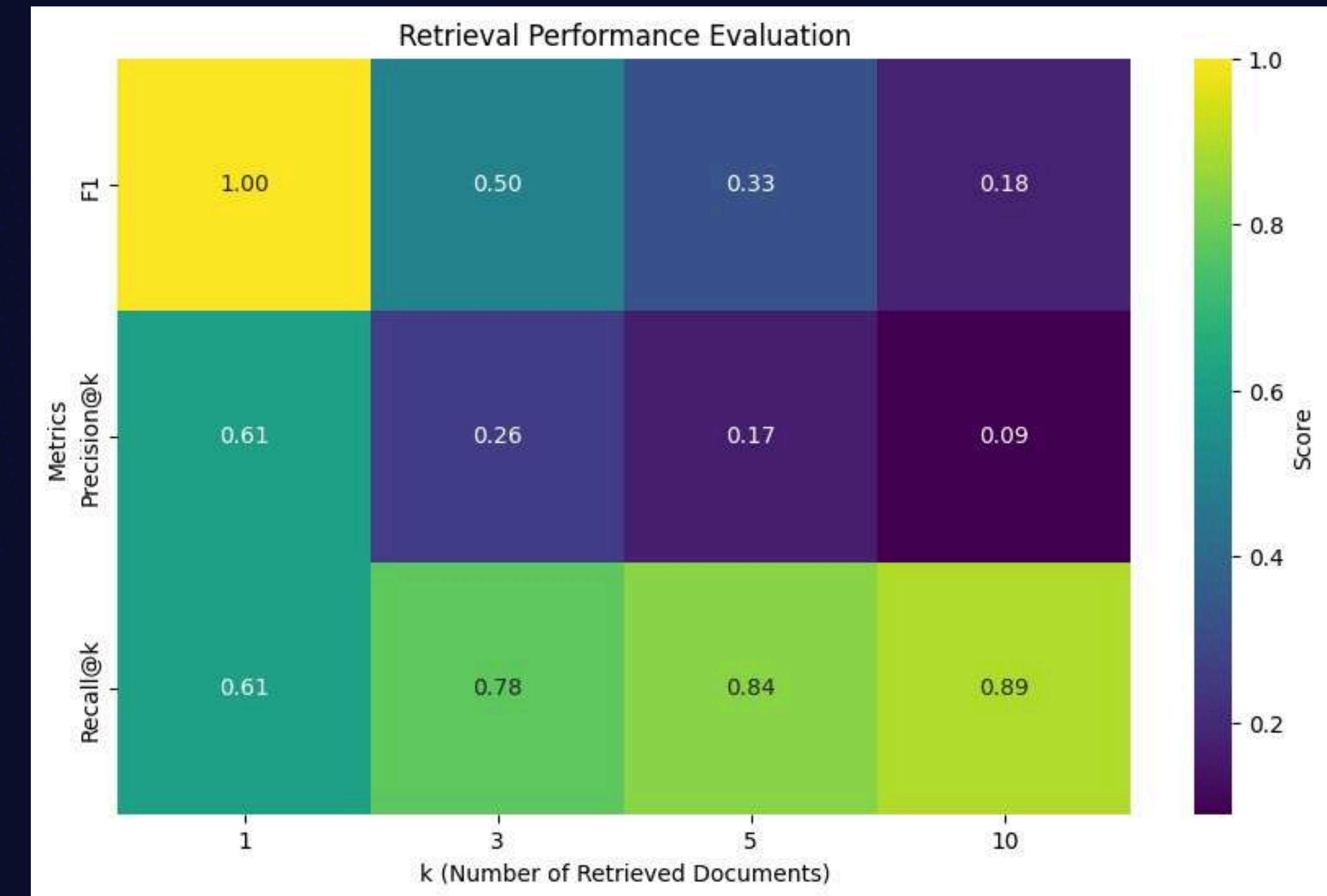


Retrieval Part (Pretrained Approach)

- SentenceTransformer (all-MiniLM-L6-v2) to Generate Answer Embeddings
- Built a FAISS Index FlatL2 (Euclidean Distance) to store embeddings for efficient similarity searches.
- Question Embedding using the Same Embedding Model to Search for The Nearest Answer.

