FREIE UNIVERSITÄT BERLIN

AI week

Engaging Exam Evaluations

Making Use of Learning Materials for LLM-based QA Assessments

Andrea Palmini, Tunc Yilmaz

Freie Universität Berlin, FUB-IT, Abt. Lehre, Studium und Forschung
Arbeitsbereich E-Learning und E-Examinations (EEE)

VERANSTALTET VON:

BHDL BERLINER HOCHSCHULNETZWERK DIGITALE LEHRE

IM RAHMEN EINES PROJEKTES VON:

Netzwerk Landeseinrichtungen für digitale Hochschullehre

GEFÖRDERT VON:

Stiftung Innovation in der Hochschullehre

Engaging Exam Evaluations

Making Use of Learning Materials for LLM-based QA Assessments

Agenda

1 **The IMPACT Project**
Why relevant to essay type exam evaluations?

2 **Large Language Models**
Basic definition, purpose and use cases

3 **Technical Background**
Transformers, vectorstores, RAG…

4 **Concerns**
What are some issues to consider, and how to mitigate them?

5 **Showcase & Discussion**

1 The IMPACT Project

Objective

I **Project as Whole**
AI Based Implementation of Feedback and Assessment with Trusted Learning Analytics

II **Working Package 5**
- How to assist professors evaluate and grade exams
- Summative assessment

Stakeholders

I **Funding**
Bundesministerium für Forschung und Bildung

II **Joint Project**
FUB , GUF, UB, HUB & FeU participate with different roles in different WPs

Project Framework

I **Methodology**
- Ethical principles
- Data protection
- Bias and discrimination

II **NLP & Statistical Models**
How to make use of language models and computational tools to achieve goals

2 Large Language Model (LLM)

What is a large language model?

A large language model is an artificial intelligence algorithm trained on large amount of text data to create a natural language output

- **It uses neural network techniques to process and understand human language**
- **Those techniques are based on the deep learning methodologies, which can detect complex relationships in the text, and also generate text, understanding the semantic and syntactic of a language**

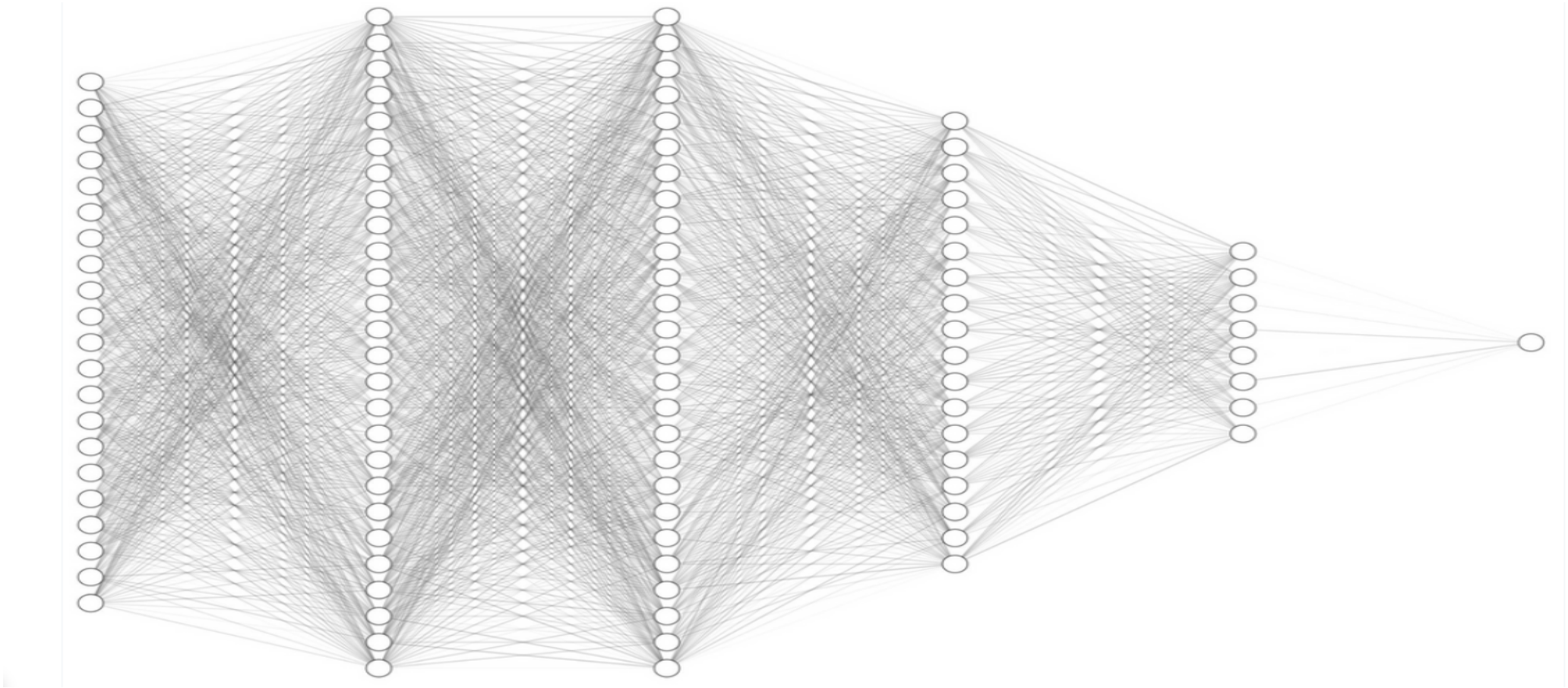
2 Large Language Model (LLM)

