21/04/2024

POLITECNICO
MILANO 1863

SCUOLA DI INGEGNERIA INDUSTRIALE
E DELL'INFORMAZIONE

Gabriele Puglisi
Mattia Vicenzotto
Simone Zacchetti
Xin Ye

AI Study

A new way of learning 

Abstract: Discover AI Study, a revolutionary learning tool that uses the Feynman technique to transform education. Improve mastery of any subject by answering questions at different levels of difficulty. AI Study offers multiple modes of interaction, including text and voice interfaces, allowing users to engage with learning in a personalized way.

1. Introduction

The convergence of artificial intelligence (AI) and large language models (LLMs) is transforming traditional learning paradigms in the realm of educational technology (EdTech). Innovative applications driven by AI, such as the AI Study software, are redefining how students engage with educational content.

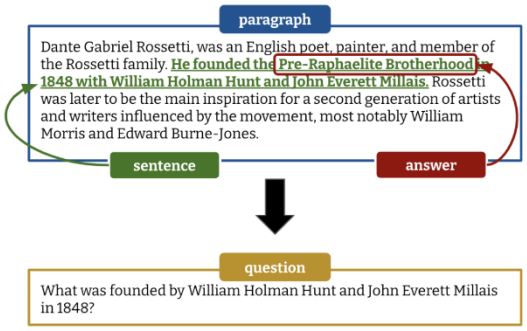
AI Study represents a pioneering effort to enhance the learning process by applying advanced AI techniques inspired by proven educational techniques, such as the Feynman Technique, implemented with a Gamification approach. This software leverages AI-powered question generation and adaptive difficulty levels to promote comprehensive understanding and mastery of subjects. Through multi-modal interaction features including speech recognition and text-to-speech capabilities, AI Study offers users diverse pathways for immersive learning experiences.

This documentation explores the development and functionality of AI Study, highlighting the transformative potential of AI in education while emphasizing the importance of leveraging technology to complement human intelligence and effort. The narrative prompts reflection on the evolving landscape of EdTech and the role of AI-driven innovations in fostering enriched learning environments.

2. Technologies used

The AI Study architecture is built upon a foundation of cutting-edge AI technologies, which are tailored to optimize the learning experience. At its core, AI Study integrates several key components to deliver a comprehensive and interactive educational platform.

1. LMQG-T5 for Question Generation:
    - AI Study utilizes the lmqg-t5 model, a variant of the T5 (Text-To-Text Transfer Transformer) architecture, for generating questions from educational topics. This model excels in understanding and transforming textual inputs, allowing AI Study to generate diverse and contextually relevant questions tailored to specific subject matter.



2. Adaptive Difficulty Levels:
    - The software incorporates adaptive difficulty levels to cater to individual learning needs. By dynamically adjusting the complexity of questions based on user performance, AI Study optimizes engagement and promotes incremental skill development.
3. OpenAI/WHISPER-Small for Speech Recognition:
    - AI Study leverages the openai/whisper-small model for robust speech recognition capabilities. This facilitates hands-free interaction, enabling users to engage with the software via spoken responses and commands.
4. Text-to-Speech (TTS) with SpeechT5_TTS:
    - To enhance accessibility and interactivity, AI Study integrates the SpeechT5_TTS model for text-to-speech functionality. This feature allows the software to verbally communicate questions and instructional content, accommodating diverse learning preferences.
5. Multi-Modal Interaction:
    - AI Study supports various modes of interaction, including text-based input/output, speech recognition, and auditory feedback via text-to-speech. This multi-modal approach caters to different learning styles and accessibility needs, fostering a flexible and inclusive learning environment.

3. Code from the demo and screenshots

To showcase the capabilities and potential impact of AI Study, we developed a demo version using Gradio, a user-friendly platform for creating interactive machine learning demos.

Following are some snippets of code and screenshots of the demo GUI.

```
from transformers import SpeechT5Processor, SpeechT5ForTextToSpeech, SpeechT5HifiGan
from datasets import load_dataset
import torch
import soundfile as sf
from datasets import load_dataset
import IPython.display as ipd

# Initialize processor, model, and vocoder
processor = SpeechT5Processor.from_pretrained("microsoft/speecht5_tts")
model_tts = SpeechT5ForTextToSpeech.from_pretrained("microsoft/speecht5_tts")
vocoder = SpeechT5HifiGan.from_pretrained("microsoft/speecht5_hifigan")

# Load speaker embeddings from a dataset
embeddings_dataset = load_dataset("Matthijs/cmu-arctic-xvectors", split="validation")
speaker_embeddings = torch.tensor(embeddings_dataset[7306]["xvector"]).unsqueeze(0)
```

