CMPT 713 Project: Explore Prototype of Enterprise Domain Question & Answer

Group: ContextCrafters **Hongying Yue, Qin Duan, Hao Sun**

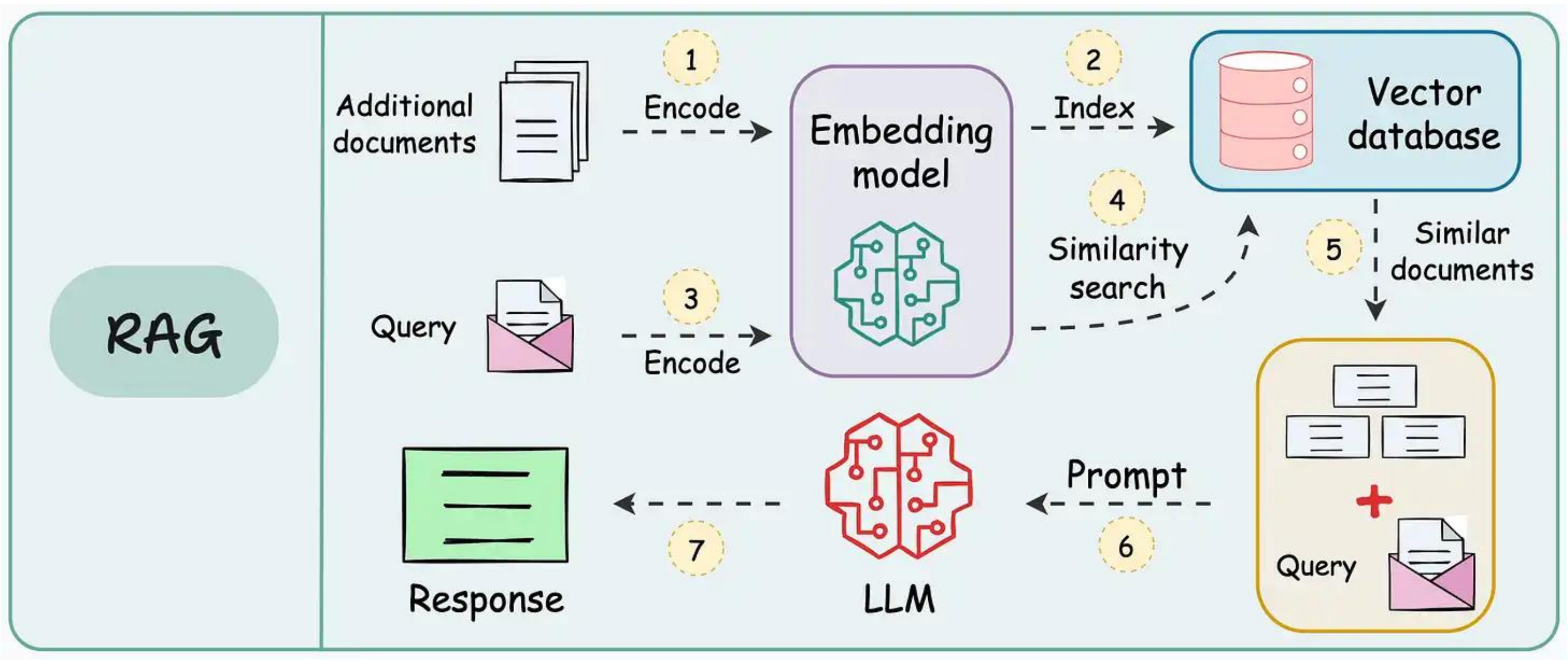
Motivation

- 70% of enterprise internal IT/Finance/HR inquiries are basic, repetitive, and resource-intensive.
- AI QA tools can reduce workload, improve response time, and enhance collaboration efficiency.
- Most inquiries contain enterprise-specific internal information, posing security risks if sent to public tools like Chat-GPT.
- A private-owned AI tool can balance efficiency and information security needs.
- Billion-dollar market potential for customized enterprise AI tools

Compare to Previous Work

- *IBM utilized elastic-search for retrieval part, our pipeline applied embedding in retrieval part
- IBM utilized bert-large-cased for answer extraction, our pipeline applied Reberta-v3, Llama2 and Deepseek API for answer generation

Overall Pipeline



- IBM TECHQA is a highly technical QA dataset; our training and testing are conducted on a smaller subset of TECHQA
- Context length impact answer accuracy
- First part will retrieve top k relevant tech notes
- Second part will generate answer according to relevant tech notes

