Application and Use Cases in AI+XR

Lecturer: Dr Allison Jing

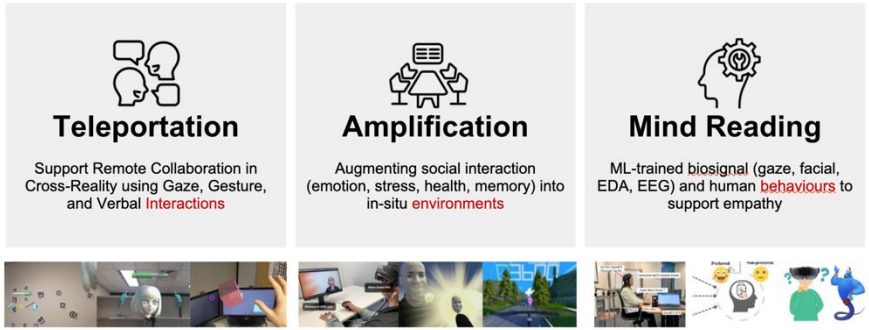
Winter School, 2025

What's next...

RMIT UNIVERSITY

About Me

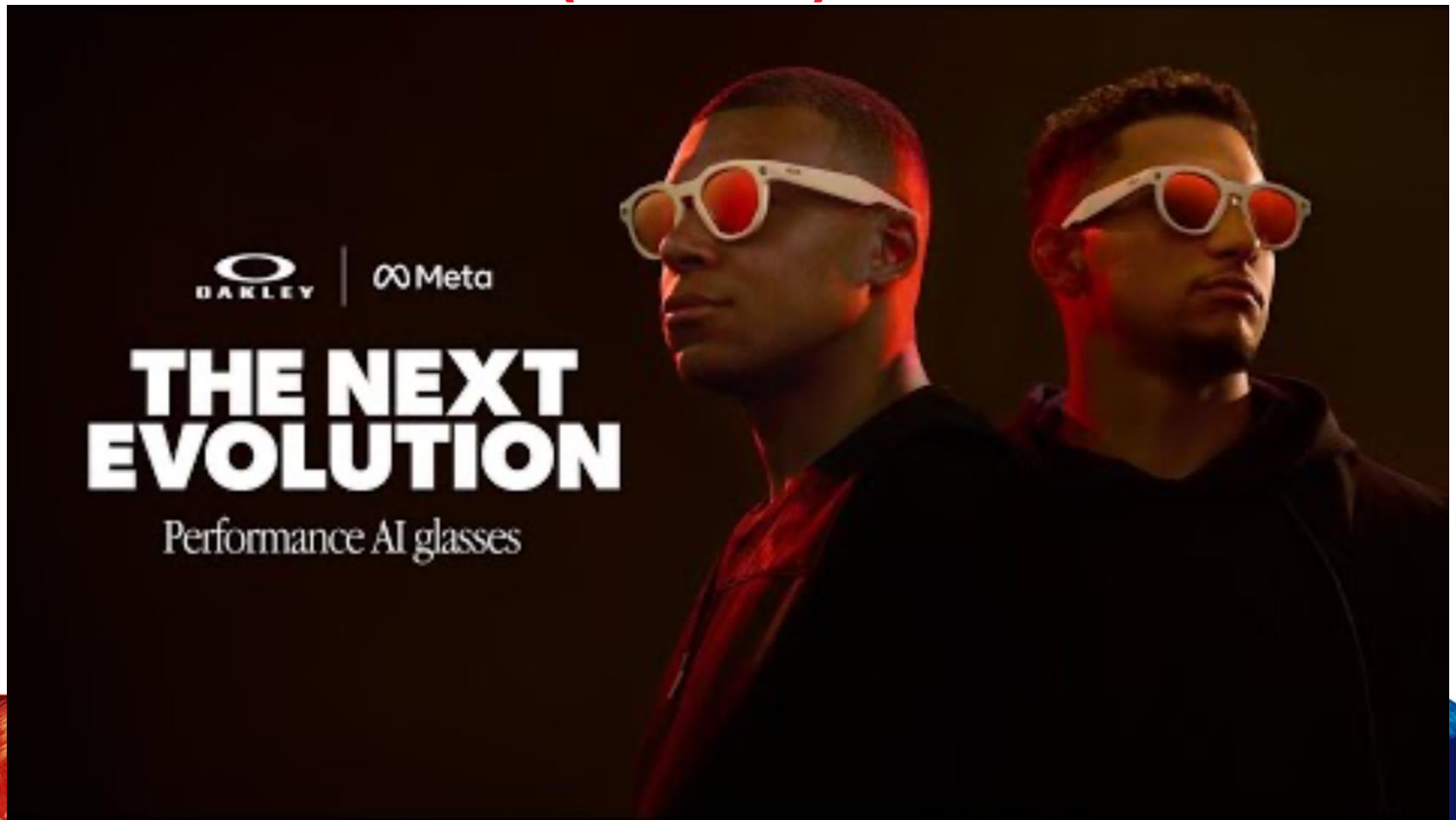
- **Lecturer** at RMIT for 2 years (teaching UG/PG programming and Mixed Reality)



- Previously **worked** with Microsoft, Xbox, and Meta as a content manager, product designer, and research scientist intern

First off let's watch some videos about AI/Smart Glasses (AI+AR)

First off let's watch three videos about AI/Smart Glasses (AI+AR)

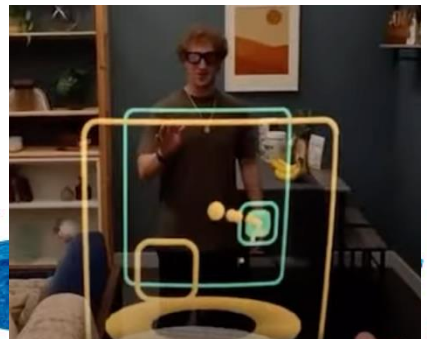


First off let's watch three videos about AI/Smart Glasses (AI+AR)

First off let's watch three videos about AI/Smart Glasses (AI+AR)

Anything in common?

- Speech command
- Conversation with the AI Agent
- Context-related information (sports performance, weather, route navigation, language translation etc)
- Real-time collaboration/Gameplay
- Others

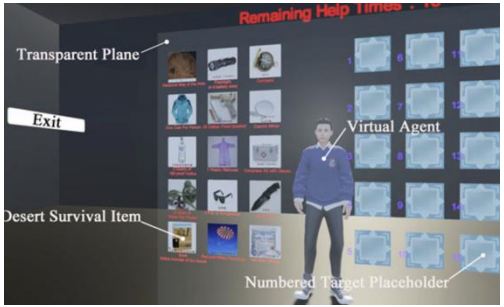


Domain use cases?

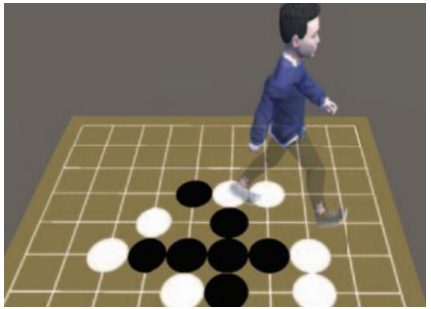
- General daily use
- Sports performance
- Navigation
- Translation
- But more! (What domain are you interested in using AI+XR?)

Intelligent Virtual Agents (IVAs)

- Emotion companion/Empathic Virtual Avatars
  - E.g. remind you of your current emotion state
- Task assistant
  - Prison's dilemma: AI-assisted Decision Making
  - IVA to provide help/intervention
  - Reinforced Learning to support personalized agent



Chang et al



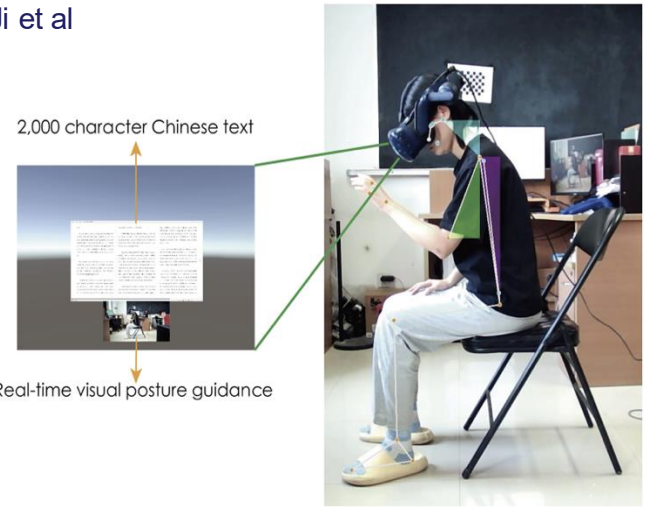
Chang et al

Gupta et al

Digital Health

- Disease or Ergonomics ML Detection in XR
  - Ji's project guide users to read in an ergonomical posture in VR using ML
- Virtual Humans (VHs) to support therapy/rehabilitation
  - E.g. Otono's project involves developing a VH to monitor the brain-injured patients and give personalised feedback. (https://www.auckland.ac.nz/en/news/2025/06/24/using-vr-for-brain-injury-recovery.html)

Ji et al



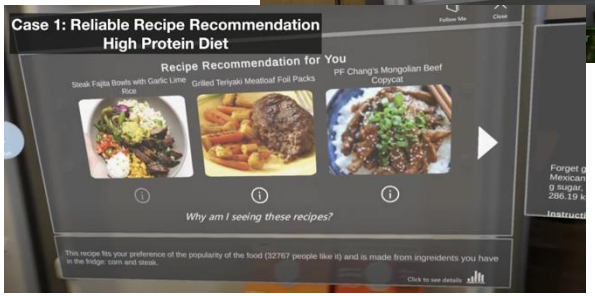
Otono et al

Multimodal XAI (Context-Aware)

