

LLM-BASED CHATBOT TO ANSWER QUESTIONS ON DEEP-LEARNING

Sara Pasquino, Valentin Pinon, and Jan Philipp Grgott

March 13th, 2024 | Hands-on Deep Learning

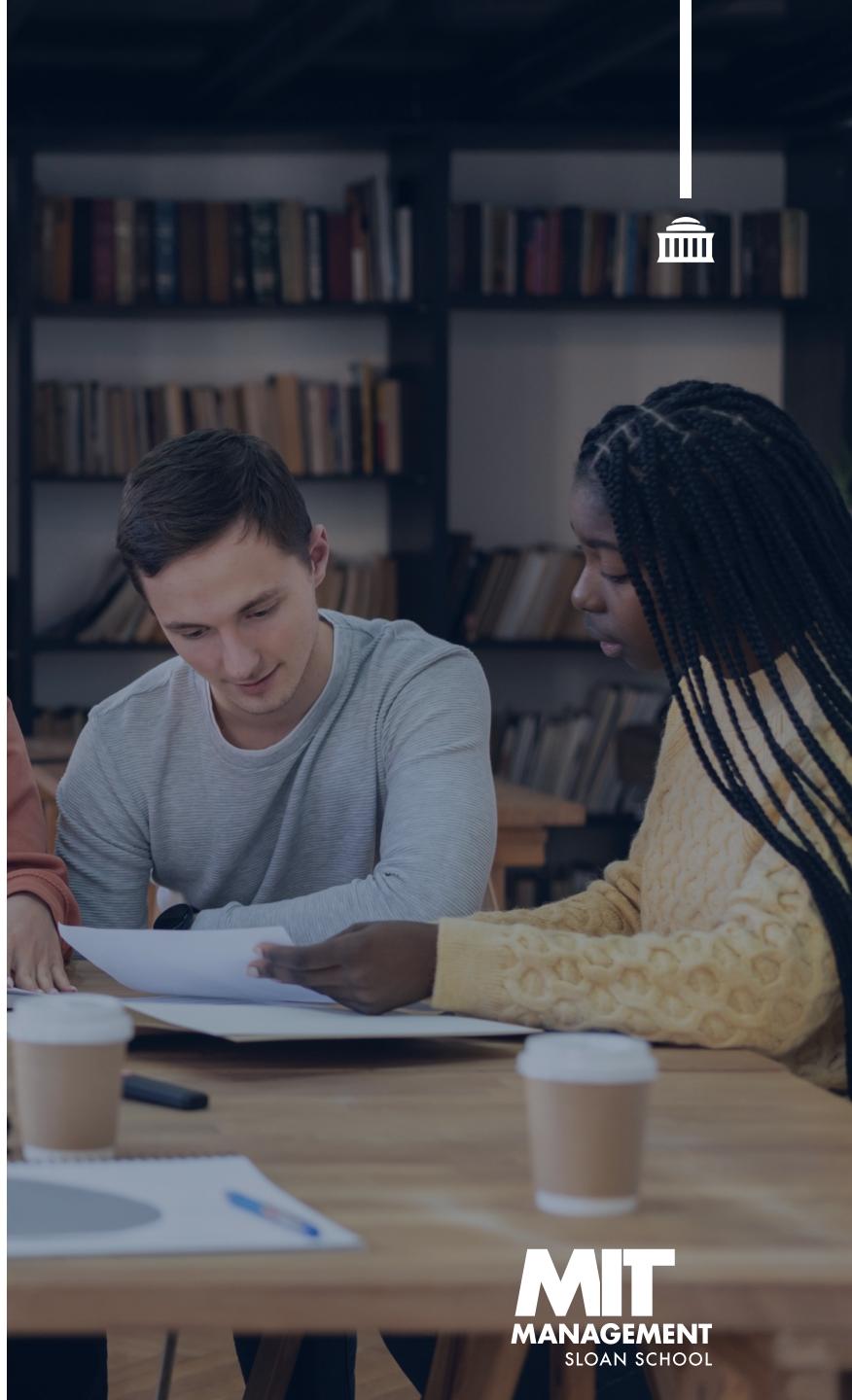
PROBLEM STATEMENT

USING LLMS FOR Q&A IMPROVES LEARNING, SAVES TIME AND COSTS



PROBLEMS

- 1 Students do not get real-time answers to their questions:**
delays in responses by teaching staff can hinder learning progress (e.g., when working on homeworks)
- 2 Answering questions is time-consuming for teaching staff:**
Time saved could be used to improve teaching material, conduct research, or grade assignments
- 3 Maintaining a platform like Piazza is costly for MIT:**
LLMs are available as open-source projects and can be used without recurring costs





