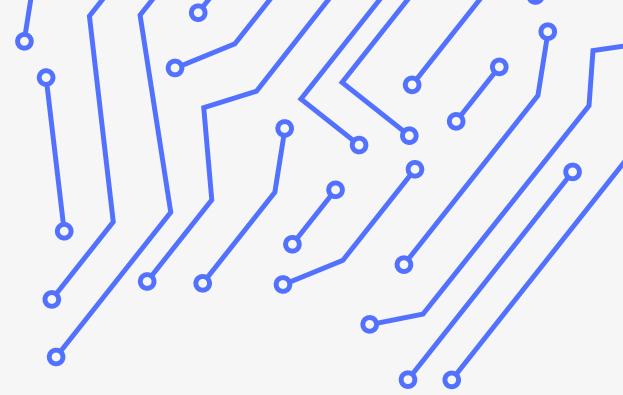


AVATAR CHATBOT

ENHANCING KNOWLEDGE AND EXPRESSION WITH AI AVATARS

Proposed By:
Prof Dr. Laura Dioşan &
Alexandru Manole



“WHY DO AI AVATARS MATTER?”

AI that turns knowledge into conversation

1. Education & Training

- Explain complex concepts in a more engaging way
- Provide interactive tutors that personalise learning.

2. Accessibility

- Turn text into speech and lip-synced video for people support learners with diverse needs and preferences
- Enable multilingual support by instantly translating and “speaking” in different languages.

3. Engagement & Communication

- Make information delivery more human-like and memorable.
- Keep users’ attention longer compared to plain text or audio.
- Adapt to user preferences, tone, and style.



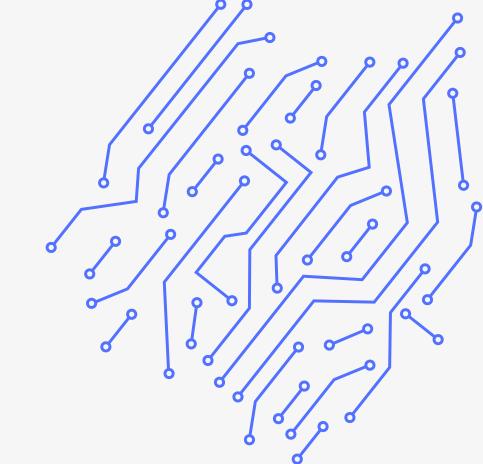
PROPOSED USE-CASE:

The task is to create an AI avatar that supports the university curriculum by answering questions and explaining concepts in an engaging way. The avatar should combine multiple AI systems. You can choose which part of the system gets the focus of your work or try to improve each component



PROPOSED FLOW:

- **Build the brain:** Fine-tune a small LLM or use RAG for subject knowledge.
- **Voice the answer:** Convert text to speech.
- **Animate the face:** Create a lip-synced avatar video (pre-rendered or real-time).
- **Integrate & evaluate:** Connect all parts and test for speed, accuracy, and realism.