Retrieve Tech Notes

Methods	Train – rRecall Score	Dev – rRecall Score
bm25-top10	3.6	3.6
bm25-top20	2.3	2.3
embed-top10-0.6	13.0	13.1
embed-top20-0.6	12.0	12.2
embed-top30-0.6	11.6	11.9
embed-top10-0.7	14.8	13.6
embed-top20-0.7	14.5	13.4
embed-top30-0.7	14.5	13.4

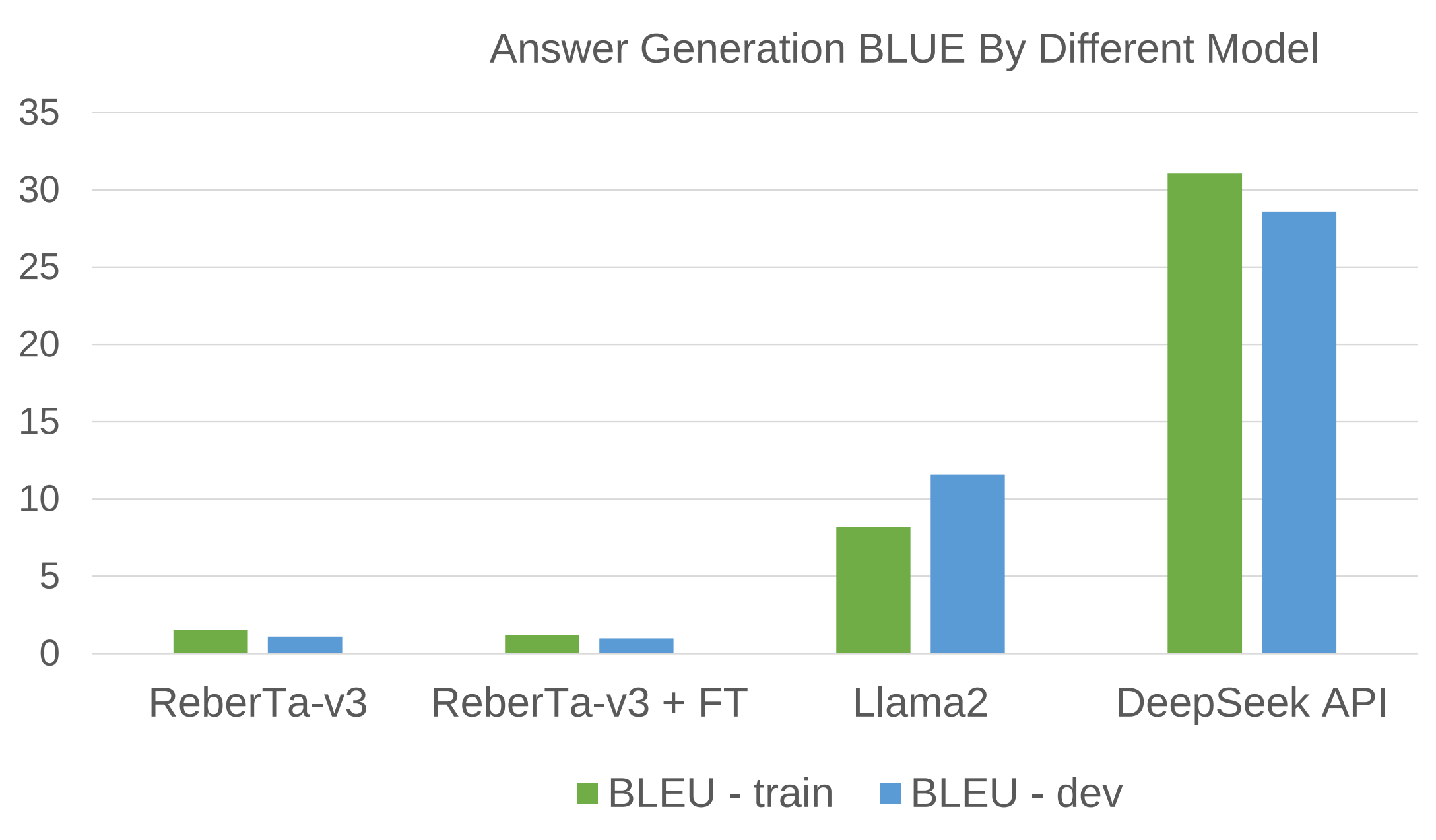
- In retrieving relevant notes, performance of embedding model all-mpnet-base-v2 is much better than bm25, which only match key words