How does an LLM work?

- The models are trained on a vast amount of data
- Their utility lies on the ability to recognise patterns and relationships they learn from languages in the training phase
- This ability is given by their structure: consisting in many layers (feed forward, embedding or attention) which collaborate to process a text and generate an output
- The architecture of LLM depends on many factors (computational resource, number of layers, task)
- One of the model that revolutionized NLP tasks is the transformer model

2 Large Language Model (LLM)

LLM Architecture Example



3 Technical Background

Transformer Models

Transformer models were introduced in 2017 achieving best performance in different tasks

Key features of this innovative models are:

Attention Mechanism

- helps to focus on important parts of the input (text)
- allows understanding connections between words or elements far from each other (context understanding)

Parallel Processing

-instead of screening input sequentially, the mechanism is employed on all input to handle larger sequences of text

Encoder-Decoder Architecture

- the encoder process the input with the mechanism
- the decoder generate the output sequence based on the encoder representation of the input

3 Technical Background

Vector Store

A vector store is a database optimized for storing and retrieving high-dimensional vector embeddings, commonly used for similarity searches in AI applications

How it works?

-Text converted into numerical vectors (Embedding generation)
- Vectors stored in a specialized database (storage & indexing)
- After a query, closest matching vectors are retrieved (similarity search)

Use Cases

- Semantic search
- Recommendation systems
- Chatbots with long-term memory

3 Technical Background

Retrieval-Augmented Generation (RAG)

RAG is an AI framework that enhances LLMs by retrieving relevant external information before generating a response, reducing hallucinations and improving factual accuracy

