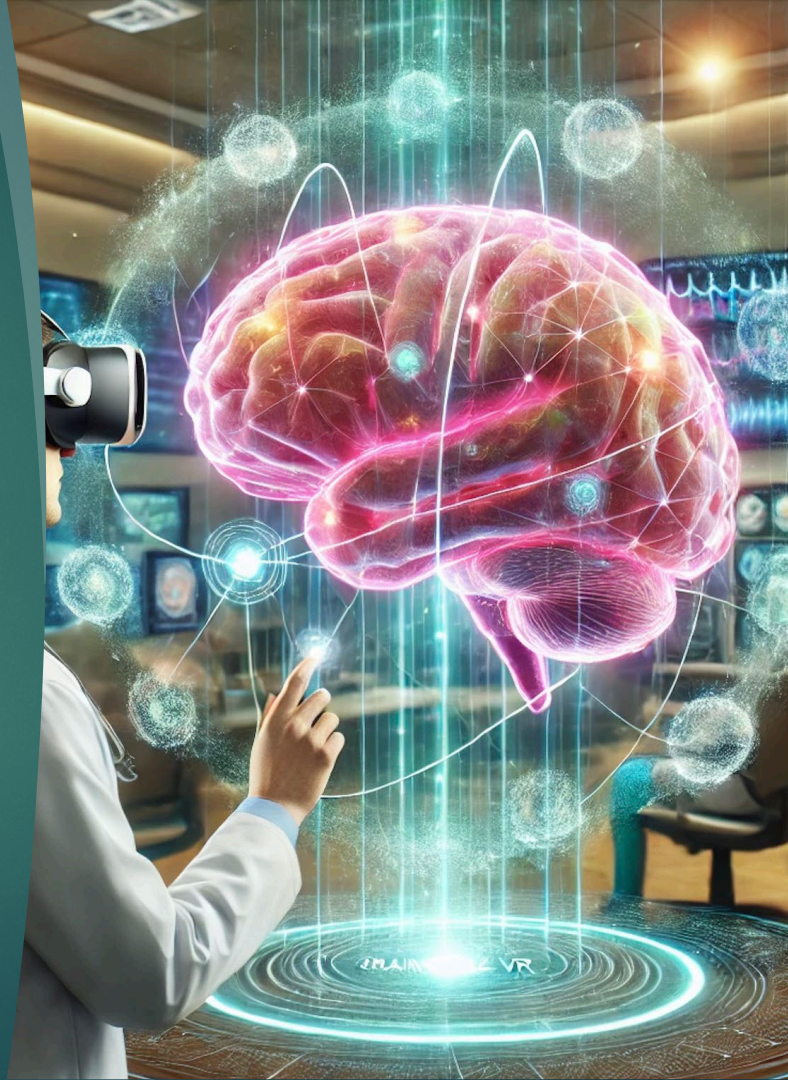
VR-Enabled Foundational Vision Language Models for Medical Training

Akhil Kasturi & Adma Gama-Krummel
Advisor: Mudjat Cetin

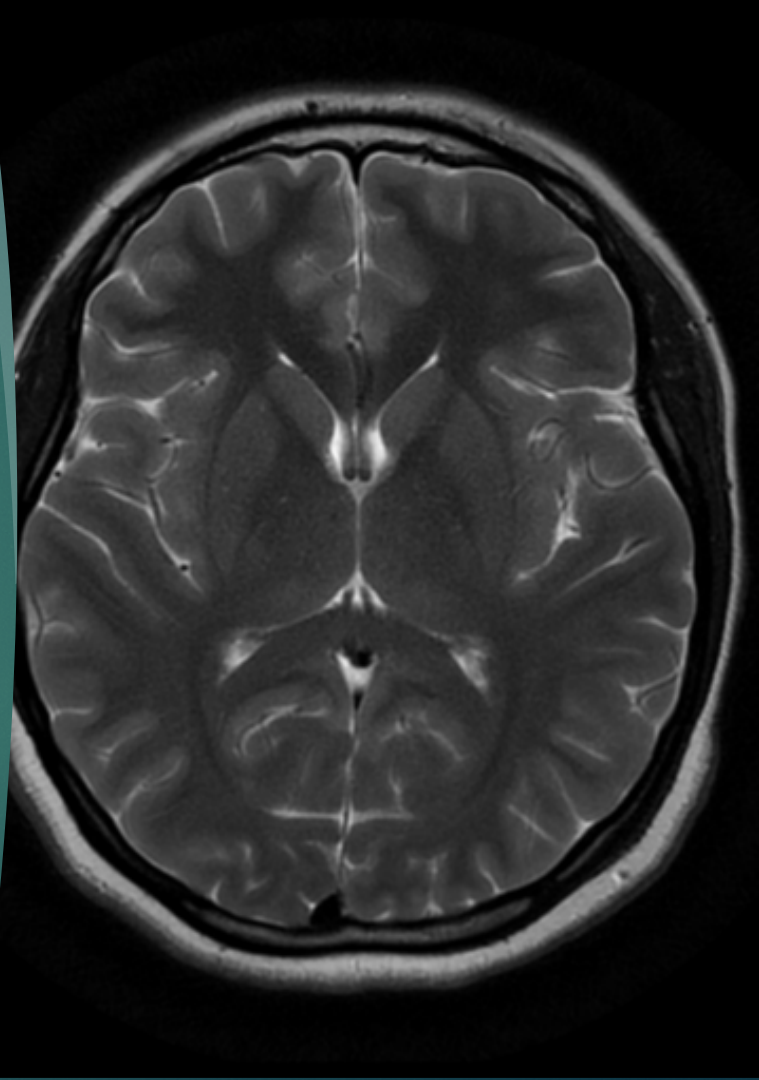
Introduction

- In this project, we designed a specialized VR environment for anatomical analysis in medical training.
- This work is developed for 3D volumetric rendering of 3D Medical images, particularly focusing on brain tumors.
- This VR application aims to facilitate medical education: Offering an interactively way to examine and analyze MRI brain scans in a fully immersive 3D space with a conversational avatar.