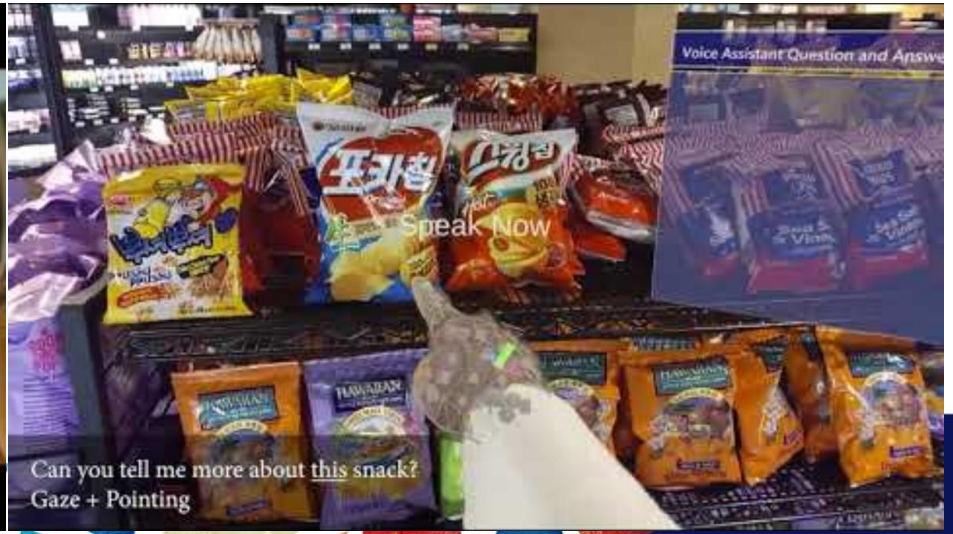
Explainable AI

- XAI – explain outputs/decisions
- Prediction and accuracy, traceability, and decision understanding
- Combines multimodal interactions

Xu et al

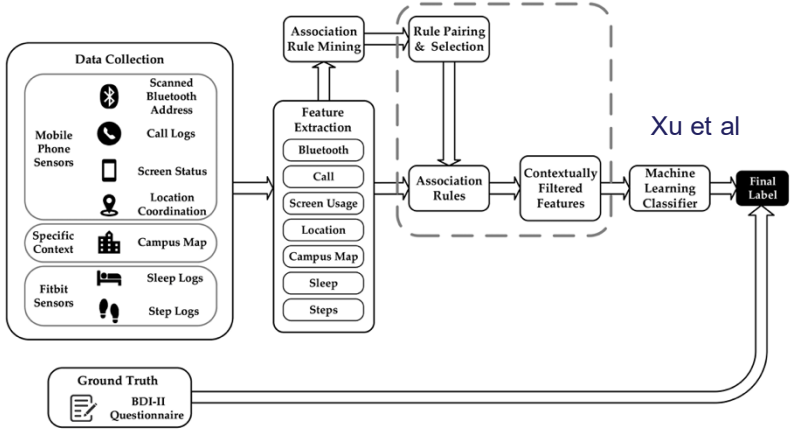


Lee et al



DL/ML Personalization and Harms in XR

- Personalised recommendation
  - Trained on your behavioural data
  - Predict recommendations for the users
- Fraudulent detection (finance tools):
  - https://www.youtube.com/watch?v=jFHPEQi55Ko
- Security and Privacy challenges in XR (especially combining AI)
  - Attacks on sensitive biometric data in GenModels (XR has integrated sensors to capture all biodata)
  - Avatar impersonation/Identity theft
  - Cyberbully/ Sexual harassment to embodied VR characters

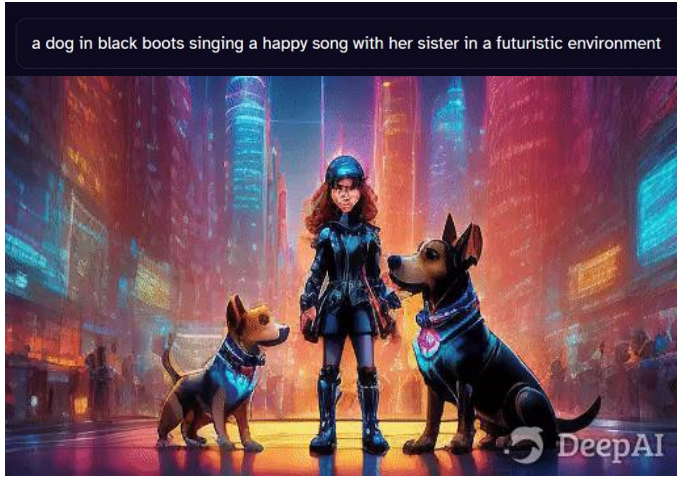


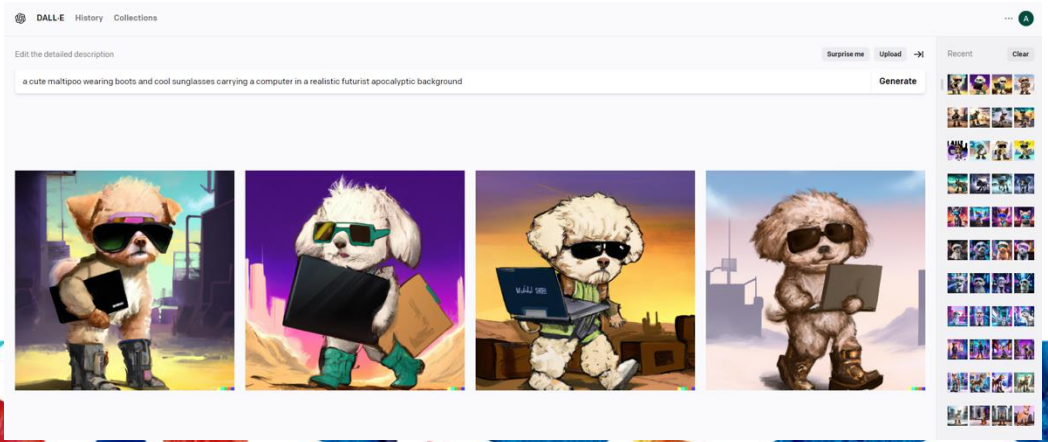
Xu et al

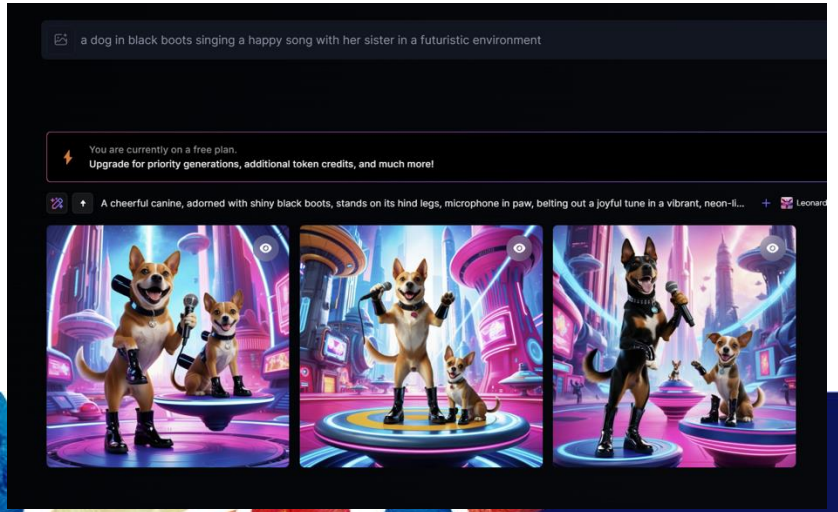
Can we build AI+XR applications to solve those issues?

GenAI/LLM Tools

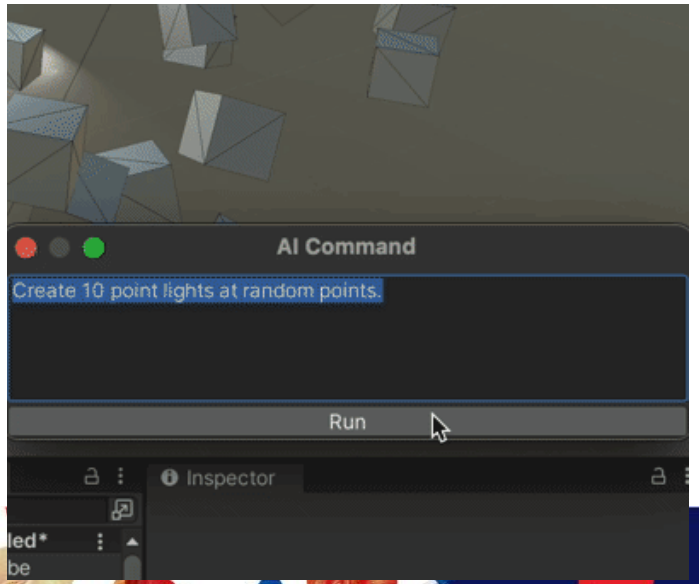
- Multimodal GenAI
  - GPT-4o: https://chat.aichatapp.ai/
  - DALL-E 3: https://labs.openai.com/
  - Pika Labs: https://chat.aichatapp.ai/
  - DeepAI: https://deepai.org/video
  - Leonardo.AI: https://app.leonardo.ai/







- Crowd shots scene generation in movies or games
  - Massive Software (https://www.massivesoftware.com/)
  - Also includes ambient agents
  - Motion agents such as horse and rider agent
- Generative AI assistant
  - Wastonx (https://www.ibm.com/Watson)
  - Code assistant, personal assistant, chatbot
  - GPT to create Unity scenes: https://twitter.com/i/status/1640677793087340544
- NLP/LLM
  - LLM: GPT (OpenAI), Llama (Meta), Gemini (Google), Claude (Anthropic), Grok (xAI)
  - Programming: Co-pilot, cursor, AI-code companion
  - Research: txyz, consensus, turnin, research rabbit



FUTURE SKILLS | TOP LARGE LANGUAGE MODELS & THIER FEATURES

CRITERIA	ChatGPT	Gemini	Claude	Mistral	LlaMA
DEVELOPER	OpenAI	Google	Anthropic	Mistral AI	Meta
RELEASE DATE	Nov. 2022	Dec. 2023	Mar. 2023	Sept. 2023	Feb. 2023
LANGUAGE MODEL	GPT 4o	Gemini 1.5 Pro	Claude 3 Opus	Mixtral 8x22B	Llama 3 (8B)
OUTPUT TOKEN PRICE	$15.00 per 1M Tokens	$21 per 1M Tokens	$75.00 per 1M Tokens	$1 per 1M Tokens	$0.1 per 1M Tokens
SPEED	74 Tokens per Second	55 Tokens per Second	32 Tokens per Second	82 Tokens per Second	866 Tokens per Second
QUALITY INDEX	100	88	94	63	65
KEY FEATURE	Generates human-like response in real time based on user-input.	Understand different types of information, including text, images, audio video & code.	Generates various forms of text content like summary, creative works & code.	It can grasp the nuances of language, context, and even emotions.	It has advanced NLP capabilities that can handle complex queries easily.