$$rRecallScore = \frac{na\_hit + weighted(ab\_hit)}{na\_total + ab\_total}$$

$$NARecall = \frac{na\_hit}{na\_total}$$

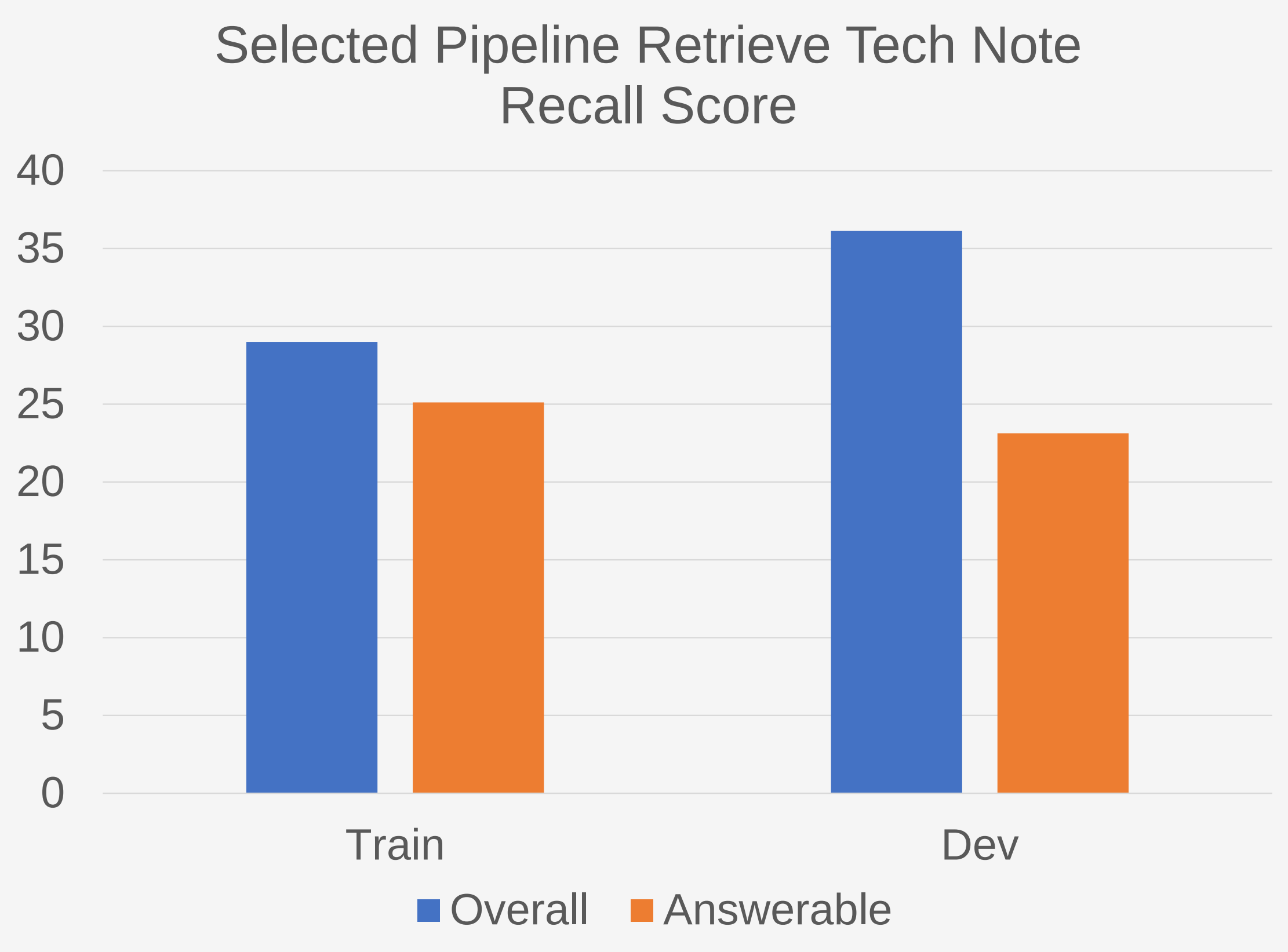
$$rABRecall = \frac{weighted(ab\_hit)}{ab\_total}$$