RESEARCH DIRECTIONS

* LLM ENHACEMENTS

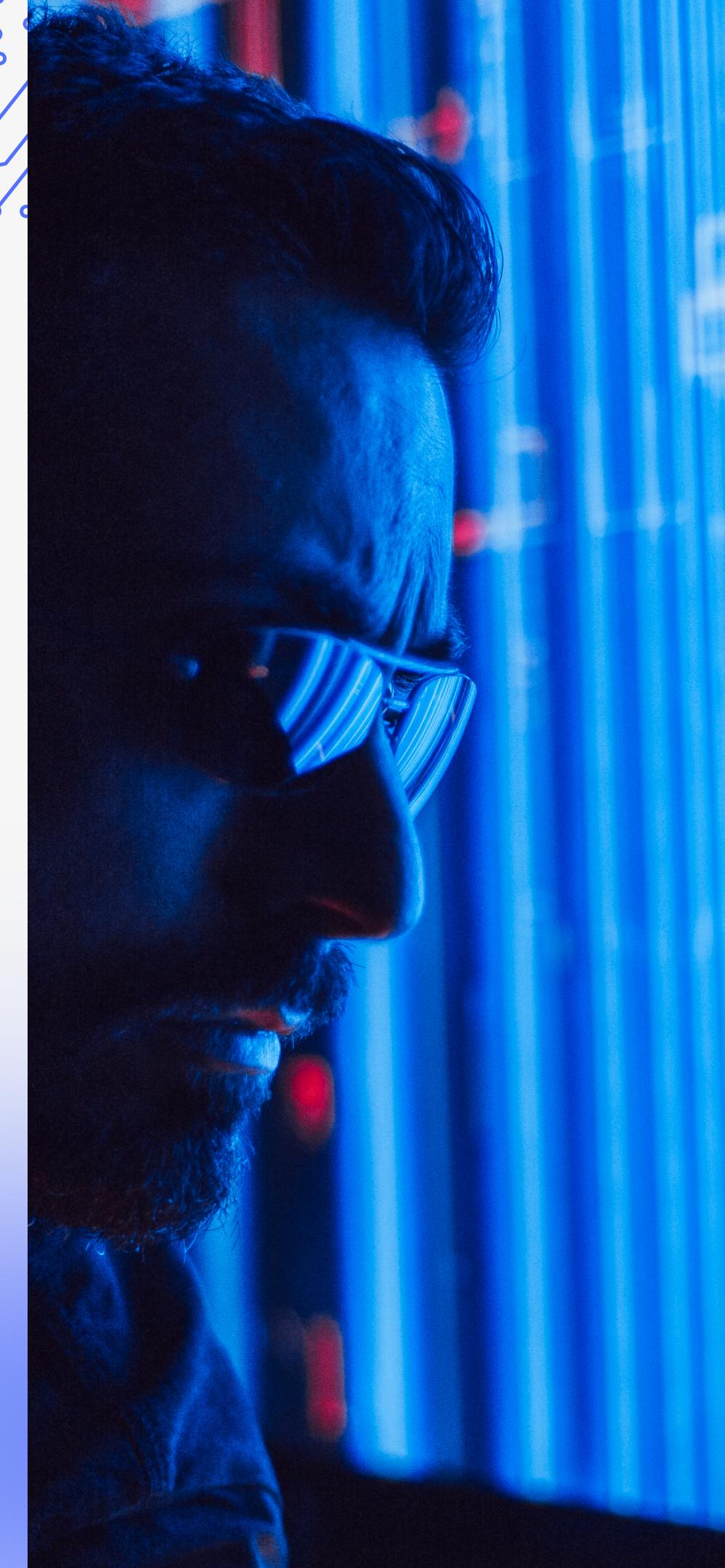
- Fine-tuning with **LoRA/QLoRA** for domain knowledge.
- Using **small, efficient models** (Phi, Mistral, TinyLLaMA) for consumer GPUs.
- **RAG** pipelines to connect the avatar with external knowledge sources.

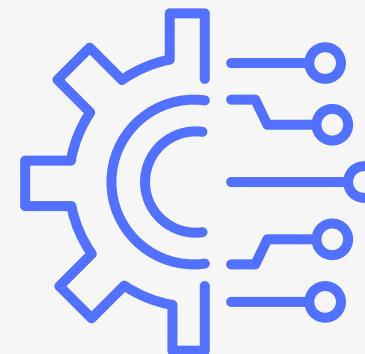
* AVATAR VOICE

- Text-to-Speech (**TTS**) models → open-source tools (Coqui TTS, VITS, Piper); focus on low-latency and multi-language support.
- Prosody & **emotion control** → Use models/datasets with emotional TTS
- **Voice adaptation** & cloning → Apply voice conversion (RVC, So-VITS-SVC) or speaker embedding methods

* AVATAR ANIMATION

- **Pre-rendered avatars** → Wav2Lip, SadTalker, AnimateDiff
- **Real-time avatars** → DeepFaceLive, RVC
- **End-to-End** → OpenTalker





RESEARCH DIRECTIONS

BETTER

Accuracy is key for both text generation and speech. Reliable answers make the avatar trustworthy and useful. Clear, natural speech ensures correct delivery and tone. Together they create a believable, human-like experience.