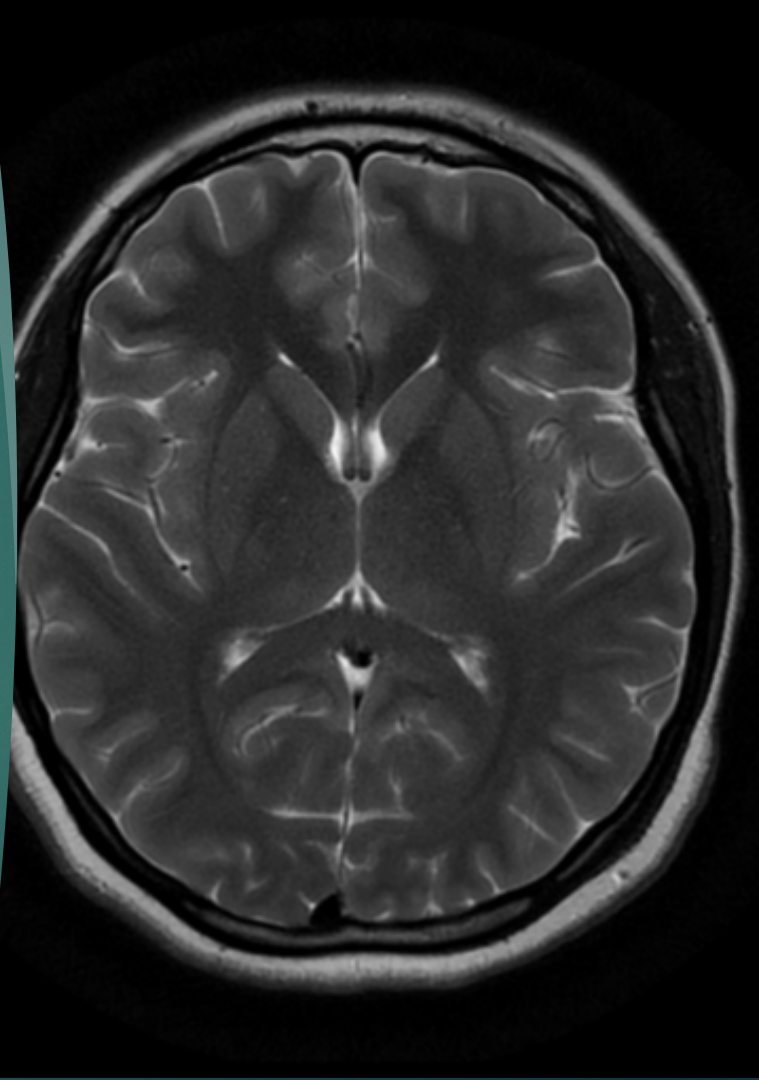
Problem Statement and Motivation

- Brain tumors impact individuals of all ages, races, and genders, with over 1.3 million Americans currently living with a primary or secondary brain tumor (ABTA, 2023).
- There are more than 100 types of brain tumors and around 27.9% are malignant.
- Annually, about 17,200 people die from malignant brain tumors.



Problem Statement and Motivation

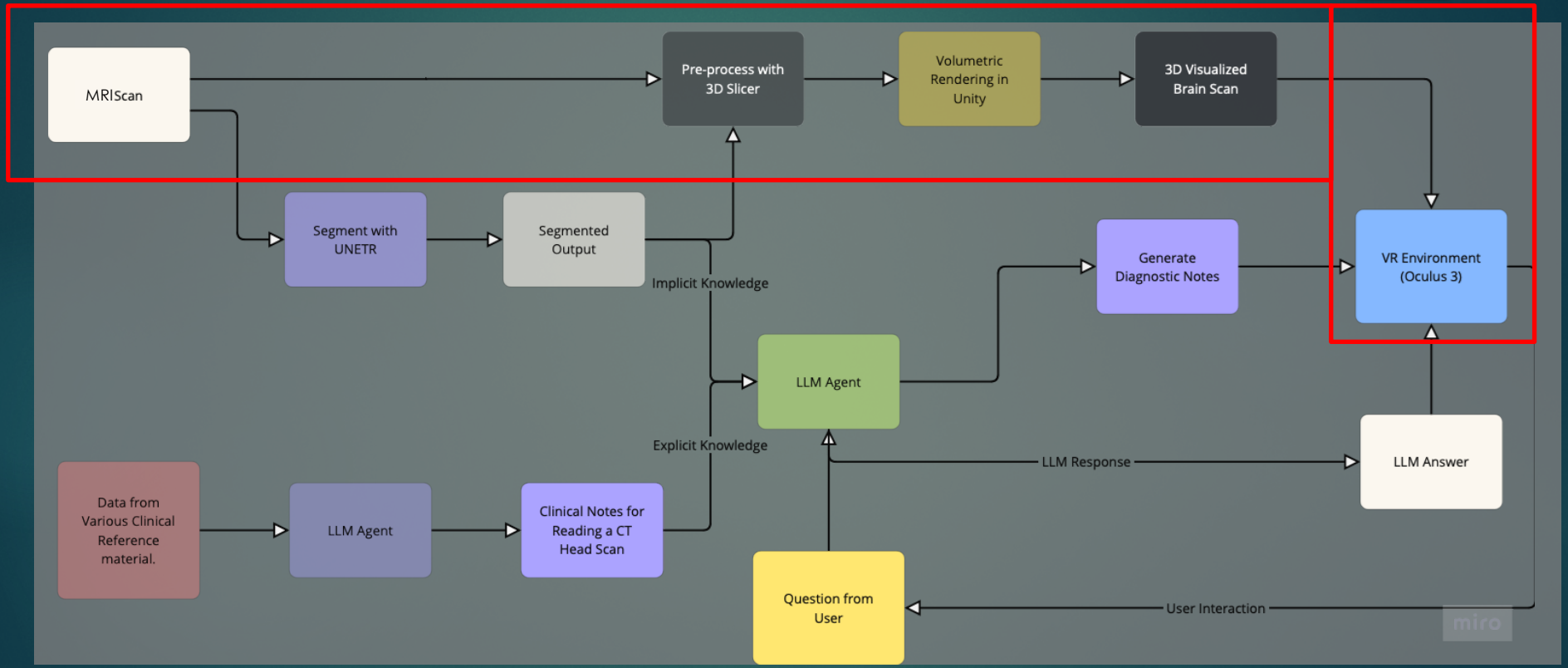
- Advanced brain imaging tools are essential for addressing critical needs in visualizing brain tumors, enhancing both education and clinical diagnostics.
- Converting 3D MRI scans into a VR-enabled format offers an engaging and intuitive method of brain visualization.
- This project offers immersive learning experiences for medical students.