CREATED BY **FUTURESKILLSACADEMY.COM** ©

Multimodal GenAI Tool Comparison

	Tool	Company	Modalities	Strengths	Limitations
1	OpenAI GPT-4o	OpenAI	Text, Image, Audio, Video (input), Text, Image (output)	Real-time multimodal reasoning, OCR, math, vision+audio	No video generation yet, limited video/audio output
2	Google Gemini 1.5 Pro	Google DeepMind	Text, Image, Video, Code, Audio (input/output varies)	Long context (~1M tokens), fast reasoning, strong coding and vision	Video input only, not strong in image/video generation
3	Anthropic Claude 3.5 Sonnet	Anthropic	Text, Image	Text+image reasoning, documents, safer output	No audio/video modalities, no generation capabilities
4	Meta LLaVA-1.5 / CM3leon	Meta AI	Text, Image	Open-source, fine-grained image understanding	Limited generation ability, slower development
5	Fuyu-Heavy / DeepSeek-VL	Perplexity / DeepSeek	Text, Image	Open-access, strong on OCR and chart reasoning	Only image+text, no video/audio support
6	Sora (demo stage)	OpenAI	Text to Video (output)	High-quality, realistic video generation	No public access yet, input only text or images
7	Runway Gen-3 Alpha	Runway ML	Text/Image to Video	Cinematic quality, temporal consistency	Short duration, limited fine control
8	Pika 1.0+	Pika Labs	Text/Image to Video	Fast creative video clips, stylized motion	Short clips, artistic bias

Current trends

Step-by-step guide to support human via XR+AI

- Understand human biosignals (implicit inner feelings)
  - Sensor to capture human biosignals
  - ML/DL to understand signal behaviours/patterns
  - Train personalised models

- Understand human communications (explicit external behaviours)
  - Multimodality (eye, speech, gesture, body, facial)
- Understand the environment or context humans are in
  - Context awareness, AI/CV static and dynamic changes in the environment
- Augment the best experience to support human
  - Visual, auditory and haptic augmentation in XR
  - AI/ML to best customise or personalise experiences based on context
  - Superpower human beyond their physical abilities

Summary – following case studies

- Enhancing Remote Collaboration in XR with Gaze & LLM/ML
- Audiovisual enhanced AI Navigation Systems in VR
- Voxii: VR+AI Language App
- Pokemon TCG AR: AI voice command for Gameplay

Basic AI in XR

What's next...

RMIT UNIVERSITY

Basic AI/ML concepts we use in XR/HCI

- AI
- Supervised Learning
- Unsupervised Learning
- Reinforced Learning
- More fundamentals will be covered by Dr Lia Song tomorrow
- More technical details will be covered by Tamil in the afternoon

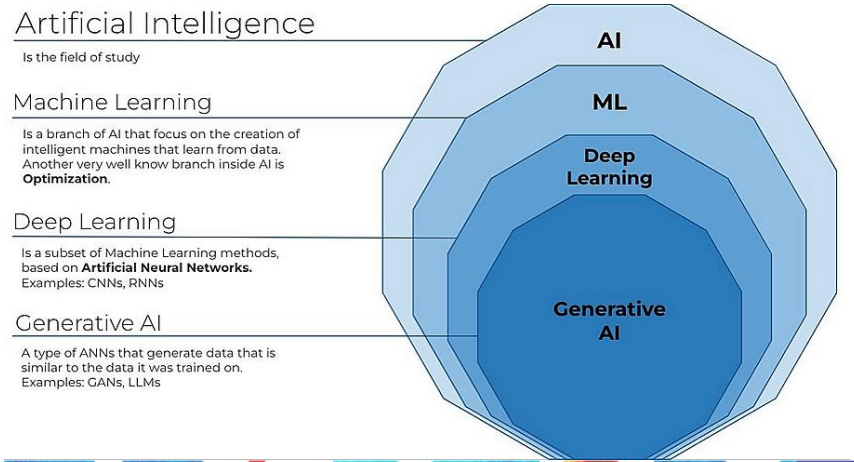
What is AI and ML?

Definition: AI is the simulation of human intelligence in machines, enabling them to perform tasks that typically require human intelligence. These tasks include reasoning, learning, problem-solving, understanding language, and perceiving the environment (e.g., vision or sound).

Machine Learning (ML) is a subset of AI. It focuses on algorithms and statistical models that enable computers to learn from data and improve their performance over time without being explicitly programmed for every task.

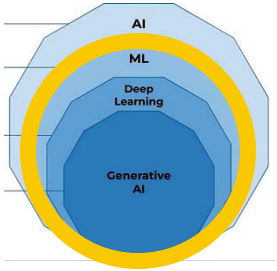
Examples:

- Virtual Assistants
- Recommender systems (Netflix, Spotify etc)
- Self-driving cars
- Fraud detections
- Chatbots (GPT)

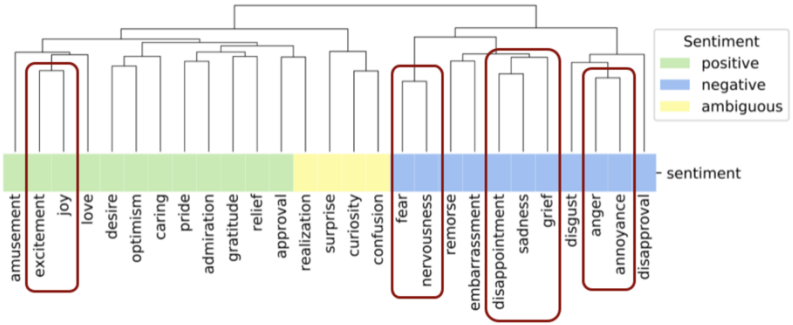


Machine Learning

Supervised learning (Classification)



- The model is trained based on the **label**: input comes with an associated output
- Correct classifications: The model can learn to map from input to output based on examples
- Image classification, face recognition, text classification etc
- Algorithms: linear regression, logic regression, Support Vector Machine (SVM), decision trees (DT), neural networks (NN)

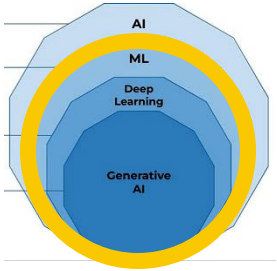


Positive		Negative		Ambiguous
admiration 👏	joy 😃	anger 😡	grief 😢	confusion 😕
amusement 😂	love ❤️	annoyance 😒	nervousness 😬	curiosity 🤔
approval 👍	optimism 🤞	disappointment	remorse 😔	realization 💡
caring 🤗	pride 😌	disapproval 👎	sadness 😞	surprise 😲
desire 😍	relief 😅	disgust 🤮		
excitement 🤩		embarrassment 😳		
gratitude 🙏		fear 😨		

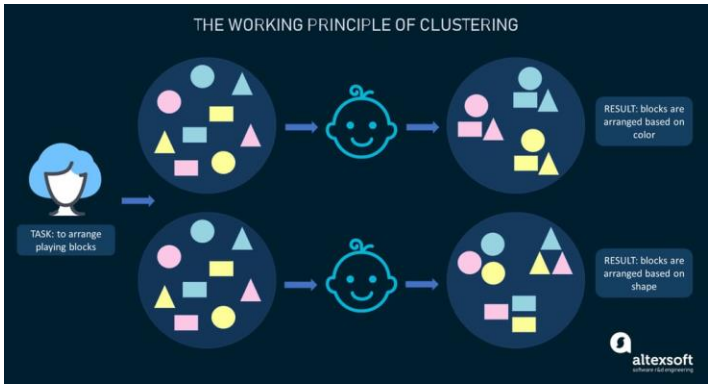
https://research.google/blog/goemotions-a-dataset-for-fine-grained-emotion-classification/

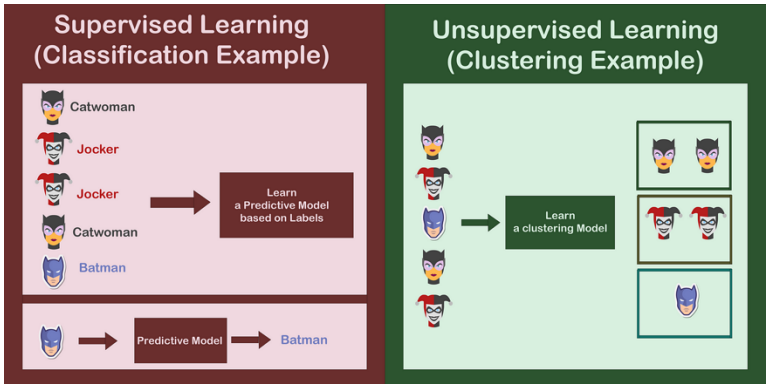
Machine Learning

Unsupervised learning (clustering)



