

**BABEŞ-BOLYAI UNIVERSITY, CLUJ-NAPOCA,
ROMANIA**

FACULTY OF MATHEMATICS AND COMPUTER SCIENCE

Face2Learn

– MIRPR Report 2025 –

Team members:

Bucur Victor Sever, Popoviciu Luca, Porcar Cezar, Potra-Rațiu Darius, Preduca Matei

2025

Rezumat

Aplicația reprezintă un sistem inteligent bazat pe inteligență artificială, denumit **Face2Learn**, care combină modele de limbaj mari (LLM), sinteză vocală (TTS) și animație facială pentru a crea o experiență de învățare naturală și interactivă. Scopul proiectului este de a transforma informațiile academice în explicații conversaționale, vizuale și auditive, crescând astfel implicarea, accesibilitatea și retenția informației. Performanța sistemului este evaluată prin acuratețea răspunsurilor, latența end-to-end și naturalețea percepției a utilizatorului.

Cuprins

1 Descrierea problemei rezolvate cu ajutorul AI	2
1.1 Contextul problemei	2
1.2 Scopul și importanța problemei	2
1.3 Utilizatorii sistemului	2
1.4 Datele de intrare și ieșire	3
1.5 Tipurile de date utilizate	3
1.6 Măsurarea performanței sistemului AI	3
2 Related work and useful tools and technologies	4
2.1 1. LoRA (Low-Rank Adaptation)	4
2.2 2. QLoRA	4
2.3 3. llama.cpp	4
2.4 4. TinyLLM (Phi, Mistral, TinyLLaMA)	5
2.5 5. RAGFlow	5
2.6 6. RAG-Anything	5
2.7 7. Whisper TTS	5
2.8 8. Piper	5
2.9 9. Wav2Lip	5
2.10 10. SadTalker	6

Capitolul 1

Descrierea problemei rezolvate cu ajutorul AI

1.1 Contextul problemei

Sistemele educaționale tradiționale se bazează predominant pe text și prelegeri statice. În contextul actual al digitalizării, apare nevoia de metode de predare interactive, personalizate și accesibile. Proiectul **Face2Learn** propune utilizarea inteligenței artificiale multimodale pentru a crea un asistent virtual care explică concepte academice prin vorbire și expresii faciale sincronizate.

Modelul AI poate înțelege întrebări formulate în limbaj natural, poate accesa informații relevante din materiale educaționale și poate furniza răspunsuri clare, exprimate vocal și vizual printr-un avatar animat.

1.2 Scopul și importanța problemei

Scopul principal este de a transforma procesul de învățare într-o experiență naturală, captivantă și accesibilă. Importanța proiectului derivă din:

- creșterea **engagement-ului** și motivației elevilor/studenților;
- asigurarea **accesibilității** pentru persoane cu deficiențe de vedere sau auz;
- sprijinirea cadrelor didactice prin automatizarea explicațiilor și a sesiunilor de Q&A;
- posibilitatea **învățării personalizate** în ritmul fiecărui utilizator.

1.3 Utilizatorii sistemului

- **Studenti și elevi** – folosesc avatarul AI pentru explicații și recapitulări;
- **Profesori și tutori** – utilizează sistemul pentru demonstrații și asistență automată;

- **Instituții educaționale** – integrează soluția pentru suport didactic 24/7;
- **Persoane cu dizabilități** – beneficiază de conținut multimodal adaptat.

1.4 Datele de intrare și ieșire

Date de intrare:

- întrebări formulate în limbaj natural (text sau voce);
- documente educaționale (manuale, cursuri, notițe);
- preferințe ale utilizatorului (limbă, voce, tonalitate).

Date de ieșire:

- răspuns text generat de LLM;
- voce sintetică naturală generată prin TTS;
- videoclip animat cu avatar sincronizat cu vorbirea.

1.5 Tipurile de date utilizate

- corporuri textuale academice și explicații didactice;
- înregistrări audio pentru antrenarea TTS;
- imagini/video cu expresii faciale pentru sincronizare (lip-sync);
- embeddings semantice pentru căutare contextuală (RAG).

1.6 Măsurarea performanței sistemului AI

Performanța sistemului este evaluată prin indicatori cantitativi și calitativi:

- **Acuratețea răspunsurilor** – procentul de răspunsuri corecte;
- **Timpul mediu de răspuns** – durata procesării end-to-end;
- **Calitatea animației** – gradul de sincronizare buze-vorbire;
- **Consum de resurse** – memorie și timp de inferență.

Capitolul 2

Related work and useful tools and technologies

Această secțiune prezintă zece proiecte și tehnologii relevante pentru construcția unui avatar AI educațional. Pentru fiecare sunt menționate tipul datelor folosite, algoritmii utilizati, performanțele și tehnologiile implicate.

2.1 1. LoRA (Low-Rank Adaptation)

Date: text educațional.

Algoritmi: fine-tuning eficient al LLM-urilor prin adaptare low-rank.

Performanță: îmbunătățire a acurateței cu cost redus.

Tehnologii: PyTorch, Hugging Face, GitHub open-source.

2.2 2. QLoRA

Date: corpus text.

Algoritmi: fine-tuning cu cuantizare pentru reducerea memoriei.

Performanță: menține calitatea modelului la 4-bit.

Tehnologii: Transformers, bitsandbytes, Hugging Face.

2.3 3. llama.cpp

Date: text.

Algoritmi: inferență locală pentru modele cuantizate.

Performanță: latență redusă pe CPU.

Tehnologii: C++, GGUF models, GitHub.

2.4 4. TinyLLM (Phi, Mistral, TinyLLaMA)

Date: text.

Algoritmi: modele compacte pentru rulare eficientă.

Performanță: raport bun între viteză și acuratețe.

Tehnologii: PyTorch, Hugging Face.

2.5 5. RAGFlow

Date: documente și note de curs.

Algoritmi: RAG (retrieval augmented generation).

Performanță: răspunsuri mai relevante.

Tehnologii: LangChain, FAISS, GitHub.

2.6 6. RAG-Anything

Date: fișiere locale (PDF, text).

Algoritmi: flux simplificat RAG.

Performanță: acces rapid la surse externe.

Tehnologii: Python, Streamlit, GitHub.

2.7 7. Whisper TTS

Date: corpusuri audio și transcriptii text.

Algoritmi: model neural de sinteză vocală bazat pe arhitectura Whisper.

Performanță: voce naturală și suport multilingv.

Tehnologii: PyTorch, Whisper TTS API (OpenAI), Hugging Face, GitHub.

2.8 8. Piper

Date: audio/text.

Algoritmi: TTS optimizat pentru dispozitive edge.

Performanță: latență foarte mică.

Tehnologii: Rust, on-device inference.

2.9 9. Wav2Lip

Date: video + audio.

Algoritmi: lip-sync bazat pe rețele CNN.

Performanță: aliniere buze-vorbire realistă.

Tehnologii: PyTorch, OpenCV, GitHub.

2.10 10. SadTalker

Date: imagine + audio.

Algoritmi: talking-face generation dintr-o singură imagine.

Performanță: expresii faciale naturale.

Tehnologii: PyTorch, DeepFace, GitHub.

Concluzii

Proiectul „Face2Learn” oferă o abordare inovatoare de integrare a AI multimodal (text, voce, video) în domeniul educației. Combinarea între LLM-uri optimizate, TTS și animație sincronizată contribuie la îmbunătățirea experienței de învățare, făcând-o mai naturală, mai interactivă și mai accesibilă.