Workflow

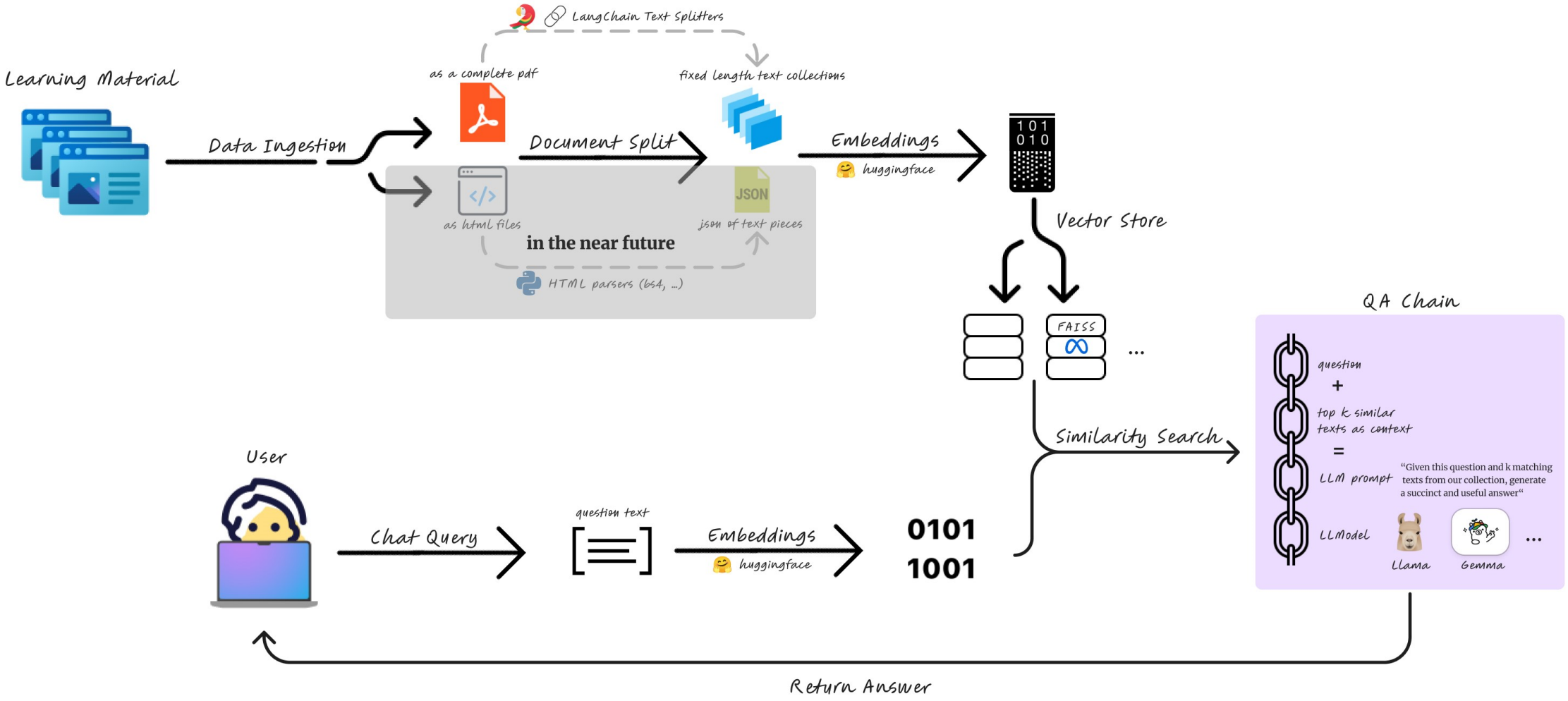
-User query transformed into vector (embedding)
-Retrieve relevant documents from a vector store
-Augment LLM input with retrieved context
-Generate response with more informed answer

Benefits

- Better accuracy incorporating real-time knowledge
-Smaller, efficient model
-Domain specific-knowledge retrieval

3 Technical Background

RAG Behind the Scenes



4 Concerns

Previous Concerns

- Bias and ethics: Data that the model trained on will always reflect the bias

- Privacy and security: Data protection and privacy is paramount in educational institutions

- Computational power and costs: Taking into account growing number of parameters and model sizes

- Hallucinations can be detrimental in the context of learning analytics

- Interpretability: The LLM can output a nice reasoning, but how exactly?

4 Concerns

Mitigation of Concerns

During the project some concerns have arised, the application created is intended to overcome them:

- The official course material of the underlying course can reduce the risk of bias informations retrieved previously from other sources during the model training process
- Stationing databases and models locally helps to secure data privacy
- The feature implemented in the application are saving time and money without undermining the efficiency
- Implementing different features such as RAG or fixed prompting reduces the risk of unwanted answers
- A user-friendly interface has been developed to facilitate output interpretation

Showcase

5 Discussion

Based on your experience in Learning Analytics and Education:

What do you think?

1. **Any other frequent use-cases during evaluations?**
   chat with pdfs, exam generation, …
2. **What are the actual needs of instructors and exam evaluators?**
3. **Control over (hyper)parameters?**
   prompts, more context,…
4. **Concerns over performance, ethics, data protection…**
5. **Other data formats:**
   images, html, etc.

Vielen Dank!

Andrea Palmini
andrea.palmini@fu-berlin.de

Tunc Yilmaz
tunc.yilmaz@fu-berlin.de

Freie Universität
Lehren, Lernen und Prüfen mit Künstlicher Intelligenz
ki@fu-berlin.de

AI Week: Organisatorisches
e-teaching@cedis.fu-berlin.de

AI week

FREIE UNIVERSITÄT BERLIN

AIQualify - Qualifizierung zu Künstlicher Intelligenz in der Hochschullehre

Koordination und Kontakt
Freie Universität Berlin
FUB-IT. Lehre, Studium, Forschung (ex-CeDiS)
E-Learning und E-Examinations (EEE)
Ihnestraße 24, D-14195 Berlin
e-teaching@cedis.fu-berlin.de
https://www.cedis.fu-berlin.de/nel-ki

VERANSTALTET VON:
BHDL BERLINER HOCHSCHULNETZWERK DIGITALE LEHRE

IM RAHMEN EINES PROJEKTES VON:
Netzwerk Landeseinrichtungen für digitale Hochschullehre

GEFÖRDERT VON:
Stiftung Innovation in der Hochschullehre