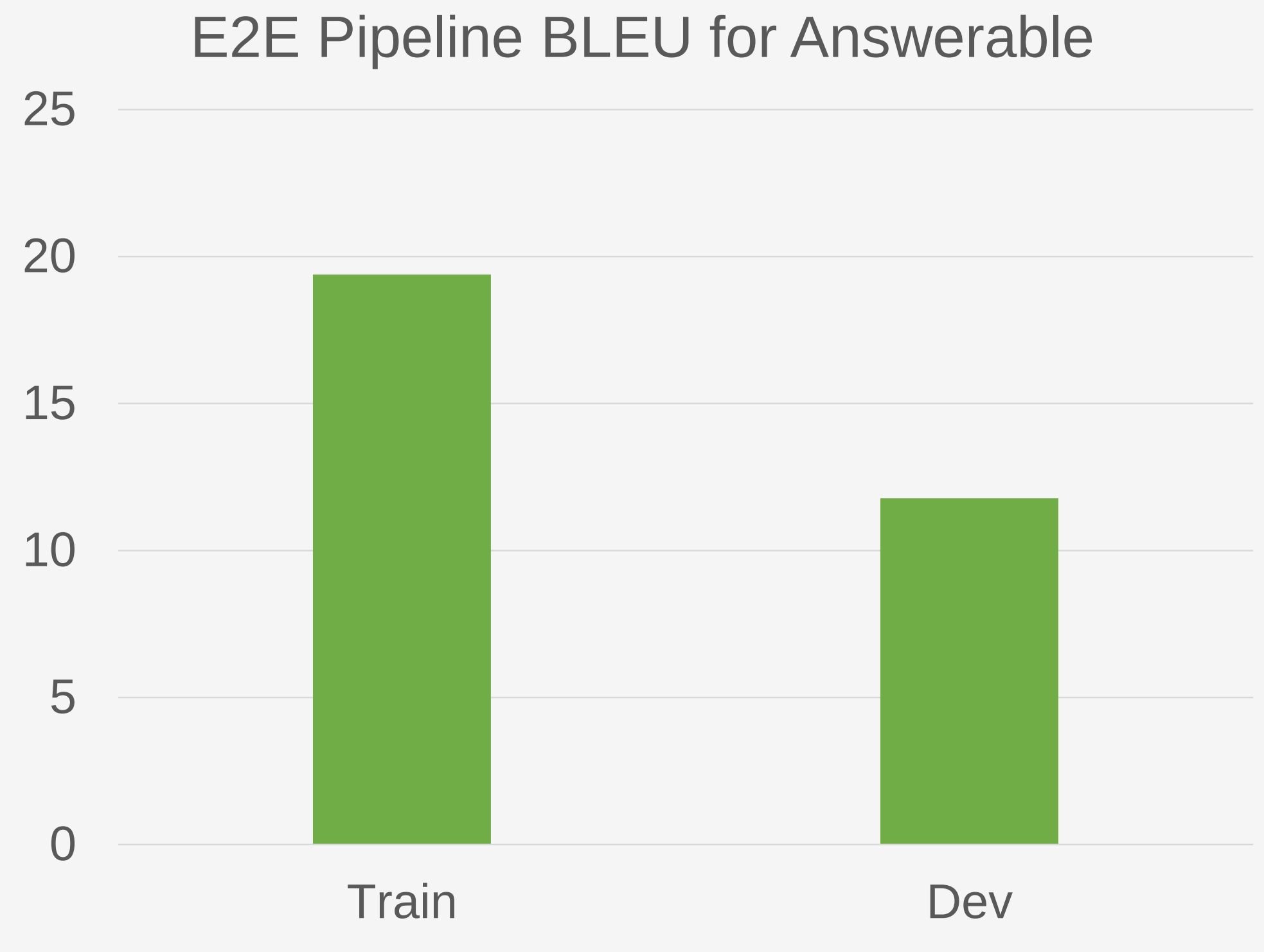
Answer Generation by Diverse Models



- ReberTa-v3 is not ideal for complicated tech QA answer generation
- LLM like Llama2 has potential, still need prompt tuning to improve BLEU
- Latest DeepSeek API perform best, though it is still public tool, applicable for QA domain with low privacy concern