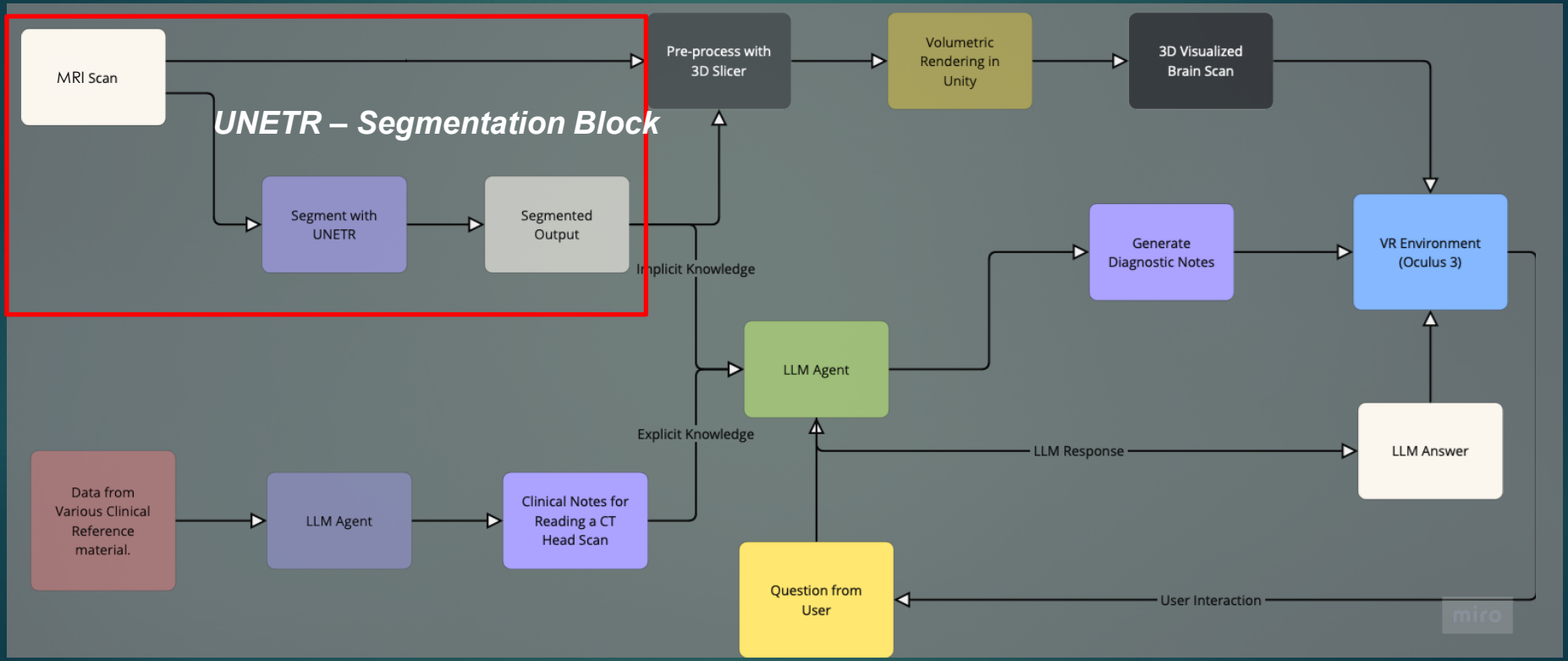
Proposed Methodology

- Key elements in our work
  - Virtual Reality Block
    - Pre-Processing of the MRI Slices using a 3D Slicer
    - Volumetric Rendering of MRI on Unity
    - VR Interactive Environment
  - AI-Assisted Segmentation Block
    - Segmentation of Tumor from MRI Scans
    - Trained a 3D-Transformer Based UNET Model for the Segmentation
  - LLM-Reasoning Block
    - Clinical Reasoning Knowledge Base
    - Diagnostic Notes Prediction
    - LLM-ChatBot for Diagnostic QA