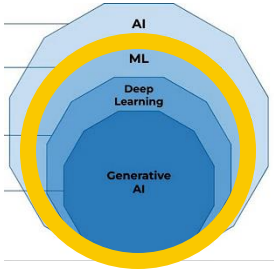
- Model is trained on the **unlabelled dataset**. The algorithm needs to discover patterns, groupings, and structures without guidance.
- Algorithms work on raw data without knowing the “correct” answers in advance
- Clustering (group to clusters), anomaly detection (detect outliers for fraud or error)
- Algorithms: Hierarchical clustering, Principal Component Analysis (PCA)



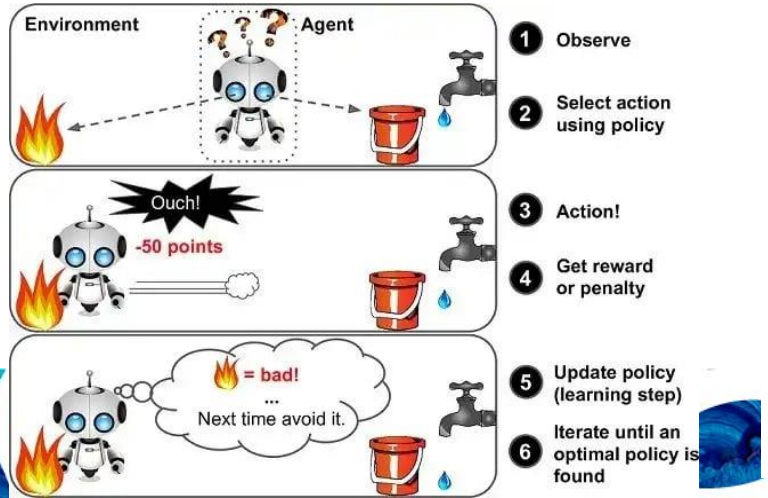


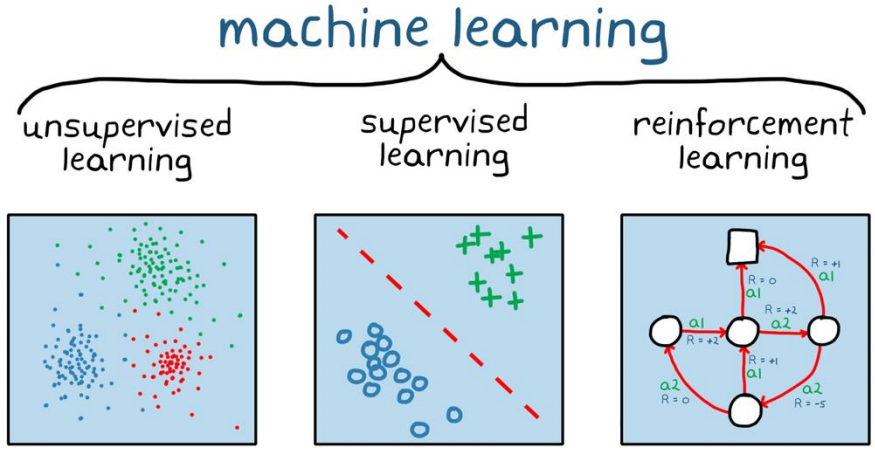
Machine Learning

Reinforced learning (learn from feedback)

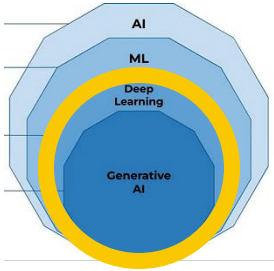


- Models learn from the interaction with an environment. It receives **feedback** through rewards or penalties.
- There is a goal associated with the environment. The environment responds to the agent's actions and gives feedback.
- Examples: prisoner's dilemma, chess (alphaGo), automation, personalized learning
- Algorithms: Q-Learning, Monte Carlo Tree Search





Deep Learning/Transformer/LLM



- **Deep Learning** (subset of ML using NN with many layers)
  - is a subset of Machine Learning that uses structures called artificial neural networks, inspired by the human brain. It's especially powerful for learning complex patterns from large amounts of data.
  - Enables learning from **unstructured data** like images, text, and audio. Learns features automatically (no need for manual feature engineering). Requires lots of data and computing power.
- **Transformer** (general purpose DL architecture)
  - is a deep learning architecture that replaced older sequence models (like RNNs) in NLP because of their efficiency and accuracy.
  - Examples including BERT, GPT, etc
- **LLM** (model trained for language tasks)
  - are large-scale **deep learning** models based on the **Transformer architecture**. Trained on massive text corpora to understand and generate language. Are the state-of-the-art in NLP, and a flagship application of DL.

With Basic AI/ML ideas Let’s see how to apply that to XR applications

What's next...

RMIT UNIVERSITY

When do you use AI + XR?

Is AI+XR the superpower I need to overcome my physical limitations?

- Do you need spatiotemporal (space and time) features?
- Does adding virtual elements/environments support your physical life/task/well-beings?
- Do you need AI-agents to understand human OR the context/environment better?

Some examples of why people think AI+XR is needed

1. Enhancing Remote Collaboration in XR with Gaze & LLM/ML
2. Audiovisual enhanced AI Navigation Systems in VR
3. Voxii: VR+AI Language App
4. Pokemon TCG AR: AI voice command for Gameplay

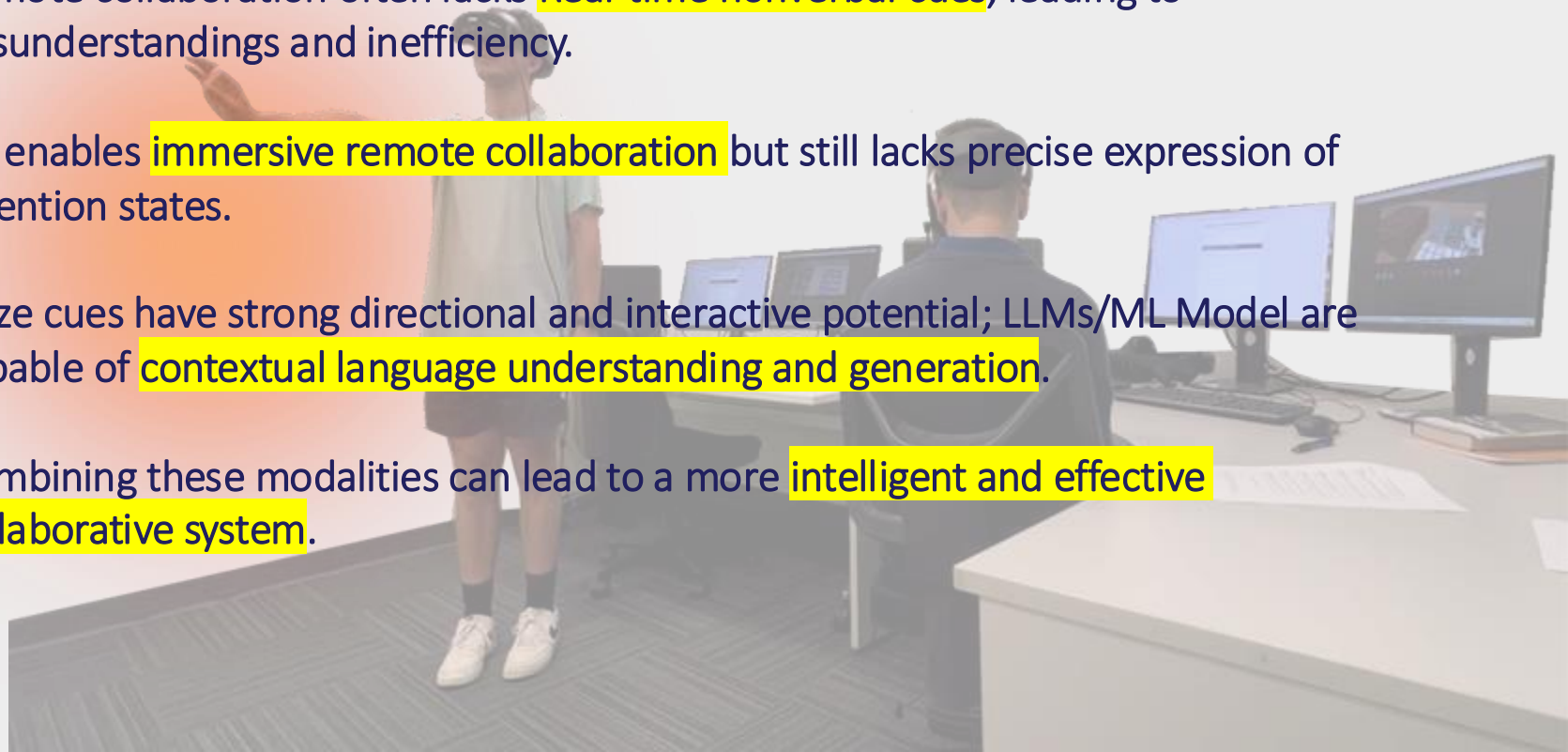
Enhancing Remote Collaboration in XR with Gaze & LLM/ML

Explore Gaze Cues Combined with LLM/ML to Improve Communication in Virtual Reality Remote Collaboration

Yi Wang

Background of the project

- Remote collaboration often lacks Real-time nonverbal cues, leading to misunderstandings and inefficiency.
- VR enables immersive remote collaboration but still lacks precise expression of attention states.
- Gaze cues have strong directional and interactive potential; LLMs/ML Model are capable of contextual language understanding and generation.
- Combining these modalities can lead to a more intelligent and effective collaborative system.