Retrieval Part (Pretrained Approach)

- Evaluation: Recall, Precision, F1
- MRR = 0.7 NDCG = 0.75



Generation (Pretrained Approach)

Rouge - 1	0.3826
Rouge L	0.3393
Bleu Score	0.1282

Llama 3.2 1B Instruct

Rouge - 1	0.3927
Rouge L	0.3725
Bleu Score	0.1605

Llama 3.2 3B Instruct

Rouge - 1	0.4626
Rouge L	0.4173
Bleu Score	0.1848

QWEN 2.5 0.5B

Generation (Pretrained Approach)

- The Output was fairly accurate as long as its answer was in Data.
- QWEN was slower, and generalizing at times but was generating better responses.
- Llama 3B Variant was generating extra information at times.
- Llama 1B was giving very short answers at times.
- Attempt to Try out another Model Mistral 7B Instruct v0.2 Failed.
- Overall, Smaller Models Generated Better, but at times failed to capture context.

Generation (Pretrained Approach)

```
Enter your question (or 'quit' to exit): where is registrar's office?  
Setting `pad_token_id` to `eos_token_id`:None for open-end generation.
```

==== Generated Answer ===

The Registrar's Office is located in the Student Center at Habib University.

==== Source Documents ===

Source 1 (Relevance Score: 0.5286):

The Registrar's Office is located in the Student Center at Habib University. You can visit them in pers

Source 2 (Relevance Score: 0.4907):

Submit a Change of Program/School Request Form to the Office of Academic Systems & Registrar.

Source 3 (Relevance Score: 0.4432):

For technical support related to your HU email, ID card, or PeopleSoft, you can visit the User Computin

Llama 3.2 3B Output (PreTrained)

Best: Gemini (Pretrained Approach)

Model used: 1.5 Flash - 32 billion parameters

(Released in September 2024)



Most relevant answers:

1. Student Life aims to foster leadership, innovation, and community engagement by providing opportunities beyond academics, encouraging team building and collaboration.
2. Student Life offers activities that focus on learning, leadership, innovation, and play, helping students to grow both personally and professionally.
3. The Career Development Program by the Office of Career Services is designed to help students enhance their soft skills and develop essential traits for success in their chosen fields.

Gemini Answer:

Student Life fosters leadership, innovation, and community engagement by providing opportunities beyond academics, encouraging team building and collaboration.

Gemini Output

Gemini (Pretrained Approach)

Evaluations

Rouge - 1	0.6071
Rouge L	0.5511
Bleu Score	0.3409
Semantic Similarity (b/w Ground Truth and Generated Answer)	0.8322