VR Block


VR Block
MRIScan
Pre-process with 3D Slicer
Volumetric Rendering in Unity
3D Visualized Brain Scan
Segment with UNETR
Segmented Output
Implicit Knowledge
Generate Diagnostic Notes
VR Environment (Oculus 3)
LLM Agent
Explicit Knowledge
LLM Response
LLM Answer
Data from Various Clinical Reference material.
LLM Agent
Clinical Notes for Reading a CT Head Scan
Question from User
User Interaction
miro

Segmentation Block

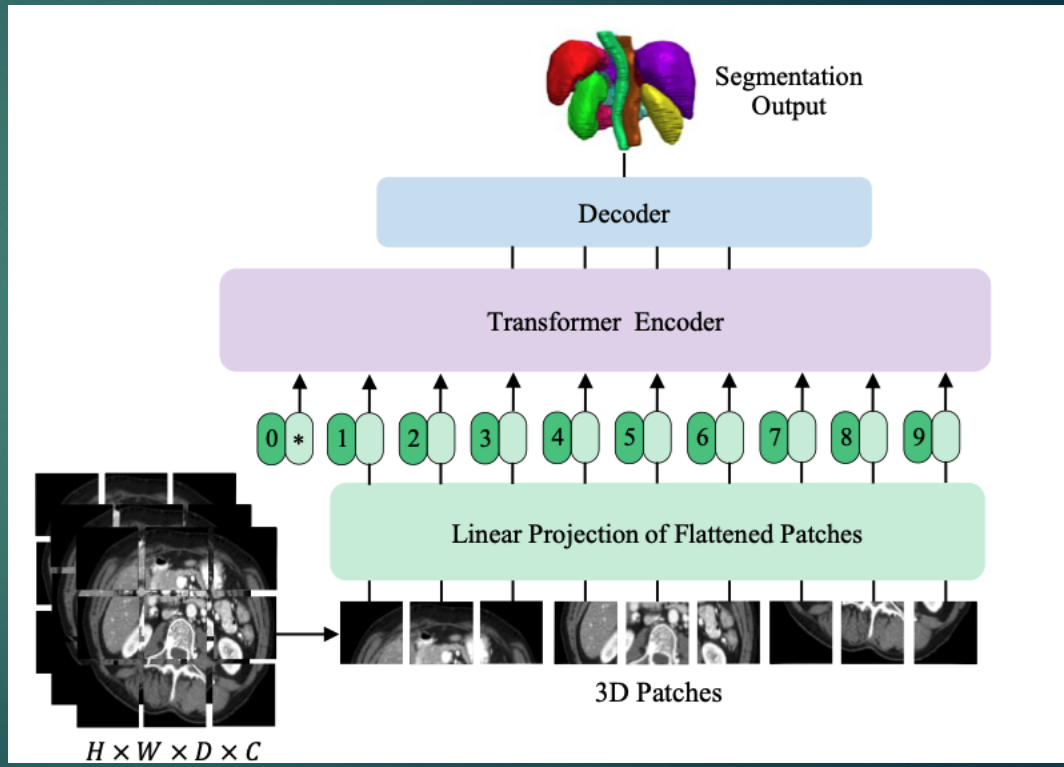


Dataset

- BraTS2020 Dataset - Part of the Brain Tumor Segmentation (BraTS) challenge, using pre-operative MRI scans from 19 institutions.
- A Total of 2349 MR Images available with manually annotated segment labels. MRI scans in multiple modalities (T1, T1Gd, T2, T2-FLAIR), stored in NIfTI format (.nii.gz).
- Focus on tumor segmentation, overall survival prediction, and progression status assessment.
- Training: 70%, Testing: 20%, Validation: 10%

UNETR

- UNETR employs an encoding-decoding structure
  - A transformer-based encoder connected to a UNet-based decoder using skip connections
- Key elements –
  - Patch-Based Embedding
  - Multi-Head Self-Attention (MSA) Block
  - Attention – Skip-connections
  - Positional Encoding



* Citation - Hatamizadeh, Ali, et al. "Unetr: Transformers for 3d medical image segmentation." Proceedings of the IEEE/CVF winter conference on applications of computer vision. 2022