Text-to-Speech (TTS) with SpeechT5_TTS

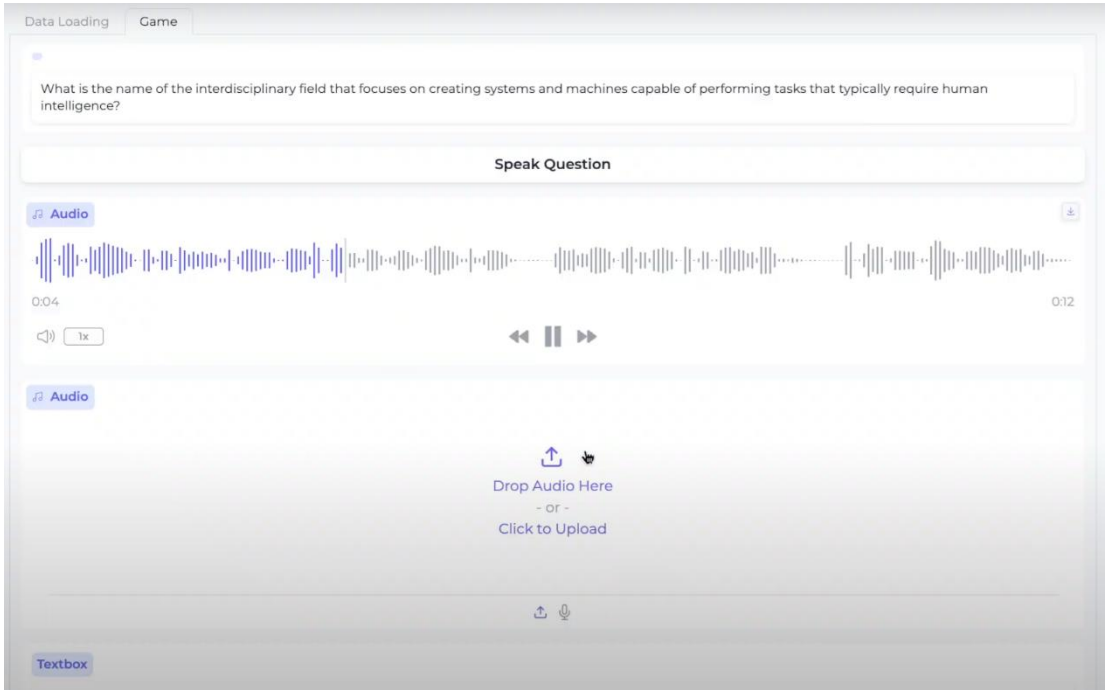
```
with gr.Tab("Data Loading"):
    input_text = gr.Textbox(label = "input")
    btn = gr.Button("Upload text")
    show_status = gr.Textbox(label = "Show status")
    btn.click(fn=load_data, show_progress=True, inputs=input_text, outputs=[show_status])


with gr.Tab("Game"):
    new_question_text = gr.Textbox(label=" ")
    speak_btn = gr.Button("Speak Question")
    speak_btn.click(from_text_to_speech, inputs=[new_question_text], outputs=gr.Audio(type="filepath"))
    audio = gr.Audio(type="filepath")
    text = gr.Textbox()
    submit_btn = gr.Button("Submit Answer")


    with gr.Row():
      score_text = gr.Label(label="Score")
      answer_correct_text = gr.Label(label="Correct Answer")
    submit_btn.click(turn, inputs=[audio, text], outputs=[score_text, answer_correct_text])

    new_question_btn = gr.Button("New Question")
    new_question_btn.click(show_new_question, outputs=new_question_text)
```

GUI components using Gradio



GUI screenshot

4. Limitations of Open Models

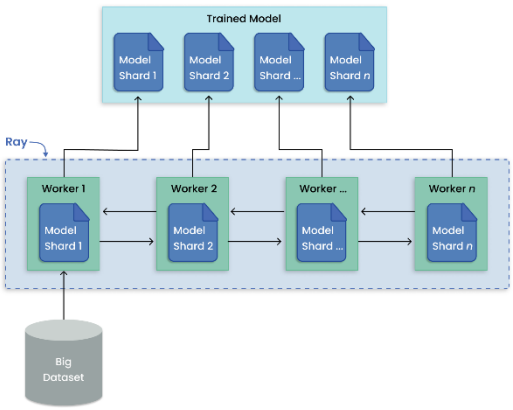
During the development phase, using smaller and less resource-intensive open models has presented certain limitations that must be addressed for a smooth transition to cloud-based production, where scalability and performance are crucial.

- Limited Compute Resources: Open models are smaller and require fewer computational resources, but they may struggle with handling large volumes of data efficiently in a production environment.
- Functional Limitations: Open models may lack certain functionalities needed for a comprehensive and scalable solution. For example, specific pre-trained models or advanced data management tools might be missing.
- Maintenance and Updates: Open models may require more effort for maintenance and updates, as they may not be actively supported by the community or providers.

5. Proposals for Cloud-based Production

Leveraging cloud solutions like RAY or Vertex AI will enhance the scalability and performance of the application, ensuring efficient workload management and increased reliability during production operation.

- **Using RAY for Distributed Model Deployment**:
  RAY provides a scalable infrastructure for running machine learning models in a distributed manner, enabling efficient handling of higher workloads and leveraging available cloud resources effectively.



- **Adopting Vertex AI for Managed ML Model Implementation**:

  Vertex AI offers a managed experience for developing, training, and deploying machine learning models on Google Cloud Platform. Utilizing Vertex AI could be beneficial thanks to advanced autoscaling, monitoring, and model management features in production.

6. Conclusions/final remarks

The development of AI Study using open models and Gradio exemplifies the rapid innovation achievable through AI-driven educational tools. This proof of concept demonstrates the feasibility of personalized, interactive learning experiences facilitated by advanced natural language processing and adaptive learning methodologies.

Transitioning AI Study to production with cloud-based solutions like RAY or Vertex AI will enhance scalability and performance, addressing the limitations of open models. This evolution promises to enhance traditional learning paradigms by empowering learners through tailored question generation and interactive study methods. With further development and refinement, AI Study has the potential to become a transformative tool in education, enriching learning experiences globally. This prototype marks a significant step toward harnessing AI for enhanced studying and knowledge acquisition.