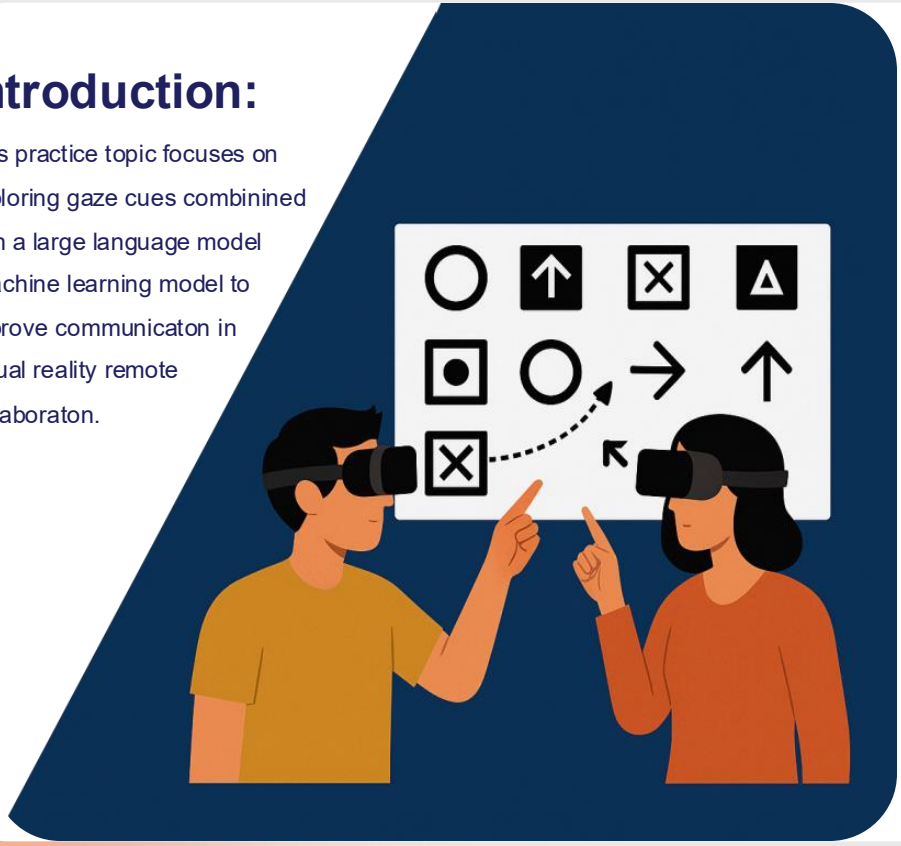


02 - Introduction Of Work - What's this project for?

Explore Gaze Cues Combined with LLM to Improve Communication in Virtual Reality Remote Collaboration

Introduction:

This practice topic focuses on exploring gaze cues combinined with a large language model /machine learning model to improve communicaton in virtual reality remote collaboraton.



· Explore the behavioral characteristics and feedback mechanisms of gaze cues in multi-user collaboration.

· Evaluate the effectiveness of machine learning–generated adaptive prompts under gaze-aware conditions;

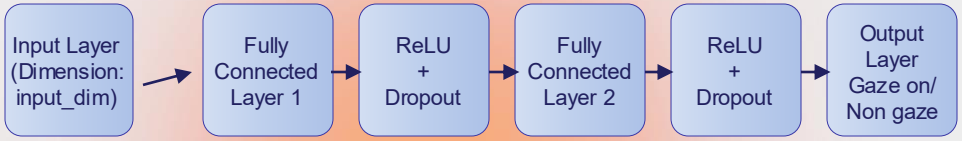
Analyzing how ML models interpret gaze data to generate real-time supportive cues that enhance mutual understanding and reduce miscommunication during collaborative tasks.

· Develop a multimodal fusion system that integrates gaze tracking with machine learning models;

Enabling dynamic feedback and intelligent guidance based on users' gaze behaviors. A collaborative VR prototype will be deployed for user evaluation, assessing usability, coordination efficiency, and user experience through both quantitative metrics and qualitative feedback.

03 - Introduction of Work –
Model Design and Evaluation in VR Collaboration?

Custom MLP Model Design and Comparison with LLM



Model Selection & Design

· In this project, we independently designed a lightweight customized Multi-Layer Perceptron (MLP) architecture tailored for real-time gaze classification in collaborative VR environments.

· The MLP model was trained to detect gaze interaction states (binary): 'Gaze On' and 'Gaze Off'.

Comparison with Large Language Models (LLMs):

· We conducted a comparative evaluation between our custom MLP model and existing LLM-based solutions
(e.g., GPT-4 prompts combined with gaze context).

· While LLMs showed flexibility in language generation, they lacked the responsiveness and lightweight efficiency required for real-time gaze-based adaptation in VR.

Final Decision:

✅ Based on performance, latency, and system integration requirements, we selected our custom-designed MLP Machine Learning Model for deployment in the final prototype.

04 - Introduction of Work –
Why MLP? Advantages of the ML Model in VR Collaboration

✅ Key Advantages:

⚡ Real-Time Responsiveness

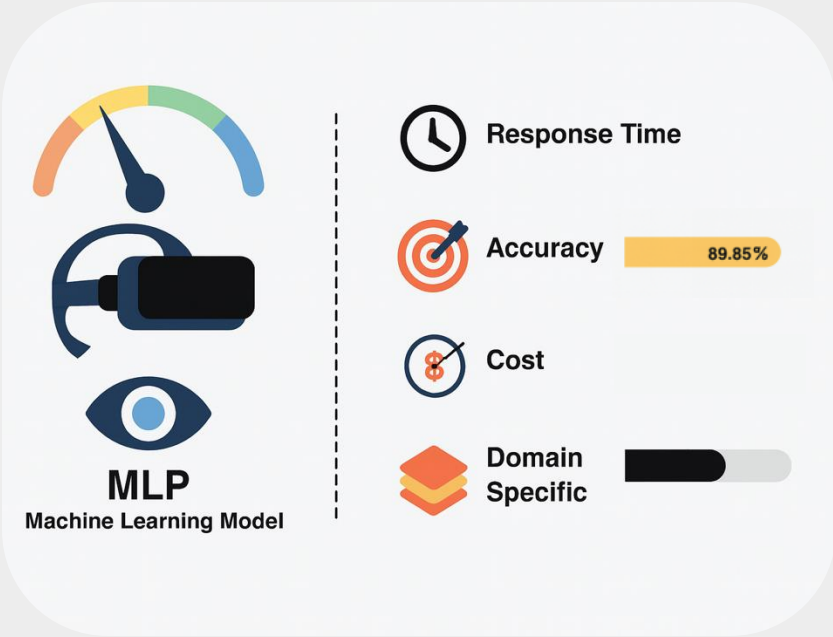
The lightweight architecture of the MLP model enables fast inference, which is crucial in time-sensitive VR collaborative tasks.

🎯 High Accuracy on Domain-Specific Tasks

Achieved an accuracy of 89.85% in detecting gaze-related interaction states (e.g., joint focus, misalignment) within our designated VR collaboration task environment.

💰 Cost-Effective & Offline Deployment

Unlike LLMs (e.g., GPT-4o, LLaMA), which require paid APIs or cloud infrastructure, the MLP model can be deployed locally without extra cost—making it a more scalable and sustainable choice.



04 - Introduction of Work –
Integration: Connecting the ML Model to Unity VR Collaboration

🎮 System Architecture Overview:

[User Interaction in VR]

↓

[Gaze + Voice Data Collection (Unity C# Scripts)]

↓

[Preprocessing (Python socket server / local API)]

↓

[ML Model Inference (Optimized MLP in Python)]

↓

[Prediction Output → Returned to Unity]

↓

[Real-time Visual Feedback in VR (e.g., changing gaze ball color/size)]

📦 Tools/Frameworks Used:

· Unity C# + Photon (for multi-user VR)

· Python (ML Inference + Socket Server)

· ONNX (optional for model export)

· Meta XR SDK / Eye-Tracking SDK

🛠️ Technical Integration Steps:

Gaze & Dialogue Capture in Unity:

· Use eye-tracking APIs (e.g., Meta Quest Pro / Tobii XR SDK) and voice input.

· Collect short dialogue segments vectors in real time.

Send Data to Python ML Server:

· Unity sends serialized gaze/context features via TCP/UDP socket or REST API to a local Python server.

Run MLP Prediction in Python:

· Preprocessed input is fed into the trained OptimizedMLP model.

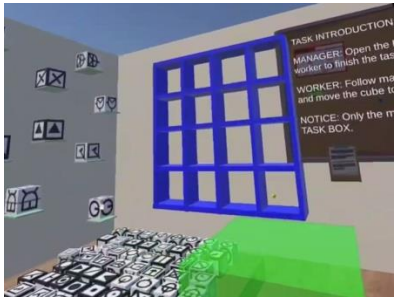
· Model returns prediction: gaze / non-gaze + confidence score.

Receive & Act in Unity:

· Unity receives prediction and updates gaze indicators or prompt display accordingly (e.g., color shift, task feedback).

05 - User Study Design - Development Work

Using unity to build up a VR collaboration environment.
Two participants work together to complete a task.

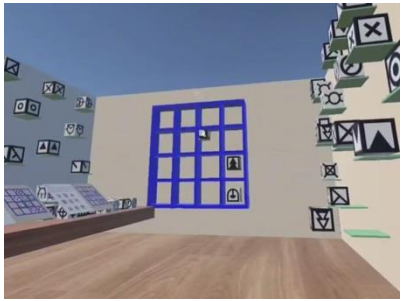


VR Environment

Two participants can collaborate in this VR environment and finish their task inside of the VR.