Architecture

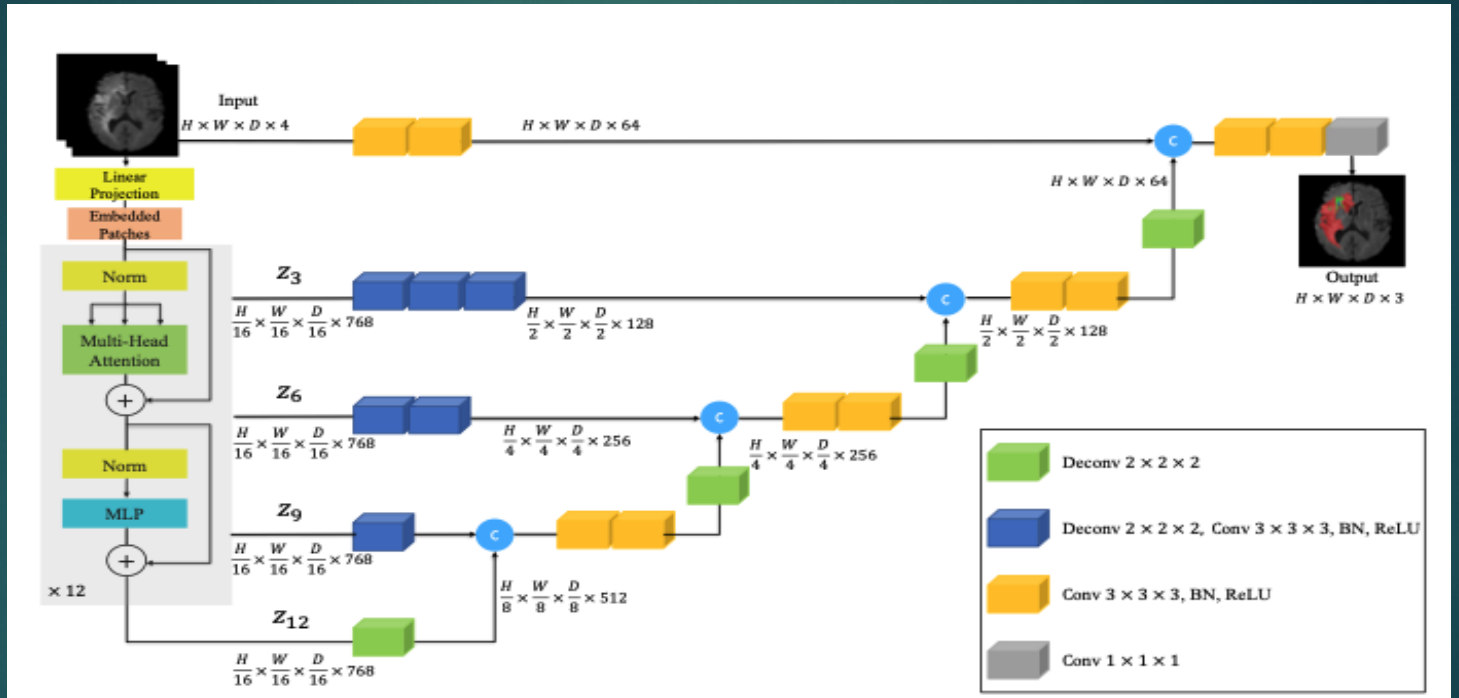
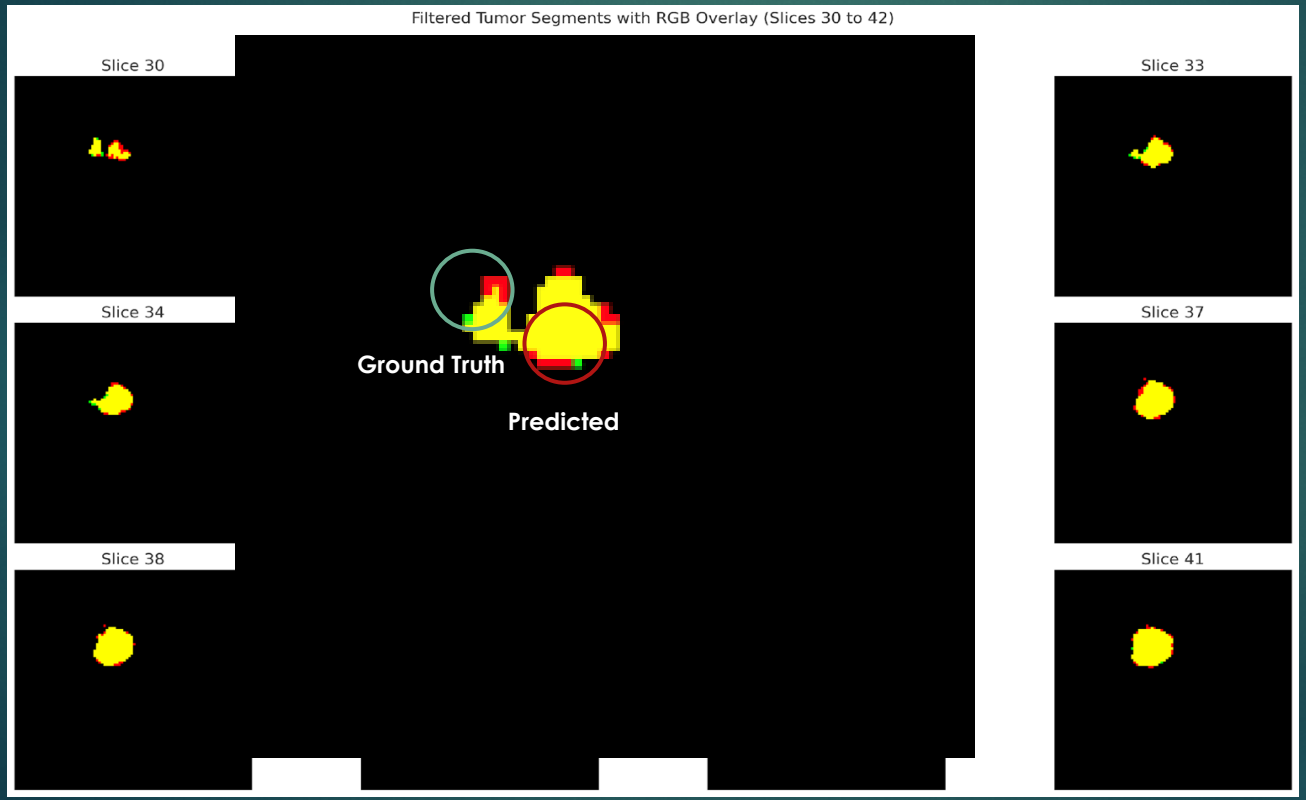


Figure 2. Overview of UNETR architecture. A 3D input volume (e.g. $C = 4$ channels for MRI images), is divided into a sequence of uniform non-overlapping patches and projected into an embedding space using a linear layer. The sequence is added with a position embedding and used as an input to a transformer model. The encoded representations of different layers in the transformer are extracted and merged with a decoder via skip connections to predict the final segmentation. Output sizes are given for patch resolution $P = 16$ and embedding size $K = 768$.