WE PRE-PROCESSED TRANSCRIPTS AND THE COURSE BOOK TO MAKE IT MORE READABLE TO THE MODEL

DATA SOURCES



Lectures

All lectures of section A before March 11th



Previewings

Previewings on MNIST, MaxPool, and fine-tuning



Recitations

All recitations held on Zoom



Course Book

„Deep Learning with Python“ by François Chollet



PRE-PROCESSING

1. Normalization

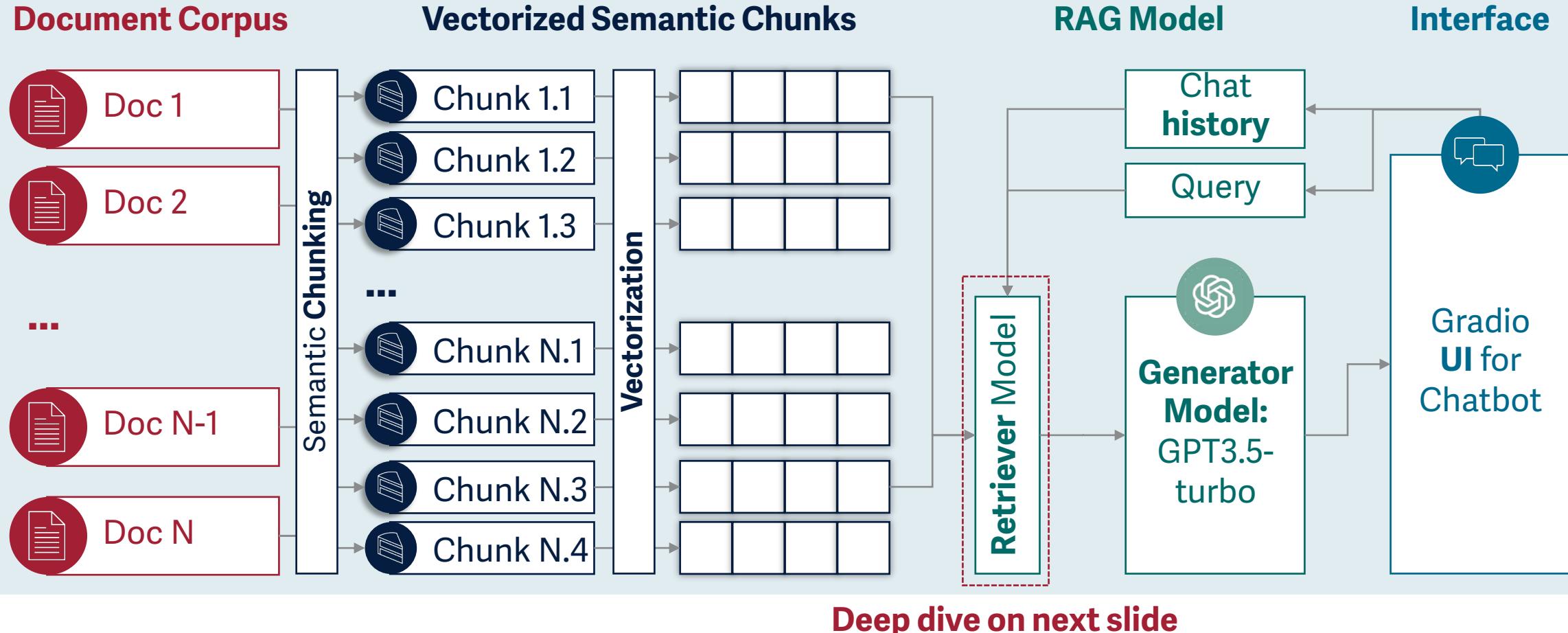
- Remove special characters and non-period punctuation
- Delete headers, footers