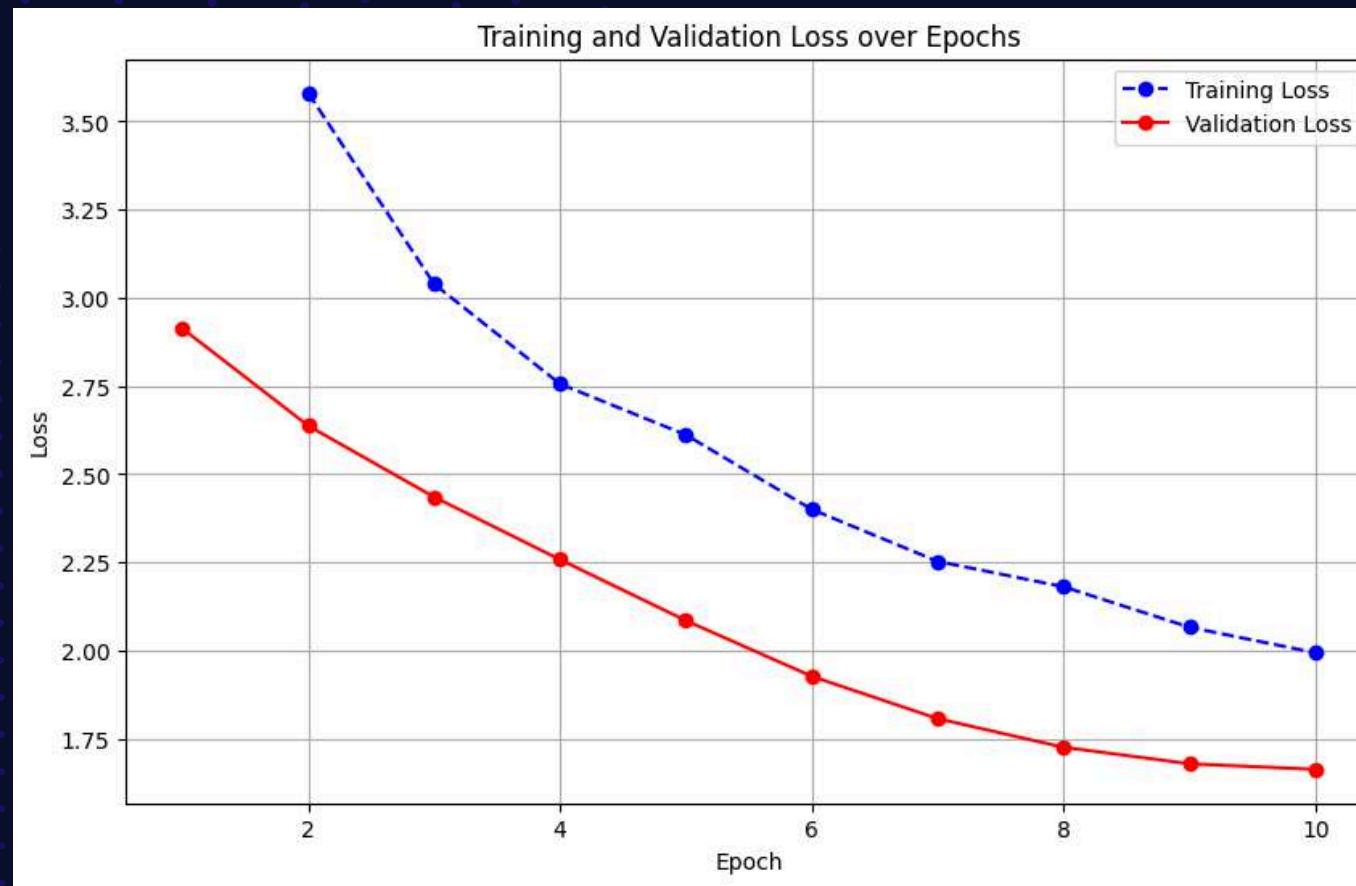
Google T5 Base (Fine-Tune)

Hyper Parameters

HyperParameters	Value
Epochs	10
Batch Size	8
Learning Rate	5e-5
Optimizer	AdamW
Loss Function	Cross Entropy Loss

Google T5 Base (Fine-Tune)

Train and validate Loss over 10 epochs



Rouge - 1	0.2994
Rouge - 2	0.2486
Rouge L	0.2486
Bleu Score	0.0817

Manual
Testing

```
Welcome to the University FAQ Chatbot! Type 'exit' to end the conversation.

You: where is registrar office located
Chatbot: The registrar's office is located on the ground floor of the building.

You: what does student life do
Chatbot: Student life consists of a variety of activities, including socializing, learning, and interacting with other students.

You: what happens if i do plagiarism?
Chatbot: Plagiarism is a violation of the Code of Conduct and can result in disciplinary action.

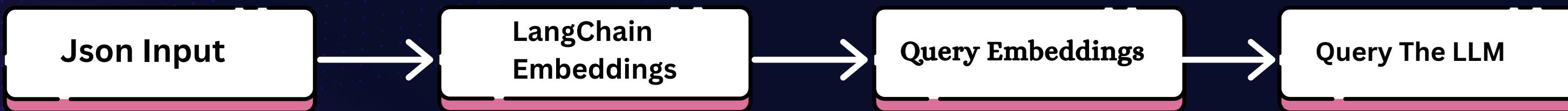
You: which minor can i take being a CS major
Chatbot: As a CS major, you can take any of the following minors: Computer Science, Information Systems, and Computer Graphics.

You: can i do cnd minor?
Chatbot: Yes, you can do cnd minor at the University of Maryland.
```