Results – GT vs Prediction Comparison Tumor Segmentations

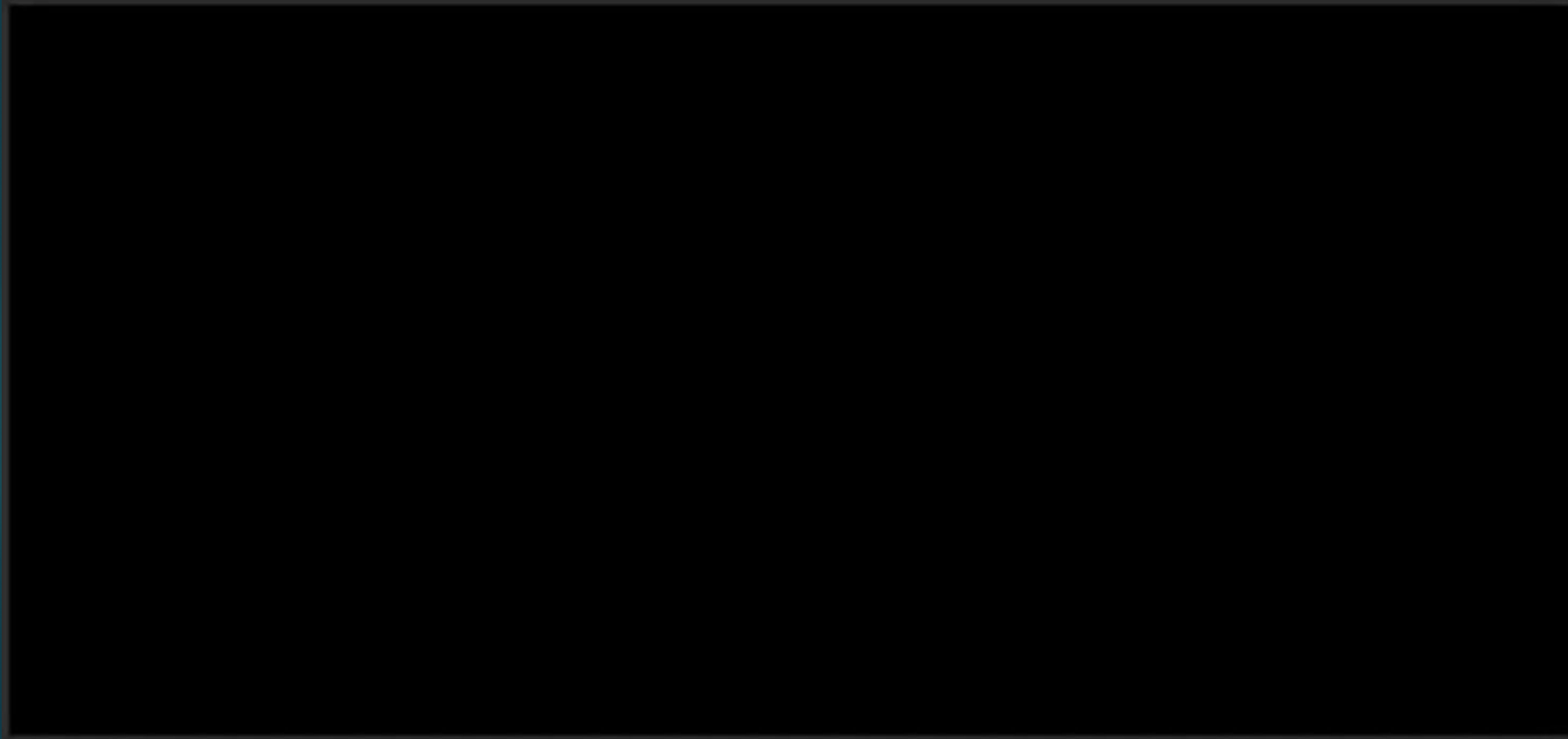


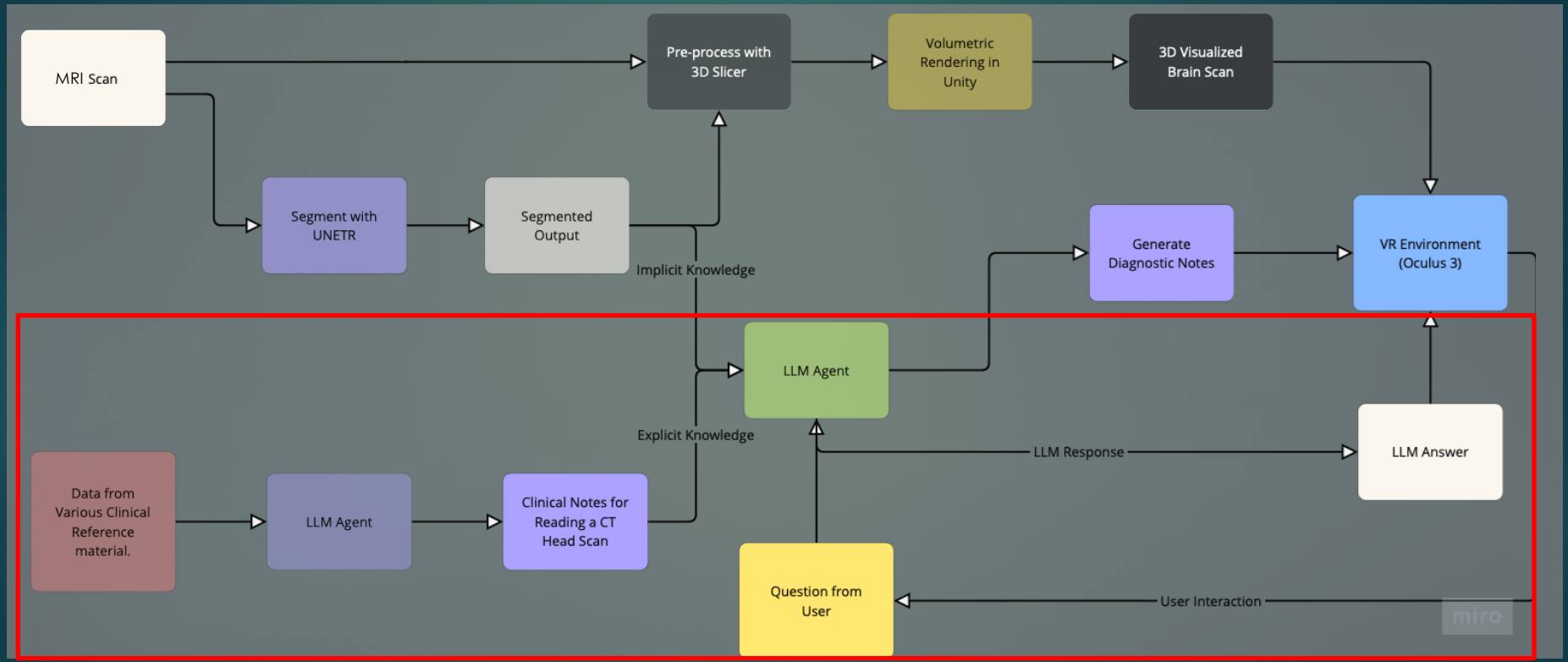
Red - Predicted Region

Green - Ground Truth

Yellow Correctly Predicted Region

Dice Score	Dice Score
Full CT Scan	0.763




MRI Scan
Pre-process with 3D Slicer
Volumetric Rendering in Unity
3D Visualized Brain Scan
Segment with UNETR
Segmented Output
Implicit Knowledge
Generate Diagnostic Notes
VR Environment (Oculus 3)
LLM Agent
Explicit Knowledge
LLM Response
LLM Answer
Data from Various Clinical Reference material.
LLM Agent
Clinical Notes for Reading a CT Head Scan
Question from User
User Interaction
miro

LLM QA Block

VR in Medical Education

- **Immersive Learning:** Simulates real-world clinical scenarios for safe, controlled skill practice (Gloy et al., 2022).
- **Enhanced Engagement:** Increases student motivation and enjoyment compared to traditional methods (Barteit et al., 2021).
- **Repeat Practice:** Enables unlimited rehearsal of procedures across disciplines without risk to patients (Barteit et al., 2021).
- **Improved Anatomy Education:** Tools like the *Immersive Anatomy Atlas* help students prepare for cadaver dissections (Gloy et al., 2022).
- **Surgical Training:** VR systems enhance psychomotor and procedural skills (Peng et al., 2021).
- **Soft Skills Development:** Interactive environments teach empathy and communication through virtual patient interactions (Jiang et al., 2021).