VR Environment

Participants' task is to move the cube to the correct place and finish all the tasks together.



VR Environment

Participants will play two roles - Task Manager and Task Worker, and try to find the most comfortable way to collaborate.

07 - Project Evaluation - Application Value

Education

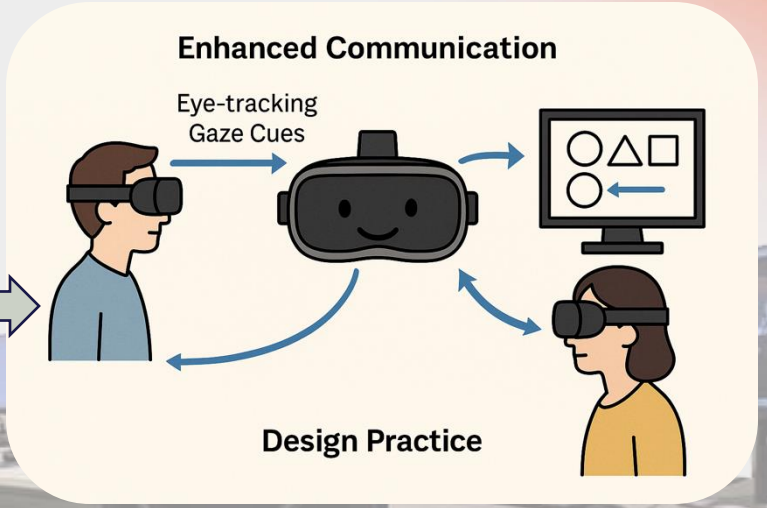
Teachers guide student focus remotely using gaze + ML Support;

Remote Work

Experts assist workers in high-stakes environments through efficient gaze-based communication;

Healthcare

Surgeons improve training and collaboration through enhanced attention sharing.



RMIT UNIVERSITY

MULTIMODAL INTERACTION FOR ENHANCING NAVIGATION SYSTEMS VIA VR

Addressing User Attention Distraction through Audiovisual Feedback

Sisi Zhang

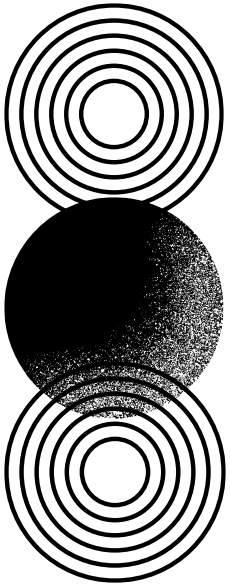
INTRODUCTION

Motivation

- Traditional navigation (GPS) can cause safety issues when attention is almost fully directed to mobile phones. Too much info can distract users.
- MR is immersive and multimodal to prioritise user attention based on context, and it also frees your hand.
- We need smarter, safer navigation tools.

Contributions

- Built an AI navigation system prototype
- Compared different guidance styles.
- Implement multimodal audiovisual interaction to drive attention based on task priorities



Research Gap

- Most VR nav uses single cues; little exo/ego-centric comparison
- Sparse evidence on haptic-audio-visual synergy
- Use multimodal to enhance user attention

Key Questions

- Which User Interface (arrow, path, minimap, audio, combo) maximizes speed & accuracy?
- Does multimodal integration reduce cognitive load?
- How do cues shape attention, confidence, and user comfort?

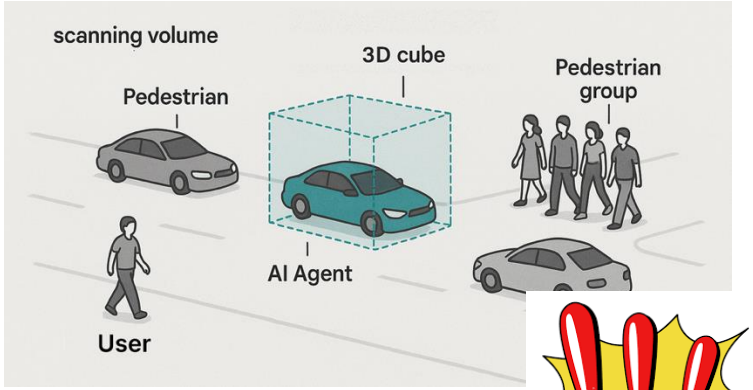
IDEAL BLUEPRINT

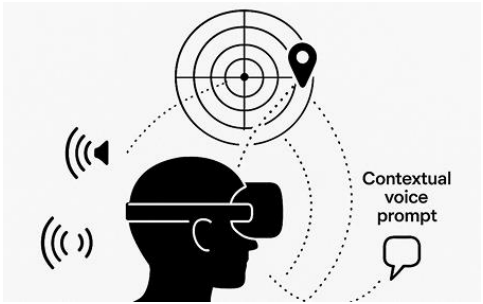
Real-time Attention Distraction Mechanisms

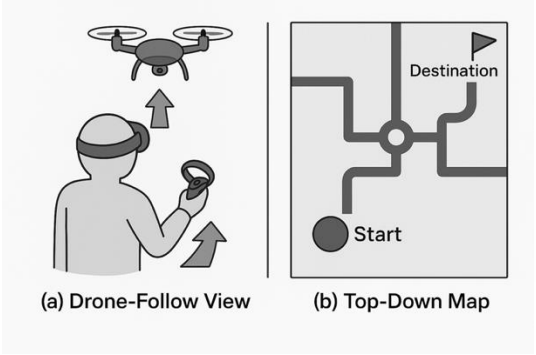
- Dynamic Traffic Simulation with AI Agent Interactions
- Environmental Context and Attention Demands: phone-call interruptions, fog & lighting changes

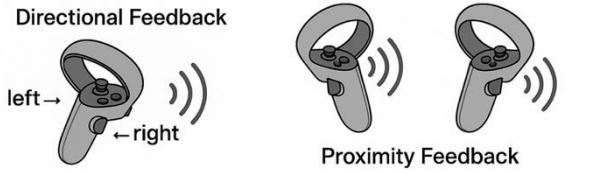
Theoretical Multimodal Navigation Framework

- **Audio Subsystem:** Spatial “radar” beeps + voice prompts (FMOD + Unity)
- **Visual Subsystem:** Egocentric 3D guidline; exocentric minimap & drone view
- **Haptic Subsystem:** Left/right pulses for direction; intensity gradient for proximity





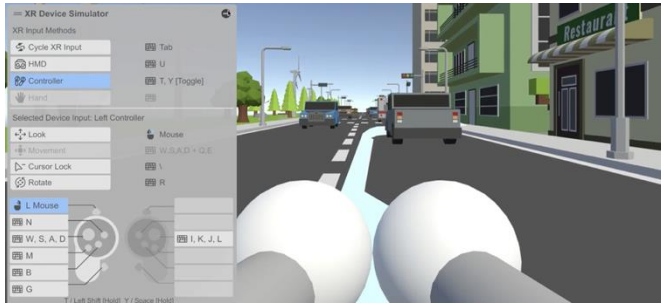




SCOPE-ADJUSTED IMPLEMENTATION

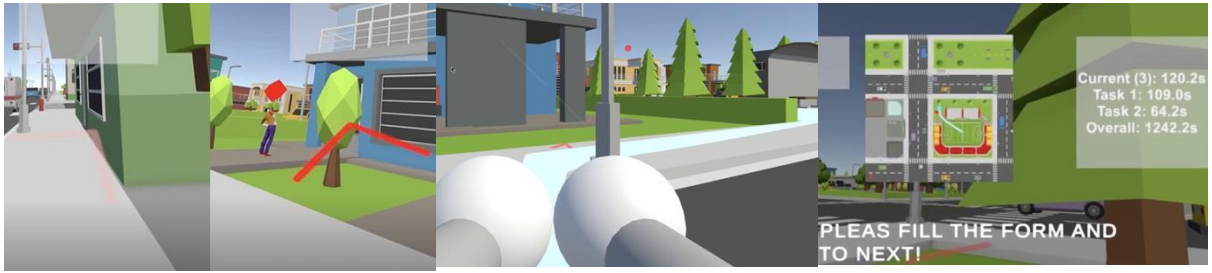
Constraints → Adaptations

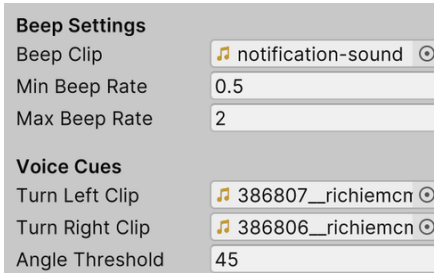
- Time in 1 semester: Real-time scanned campus → Low-poly static city
- Audio: FMOD → Unity native spatial audio
- Haptics: Complex vibration feedback → basic left/right XR motion
- Environment: No moving crowds/traffic → simple landmark layout



Simplified Architecture

- Visual Cue: Dynamic Path Visualisation(Ego) + Minimap(Exo)
- Audio Cue: Directional Voice Guidance + Proximity-Based Audio Feedback

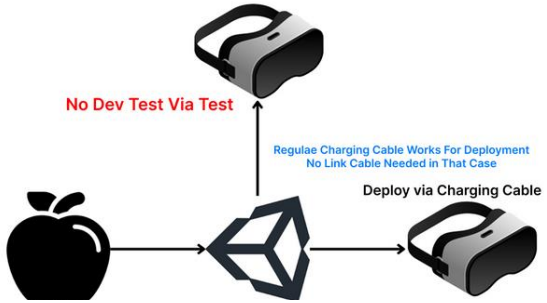




IMPLEMENTATION

Project Platform

- Development: Unity 2022.3 LTS + XR Interaction Toolkit + Quest 3 (Link cable).
- Build pipeline: Meta Quest Developer Hub + XR Simulator on Intel macOS → final testing on PC → APK deploy.
- Github : https://github.com/Selii341/MetaCityNavProj.git [12].