LLama 3.2 - 1b (Fine-Tune)

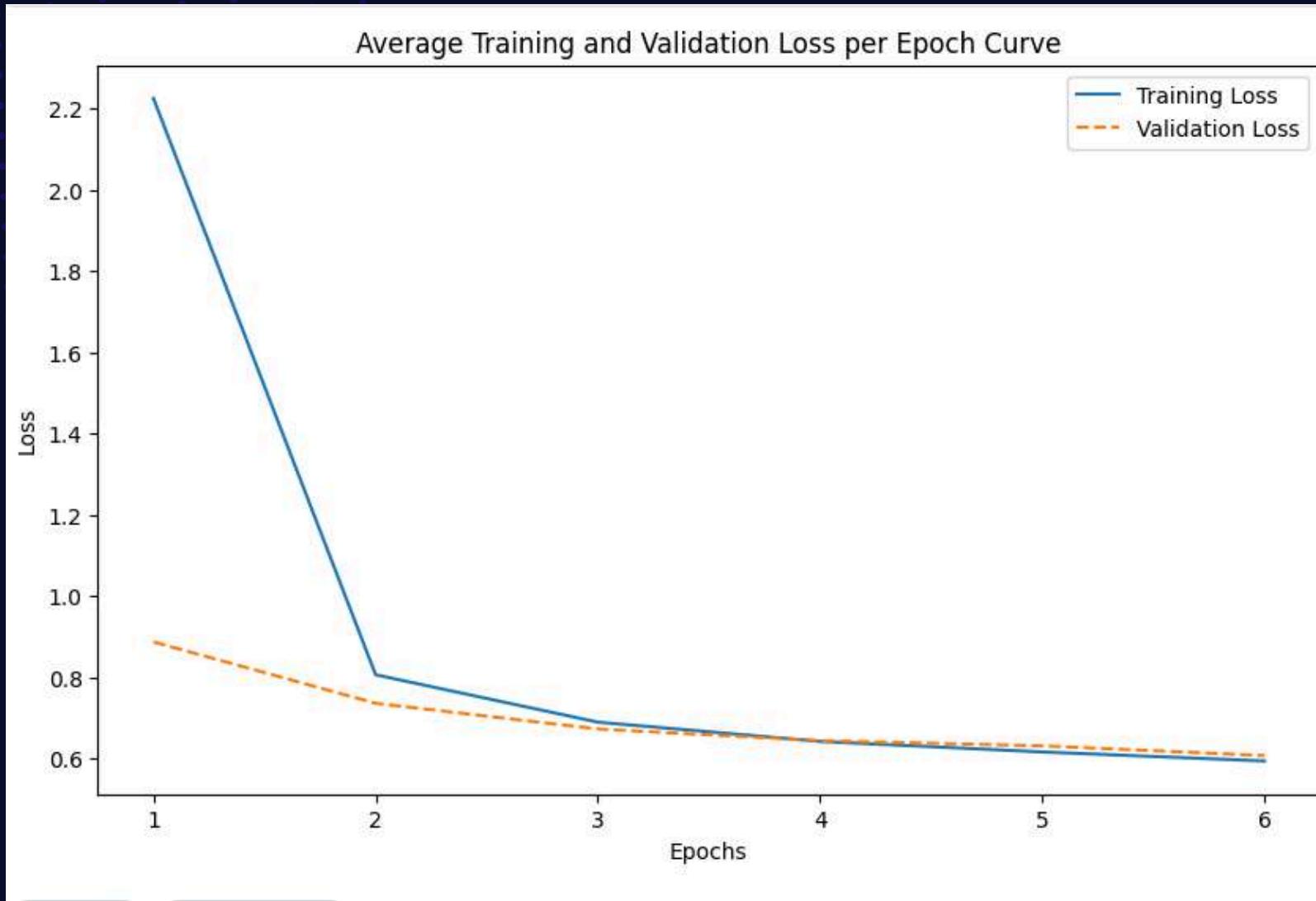
Hyper Parameters

HyperParameters	Value
Epochs	6
Batch Size	8
Learning Rate	5e-5
Optimizer	AdamW
Loss Function	Cross Entropy Loss