LLMs in Medical Education

Development of specialized medical LLMs such as:

- **DoctorGLM**: Fine-tuned on Chinese medical dialogues, offering cost-effective and bilingual healthcare solutions (Xiong et al., 2023).

- **HuatuoGPT**: Combines real-world doctor data with ChatGPT outputs to enhance diagnostic precision and patient communication (Zhang et al., 2023).

- **DISC-MedLLM & ClinicalGPT**: Incorporate knowledge graphs and electronic medical records to improve common-sense reasoning and diagnostic capabilities (Dou et al., 2024; Wang et al., 2023)

LLMs in Medical Education

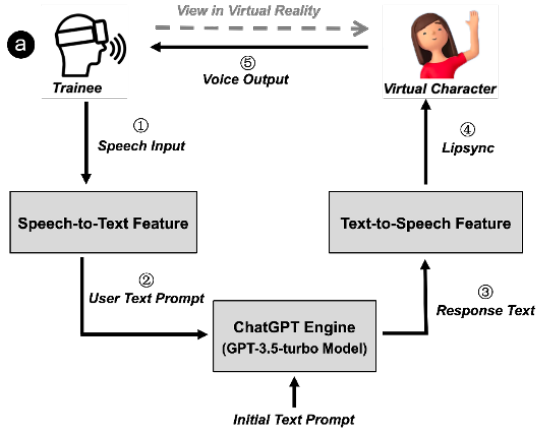
- **Dialogue and Diagnostic Training**: Simulates medical consultations with multi-turn dialogues, fostering critical thinking (Yang et al., 2024).
- **Augmented Learning Resources**: Provides structured and unstructured datasets for realistic medical scenarios (Wang et al 2023; Zhang et al 2023).
- **Assessment and Feedback**: Incorporates reinforcement learning to align with expert feedback and evaluate learner performance effectively (Dou et al., 2024)

Medical LLMs:

- **DoctorGLM**: Fine-tuned on Chinese medical dialogues, offering cost-effective and bilingual healthcare solutions (Xiong et al., 2023) .
- **HuatuoGPT**: Combines real-world doctor data with ChatGPT outputs to enhance diagnostic precision and patient communication (Zhang et al., 2023) .

