

# Group presentation: Eduatar – the legal chatbot

Spring Semester  
St.Gallen, 23rd May 2023

## Group A

Kaan Aydin  
Sven Schnydrig



# Background & project description



Many apprentices and their parents have **legal questions regarding apprenticeships**, for instance:

- I am 16 years old. Can I work more than 9 hours daily?
- Can my boss request that I work every Saturday?



To answer those questions, apprentices **reach out to their local secondary education and vocational training office**  
- typically via phone or email



We partnered up with the vocational training office in Bern to build **Eduatar, an intelligent chatbot** designed to assist users in finding information related to apprenticeships in Switzerland



By using LLMs, Eduatar aims to **simplify access to essential information and serve as a reliable resources** for apprentices and parents to navigate the Swiss vocational education system



# Project Approach



**Generate dataset  
of Q&A pairs**

*Factsheets*

*Partner Data*



**Set up pipeline &  
fine-tune**

*BERT*

*GPT3.5/4*

*Vicuna*



**Evaluation of  
models & pipelines**

*Manual*

*GPT4 Evaluation*

# Project approach | Generate dataset of Q&A pairs

## Data collection

Overall, we used two primary sources:

**1st source:** Fact sheets & legal docs manually downloaded from «Berufsbildung»

**2nd source:** Collection of previous Q&As provided by our partner

## Preparation of input to ChatGPT

First, we extracted (relevant) paragraphs from our data

Then, we developed a prompt that did well in generating Q&A pairs and appended to each paragraph

## ChatGPT query & Q&A generation

We queried all paragraphs / prompts in ChatGPT Q&A pairs and saved them in a separate file

Overall, we created a dataset of ~500 Q&A pairs

## Reasons for building a Q&A dataset

Fine-tuning of BERT and Vicuna (tbd) requires Q&A pairs (with context)

Performance evaluation of pipelines (more details later)





# Project approach | Set up pipeline & fine-tune

## Description

## Details on Pipeline setup

## Considerations (i.e., advantages / disadvantages)



### GELECTRA

- Pre-trained BERT-based German language model & finetuned w/ GermanQuAD
- HuggingFace to download pre-trained Gelectra
- Haystack for the reader-retriever pipeline
- Smaller model, not built for Chatbots
- Poor results, requires more fine-tuning with limited upside



### GPT

- We use the OpenAIs GPT 3.5/4 as it is shown to be most effective
- OpenAI's API access to deploy GPT 3.5/4 model
- Use of LangChain for pipeline setup, incl. memory buffering
- Works very well as more chatbot like functionality
- Rather costly



### VICUNA

- LLM based on LLaMa & finetuned on human convos – results close to GPT4
- Download weights and run model with GPU in Colab Pro
- Potential use of LangChain for memory buffering
- Good performance despite being smaller model
- Cheaper vs. GPT, but requires own GPU setup to deploy





Evaluating AI chatbots is **challenging** and requires examining language understanding, reasoning, and context awareness

It still remains an **open question** requiring further research

We will take **two approaches**:



# Project approach | Evaluation of models & pipeline



## Manual Evaluation

- Manually verify if chatbot provides legally correct & pertinent responses to user query



Binary classification (quantitative) of question into good and bad



## GPT-4 Evaluation

- Use GPT-4 with predefined query asking for evaluation of answers by LLMs



Qualitative and quantitative (score) evaluation of answers

*Approach as suggested by the Vicuna team*

# Preliminary results



**Gelectra**



**GPT 3.5**

## **Q1: Ist es möglich, Lernenden wegen Krankheitstagen die Ferienzeit zu kürzen?**

**A:** Ist nicht möglich

**C:** ist nicht möglich, da die Ferien zu Erholungszwecken dienen und nicht „verrechnet“ werden dürfen.

**A:** Nein, es ist nicht möglich, Lernenden wegen Krankheitstagen die Ferienzeit zu kürzen. Im GAV der Polymechaniker sind bei Krankheit zwei Karenztage vorgesehen. Das heißt, die ersten zwei Krankheitstage werden nicht bezahlt. Ein Ferienabzug darf nicht erfolgen.

## **Q2: Wie lange darf man bei der Arbeit Pausen machen?**

**A:** um die Mitte der Arbeitszeit

**C:** die gleichen Pausen wie für Erwachsene: Die Arbeit muss um die Mitte der Arbeitszeit durch Pausen von folgender Mindestdauer unterbrochen werden.

**A:** Laut Arbeitsgesetz muss eine Pause von mindestens einer halben Stunde nach 9 Stunden Arbeit gewährt werden. Am Samstag muss nach spätestens 51/2 Stunden eine Pause von einer Viertelstunde gewährt werden.



University of St.Gallen

23 May 2023

# Current progress...

Generation of Q&A pairs is mostly completed (~90%)

BERT & GPT pipeline have been largely setup (~90%) with minor fine-tuning required

Evaluation started, but at the beginning

## ... and project plan going forward

*Dependent on outcome of Thursday meeting*

Activity	May	Jun
 Finalize Q&A dataset		
 Finalize BERT fine-tuning		
 Implement Vicuna pipeline		
 Evaluate all models		
 Write paper		