Selected Pipeline Performance



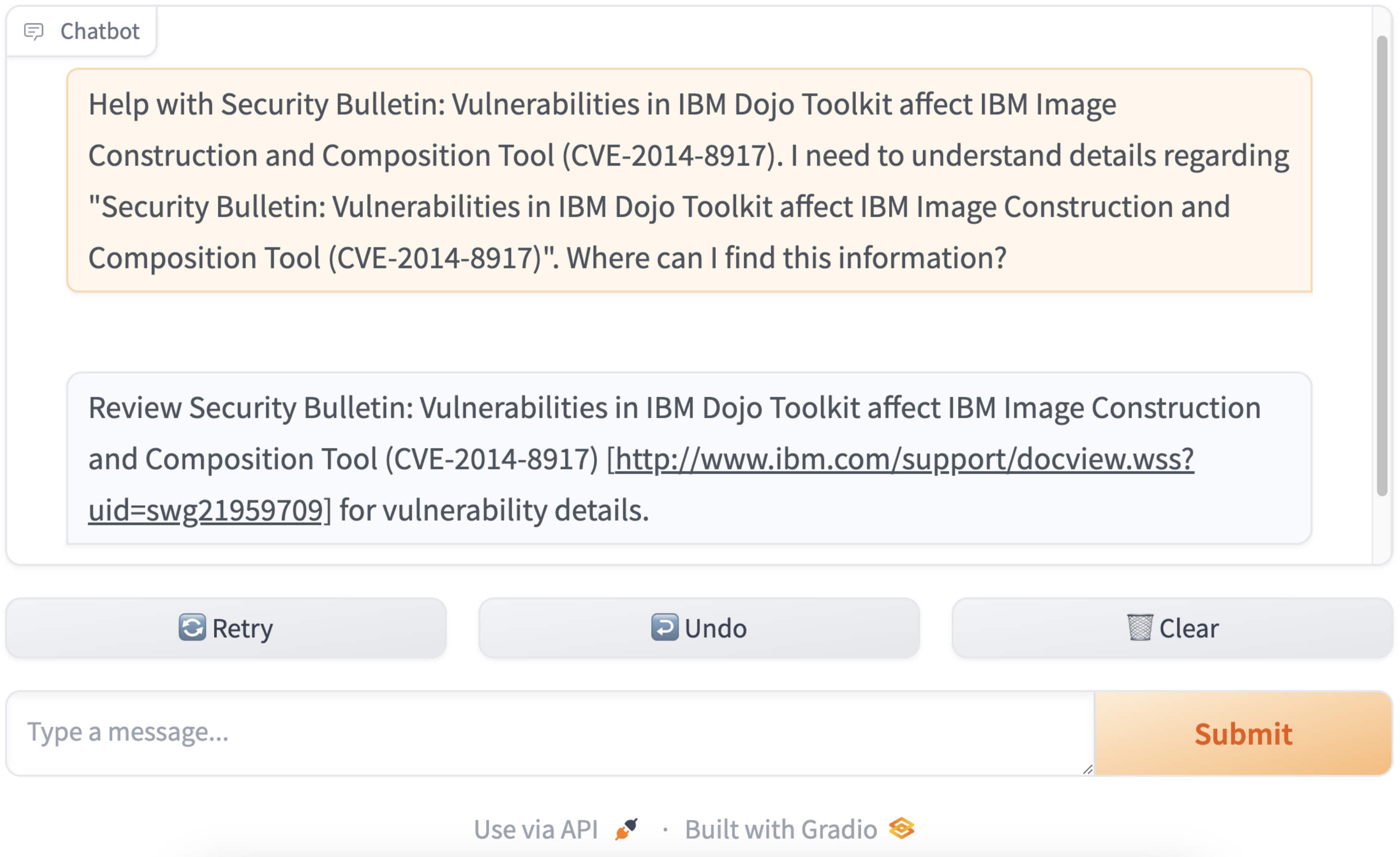


Selected Pipeline

- Model all-mpnet-base-v2 for embedding
- Deepseek API for answer generation
- Generated Answer length is much longer than reference answer

Compare	Train	Dev
Answerable BLEU	19	12
Refer Answer W. Len.	50	43
Generate Answer W. Len.	73	80

Sample & User Interface



Help with Security Bulletin: Vulnerabilities in IBM Dojo Toolkit affect IBM Image Construction and Composition Tool (CVE-2014-8917). I need to understand details regarding "Security Bulletin: Vulnerabilities in IBM Dojo Toolkit affect IBM Image Construction and Composition Tool (CVE-2014-8917)". Where can I find this information?

Review Security Bulletin: Vulnerabilities in IBM Dojo Toolkit affect IBM Image Construction and Composition Tool (CVE-2014-8917) [http://www.ibm.com/support/docview.wss?uid=swg21959709] for vulnerability details.

BLEU: 51.8406

Limitation & Future Work

- The IBM TECHQA dataset's technical Q&A content may have been underrepresented in LLM training, affecting performance.
- Current selected pipeline could be promoted for enterprise QA domain with low information security concern
- Pipeline with Llama2 has potential in domain less complicated than tech QA
- There is potential to develop pipeline tailored for technical Q&A tasks to improve accuracy and relevance.

References:
1.*The TechQA Dataset , Vittorio Castelli, Rishav Chakravarti, Saswati Dana, Anthony Ferritto, etc
2.Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks, Aleksandra P., Fabio P., etc
3.Enhancing Question Answering for Enterprise Knowledge Bases using Large Language Models, Feihu Jiang, Chuan Qin, Kaichun Yao, Chuyu Fang, Fuzhen Zhuang, Hengshu Zhu, Hui Xiong

Reach us ▶

Hongying Yue Qin Duan Hao Sun