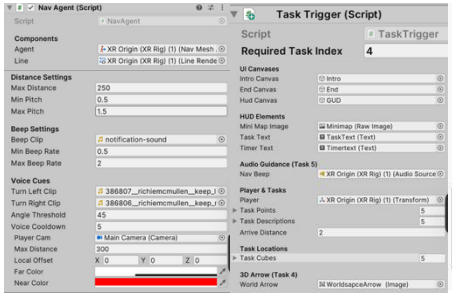


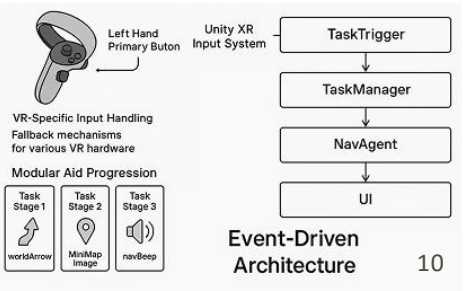
Project Setup

- Scene: low-poly city, red cube targets, AI Navigation.
- UI: World-space canvas for arrow, HUD canvas for minimap and timers.

Core Scripts

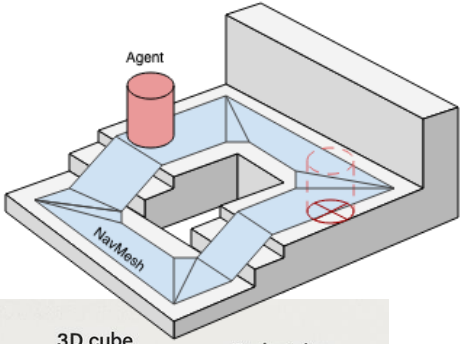
- TaskManager, TaskTrigger, NavAgent, WorldSpaceArrowController, AudioSpatialManager.

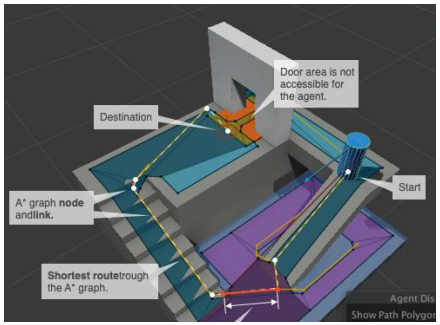


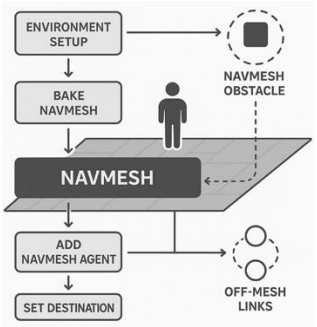


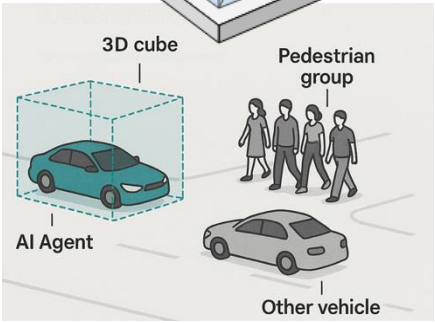
AI NAVIGATION(UNITY)

A data structure (NavMesh) representing walkable surfaces in our game world that AI agents use to automatically calculate paths, avoid obstacles, and navigate complex environments intelligently.









Foundation Layer

- NavMesh: 3D walkable surface mesh (baked from static geometry)
- Agent: Movement controller (speed, acceleration, obstacle avoidance)
- Algorithm: A* (shortest path with obstacle/cost analysis)

Workflow Overview

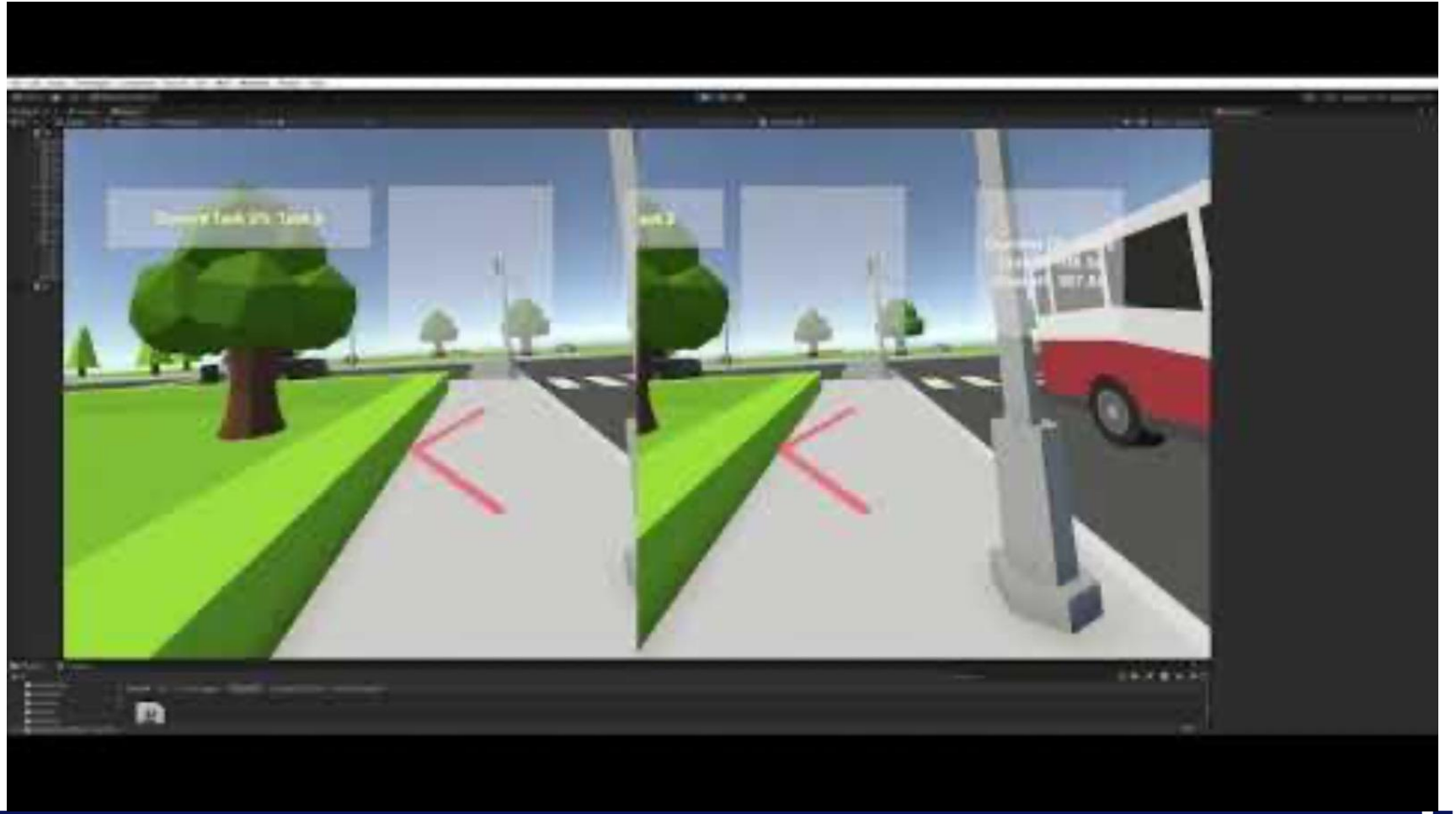
- Mark objects → "Navigation Static"
- Add NavMeshSurface → Bake
- Add NavMeshAgent → AI character
- agent.SetDestination(target.position) → script

Advanced Features

- Pedestrian Groups - Formation & flocking behaviors for realistic crowd dynamics
- Hybrid AI State Machines - Dynamic parameter adjustment responding to hazards & events
- Tactical Navigation - Strategic traffic rule enforcement and violation detection


x
z
Persp
Orientation
AI Navigation
Surfaces
Show Only Selected
Show NavMesh
Show HeightMesh
Agents
Show Path Polygons
Show Path Query Nodes
Show Neighbours
Show Walls
Show Avoidance
Obstacles
Show Carve Hull

DEMO

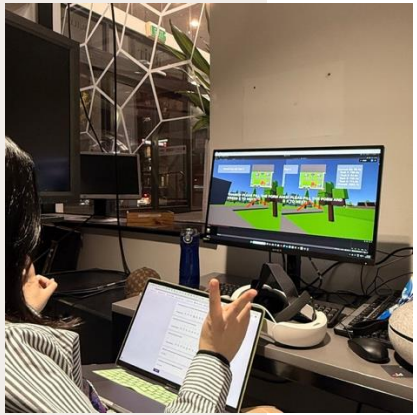


USER FEEDBACK

"I actually started liking the audio after a few tries... It reminded me of using GPS in my car—I could keep my eyes on the road and just listen for the directions."

"The audio kept interrupting my flow... It was like having a backseat driver who couldn't see what I was seeing. I finished faster but felt angry the whole time."

"When everything was turned on... I didn't know which thing to pay attention to... It felt chaotic. I kept thinking, 'Which one is right?'"



Behavioral Insights

- Familiarity Bias: audio as backup
- Flow Disruption: no continuous attention
- Safety Perception Split: cognitive load vs. long-term navigation

reduced cognitive load as safer vs. over-reliance(6:4)

LIMITATIONS & FUTURE PLAN

Current Limitations

- **Simplified Environment:** No moving traffic or real GPS data integration
- **Future Directions:**
  - Dynamic, traffic-rich VR scenes
  - Adaptive cue selection based on real-time workload (objective signals)
  - Personalized cue intensity & modality switching

Conclusion

- Multimodal VR navigation can reduce distraction and boost performance when designed judiciously.
- Prioritize clear visual paths, augment with selective audio alerts, consider optional haptics.
- **The future of navigation systems:** Moving from universal guidance tools toward **personalized cognitive partners** that adapt to individual differences while supporting both immediate navigation success and long-term spatial competence development.