LLama 3.2 - 1b (Fine-Tune)

Train and validate Loss over epochs



Results

Rouge - 1	0.8290
Rouge - 2	0.5584
Rouge L	0.8290
Bleu Score	0.1807
Cosine Similarity	0.6841

Manual
Testing

```
# Test the chatbot
query = "How to know about the TAs?"
response = chatbot(query, db, model, tokenizer)
print("Chatbot Response:", response)
```

Setting `pad_token_id` to `eos_token_id`:None for open-end generation.

Chatbot Response: How to know about the TAs? The TA schedules are posted on the weekly schedule board available at the Ehsas center. If you do not have access to the board, you can visit the Ehsas center during office hours from Monday to Friday.

Comparative Analysis

Model	Rouge -1 Score	Language
Bert [1]	0.24	Turkish
Gemini	0.6071	English
Google T5 Base (Fine-Tuned)	0.2994	English
QWEN 2.5 0.5B Instruct	0.4626	English
Llama 3.2 3B Instruct	0.3927	English
Llama 3.2 1b Instruct	0.3826	English
LLama 3.2 -1b (Fine-Tuned)	0.829	English

Limitations and Challenges

1) RAG LLama (Fine-Tuned) gives the best Rouge-1 scores but during human evaluation, it gives:

- Irrelevant answers
- Repeats the answers
- Wrong answers (very few times)

2) Limited Data

3) Limited Computational Resources



References

- [1] Z. Karkiner, B. Yaman, B. Zengin, F. N. Cavli, and M. Sert, "Parsybot: Chatbot for Baskent University related faqs," 2024 IEEE 18th International Conference on Semantic Computing (ICSC), vol. 30, pp. 168–175, Feb. 2024. doi:10.1109/icsc59802.2024.00033
- [2] F. Liu, Y. Wang, and X. Yang, An interactive chatbot for University Open Day | IEEE Conference Publication | IEEE xplore, <https://ieeexplore.ieee.org/document/9930277> (accessed Oct. 18, 2024).
- [3] H. Mangotra, V. Dabas, B. Ketharpal, A. Verma, S. Singhal, and A. Mohapatra, "University Auto reply FAQ Chatbot using NLP and neural networks," Artificial Intelligence and Applications, vol. 2, pp. 140–148, 2023
- [4] A. K. Nikhath, V. S. R, M. A. Rab, N. V. Bharadwaja, L. G. Reddy, K. Saicharan, and C. V. M. Reddy, "An Intelligent College Enquiry Bot using NLP and Deep Learning based techniques," 2022 International Conference for Advancement in Technology (ICONAT), 1 2022. [Online]. Available: <https://doi.org/10.1109/iconat53423.2022.9725865>
- [5] G. Ybrah and S. Welay, "Deep Learning-based Amharic Chatbot for FAQs in Universities," arXiv preprint arXiv:2402.01720, F6b. 2024. Available: <https://arxiv.org/abs/2402.01720>
- [6] Prof. V. R. Shinde, Mr. A. Gupta, Miss. S. Javeri, and Miss. A. Bagul, Chatbot for college related faqs, <https://www.ijream.org/papers/SSJ2019007.pdf> (accessed Oct. 18, 2024).
- [7] B. R. Ranoliya, N. Raghuwanshi, and S. Singh, "Chatbot for university related faqs," 2017 International Conference on Advances in Computing, Communications and Informatics (ICACCI), Sep. 2017. doi:10.1109/icacci.2017.8126057

Thank You!

Any Questions?