2. Data Cleaning

- Strip unnecessary white spaces, new lines and tabs
- Remove HTML tags, in-text citations, hyperlinks, metadata

METHODOLOGY: OVERVIEW

WE ARE COMBINING SEMANTIC CHUNKING, RETRIEVAL AUGMENTED GENERATION AND AN EASY-TO-USE UI



PROMPT ENGINEERING

ACCURATE CONTEXTUAL ANSWERS ARE ACHIEVED THROUGH PROMPT-ENGINEERING



CONTEXTUALIZING PROMPT

"""\Given a **chat history** and the latest user question which might **reference context** in the chat history, **formulate a standalone question** which can be understood without the chat history. **Do NOT answer the question**, just reformulate it if needed and otherwise return it as is."""

Contextualized question

Q&A PROMPT

"""\You are an **assistant for question-answering** tasks for students of a class called **Hands-on Deep Learning**. Use the following pieces of retrieved context to answer the question. If you don't know the answer, just **say that you don't know**. Use **five sentences maximum** and keep the answer **concise**.

{context}"""

PERFORMANCE

THE CHATBOT ACCURATELY ANSWERS QUESTIONS ACCORDING TO ITS KNOWLEDGE



PERFORMANCE ON DIFFERENT QUERY TYPES

1

Queries related to the class content:

- “What’s the professor’s name?” or “Explain backprop”
- Answers correctly; difficulties when asked to name source of information

2

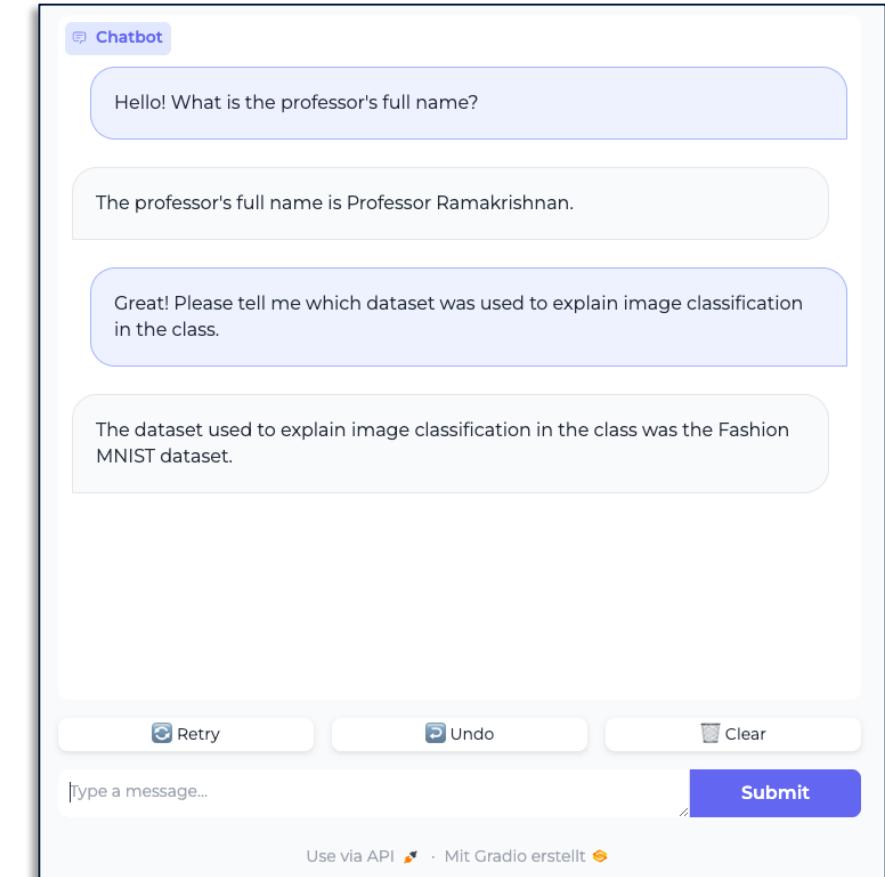
Entirely unrelated questions:

- “What is Jan’s favorite color?”
- Answers with “I don’t know” as instructed

3

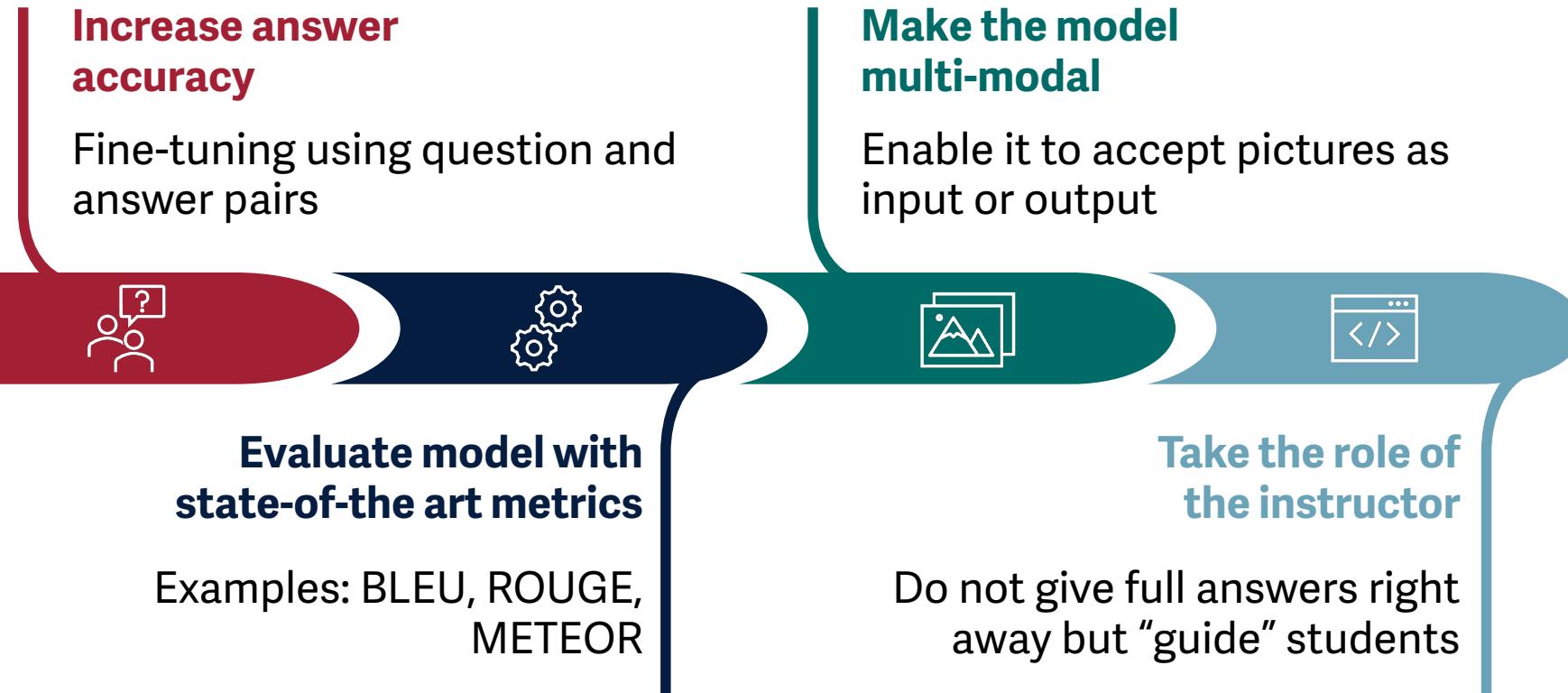
Questions GPT3.5-turbo can answer without RAG:

- “What is a tartiflette?”
- Provides an answer based on previous training, not using provided chunks



POTENTIAL FURTHER NEXT STEPS

FINE-TUNING, EVALUATION METRICS AND MULTI-MODALITY CAN BE WORKED ON



THANK YOU!

Sara Pasquino, Valentin Pinon, and Jan Philipp Grgott

March 13th, 2024 | Hands-on Deep Learning