FASTER

Speed, latency, and memory are vital because they shape how natural and practical the system feels. High latency breaks immersion in real time conversations, while heavy memory use makes models hard to run on hardware.

HUMAN

Being human-like makes the avatar more relatable and engaging. Natural voices and expressions build trust and comfort. Human qualities improve learning, support, and interaction. They transform AI into a more approachable presence.

CREATIVE

Be creative because there's always more to explore in how avatars can grow and improve. Find your own direction which would allow Avatar to be easier to implement.



LLM ENCHACEMENTS

Fine-tuning small models with LoRA/QLoRA

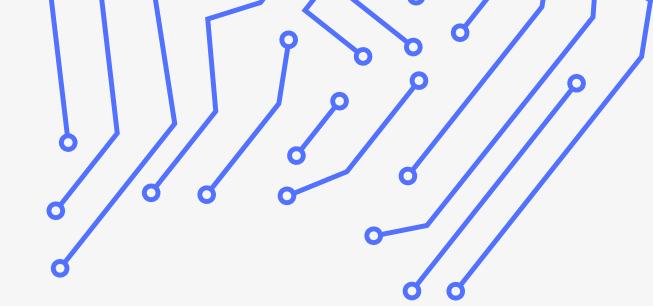
- <https://github.com/cloneofsimo/lora>
- <https://github.com/artidoro/qlora>
- https://github.com/georgesung/llm_qlora
- <https://github.com/michaelnnny/QLoRA-LLM>

RAG pipelines

- https://huggingface.co/docs/transformers/model_doc/rag
- <https://github.com/infiniFlow/ragflow>
- <https://github.com/mrdbourke/simple-local-rag>
- <https://github.com/HKUDS/RAG-Anything>
- <https://github.com/Cinnamon/kotaemon>

Running small models locally

- <https://github.com/skyzh/tiny-llm>
- <https://github.com/ggml-org/llama.cpp>
- <https://huggingface.co/blog/jjokah/small-language-model>



AVATAR VOICE

Open-source Text-to-Speech

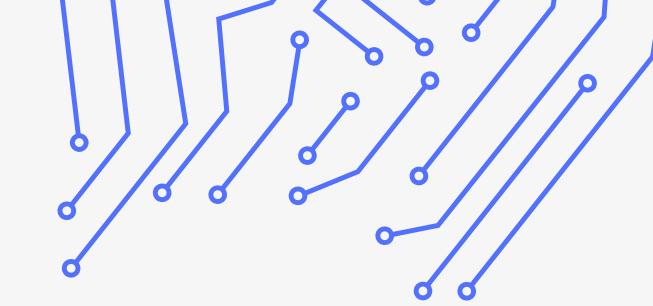
- <https://github.com/coqui-ai/TTS>
- <https://github.com/jaywalnut310/vits>
- <https://github.com/rhasspy/piper>

Voice conversion

- <https://github.com/RVC-Project/Retrieval-based-Voice-Conversion-WebUI>
- <https://github.com/svc-develop-team/so-vits-svc>

Emotion / prosody modeling

- <https://github.com/luanqz/styletts2>
- <https://github.com/suno-ai/bark>



AVATAR ANIMATION

Pre-rendered animation

- <https://github.com/Rudrabha/Wav2Lip>
[raw.githubusercontent.com](https://raw.githubusercontent.com/Rudrabha/Wav2Lip/main/README.md)
- <https://github.com/OpenTalker/SadTalker> [github.com](https://github.com/OpenTalker/SadTalker)
- <https://github.com/guoyww/AnimateDiff> [github.com](https://github.com/guoyww/AnimateDiff)

Real-time animation

- <https://github.com/iperov/DeepFaceLive>
- <https://github.com/RVC-Project/Retrieval-based-Voice-Conversion-WebUI>
- <https://github.com/FaceFusion/FaceFusion>

End-to-end / research pipelines

- <https://github.com/ZeroDes1/OpenTalker>



**THANK
YOU**