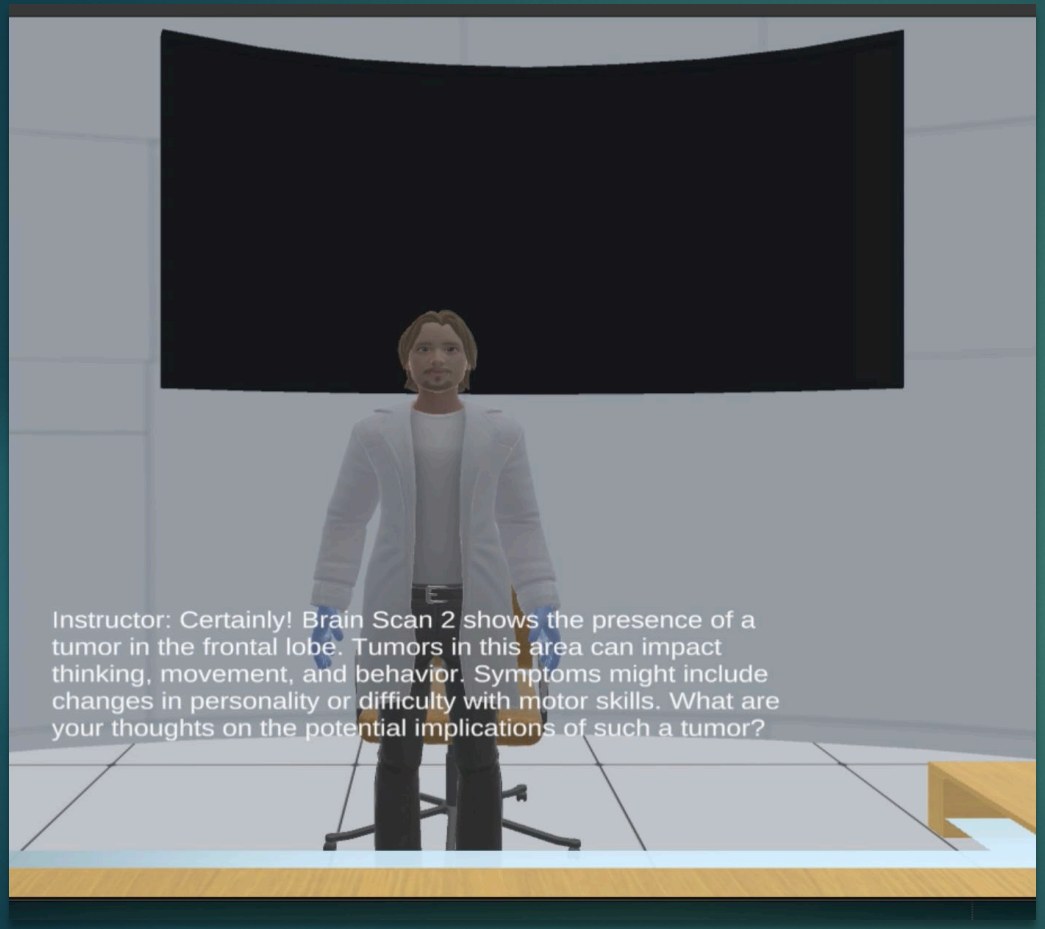
Challenges

- **Data Quality:** Ensuring reliable and accurate datasets to minimize hallucination issues (Zhang et al., 2023) .



Conversational Avatar

Inspired by previous (Li et al., 2024) works we developed the VR Instructor Chatbot application powered by ChatGPT 4. As in Li et al. (2024) project, this application allows the generation of avatar Q&A-like conversations based on the dialog context and prompt-based instructions.



With a conversational avatar we can offer the instructor-student interactions in answering general questions related to medical topics.

Clinician feedback

Qualitative Method:

- Interactive Session: 12 minutes of participant interaction recorded via video and audio to capture detailed behaviors and contextual insights.
- Semi-Structured Interview: A 20-minute interview with open-ended questions, recorded and transcribed for in-depth analysis of participant's perspectives and experiences.

Findings

Engagement & Immersive:
"*Compared to normal case studies or video lectures, it's definitely more interactive, so students can actually have some autonomy in their learning process.*"
"*Yes, it's great, it's immersive... this feels very engaging.*"

Interaction and Critical Thinking
"*It's great because it is not only giving information, but also asking some, triggering some critical thinking from the learners.*"

Pedagogical Suggestions
Evolving case study: "*Adding specific cases with questions could enhance the learning and assessment process.*"
Enhancing realism:
"*Maybe it will be very interesting if it can combine the 3D structure and also the MRI slices, because we cannot see the content of this 3D.*"

Ethical considerations: "*This is very good, I like, it is not disclosing any patient information.*"

Conclusion

Key Takeaways

- **Innovative Integration:** The project combines advanced VR environments with AI technologies like 3D segmentation models and conversational LLM-driven avatars to enhance medical education.
- **Enhanced Learning:** Immersive, interactive VR tools offer medical students a unique platform for skill development, critical thinking, and autonomy in learning.
- **Future Potential:** The integration of segmentation models and LLMs within VR holds promise for improved medical training, and patient communication.

NRT VR/AR Team

Akhil Kasturi
PhD Candidate,
Computer Vision
Graduate Research Assistant

Adma Gama-Krummel
PhD Student, Philosophy of Educational Technology.
WASP Instructor
Graduate Research Assistant

Advisor: Mudjat Cetin
Professor of Electrical and Computer Engineering
Professor of Computer Science

Collaborators

Ziming Li
Ph.D. Candidate at RIT Computing and Information Sciences / HCI x VR Researcher / Full-Stack Web Developer

Prattama Utomo
PhD Student, Higher Education (Health Professions Education)
MD, Emergency (Indonesia)
Assistant Professor, Department of Medical Education and Bioethics, Universitas Gadjah Mada, Indonesia

QUESTIONS?

Thank you!