Ethan Herpich, Kaleb Cole, Kulindu Chirantha Keeriwela Gamage, Varinder Pal Singh Bhatti

RMIT UNIVERSITY

Team

Ethan

Lead Dev
Unity Scene
Menu UI

Kulindu

AI API
Animations
GitHub

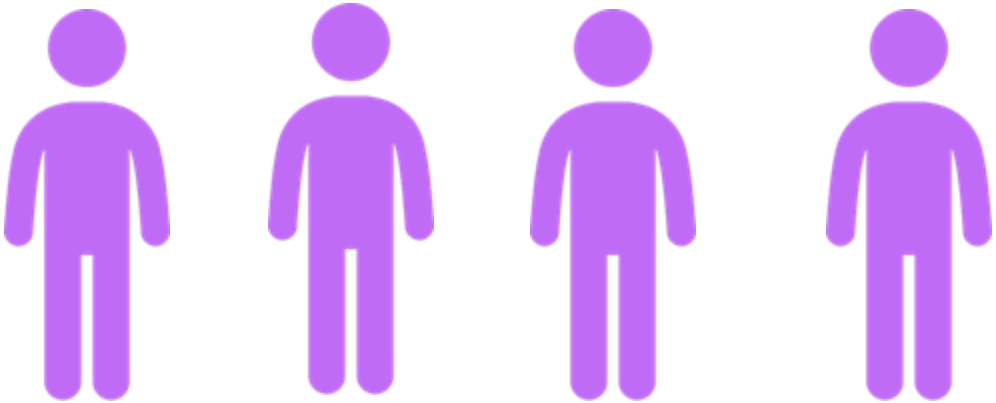
Kaleb

Lead PM
Prompt Engineer

Uday

Lead Artist

Defining the Problem



44% of Australians adults have low reading literacy (levels 1 to 2)

Australian Government Style Manual. (n.d.). Literacy and access. Style Manual. https://www.stylemanual.gov.au/acce... inclusive-content/literacy-and-access

Defining the Problem

IMMERSE

- Focuses on live instructors and community
- Lacks personalisation

duolingo

- Offers AI capability
- Lacks immersion

Key Features

1. **Conversational AI Avatar:** Engage in conversations with intelligent, responsive avatars.
2. **Grammar Checking AI:** Receive dynamic and personalised feedback and corrections based on the conversation
3. **Immersive Scene:** Immerse yourself in lifelike scenes, where you can apply your language skills in practical, scenario-based exercises.
4. **Menu and Scene Selection:** Select scenes based on specific learning goals, giving you control over the type of vocabulary and situations you want to practice
5. **Gamification:** Collect points based on accuracy and response time, encouraging consistent practice and measurable progress.

User Flow

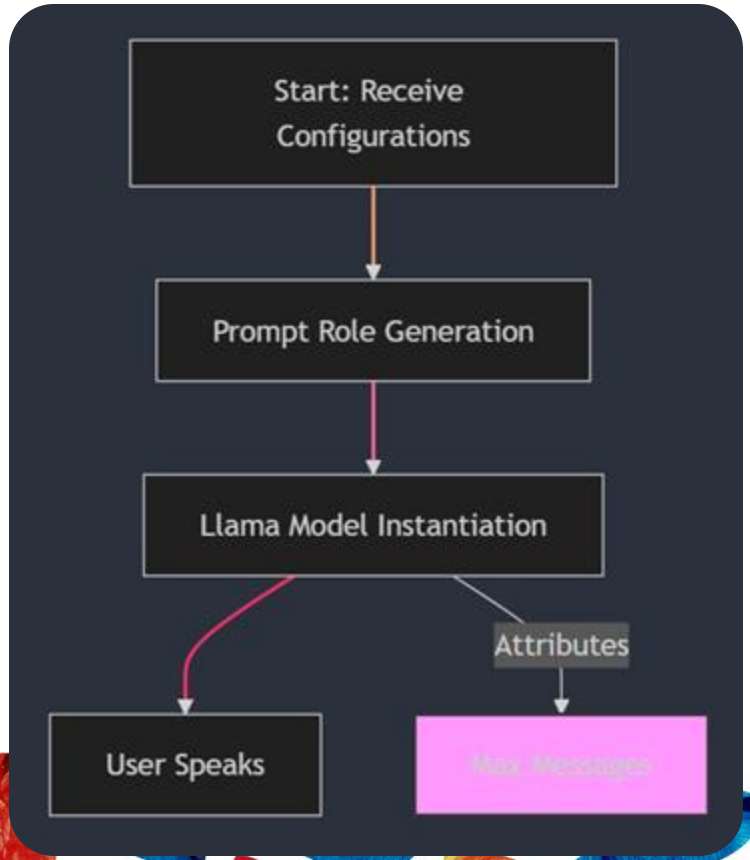
Onboarding Flow

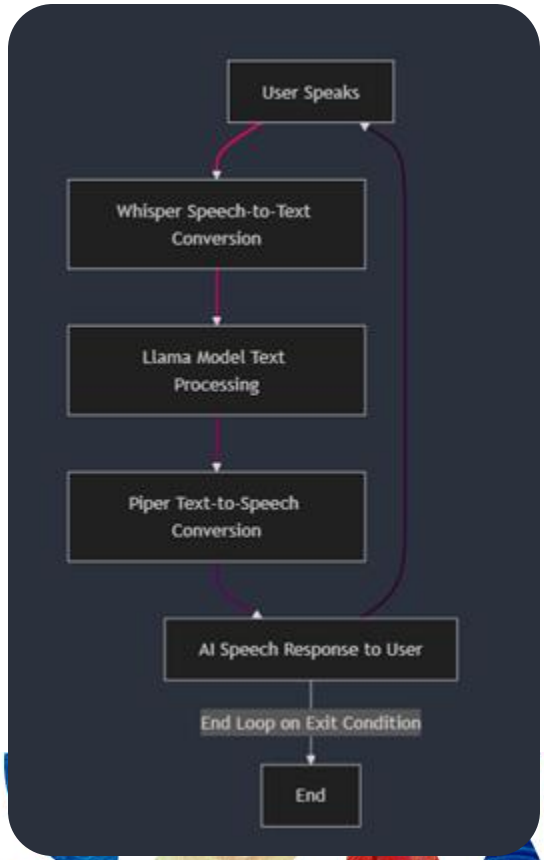
Language Proficiency
Vocabulary Content
Scene
Avatar Hostility

AI Interaction

Post Interaction Feedback and Scoring

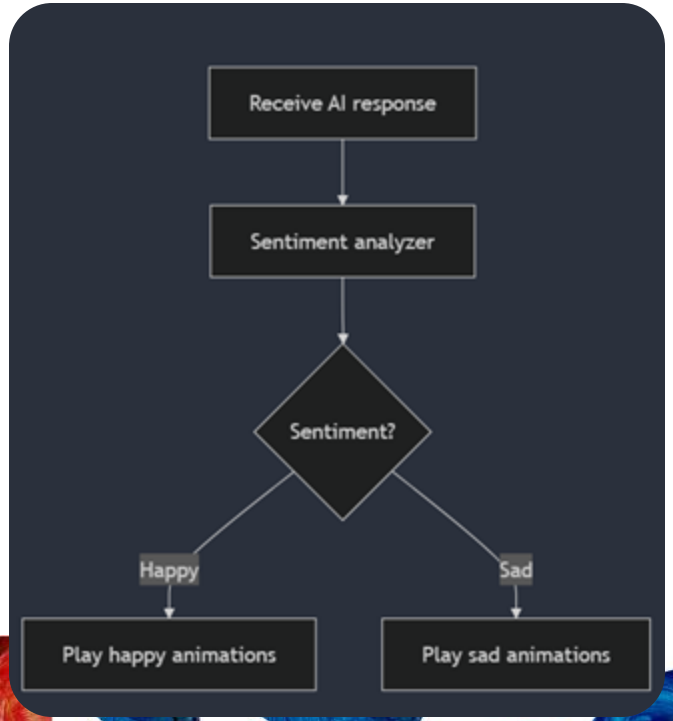
Feature 1: Conversational AI Avatar - AI

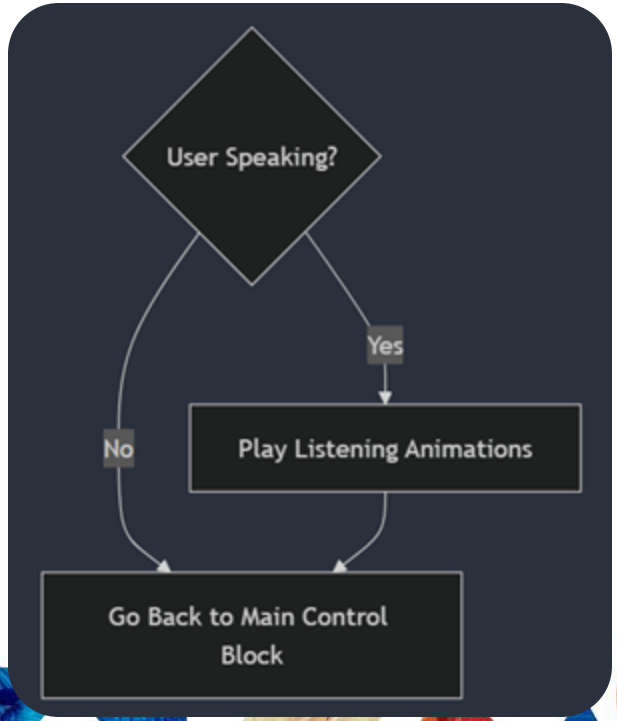




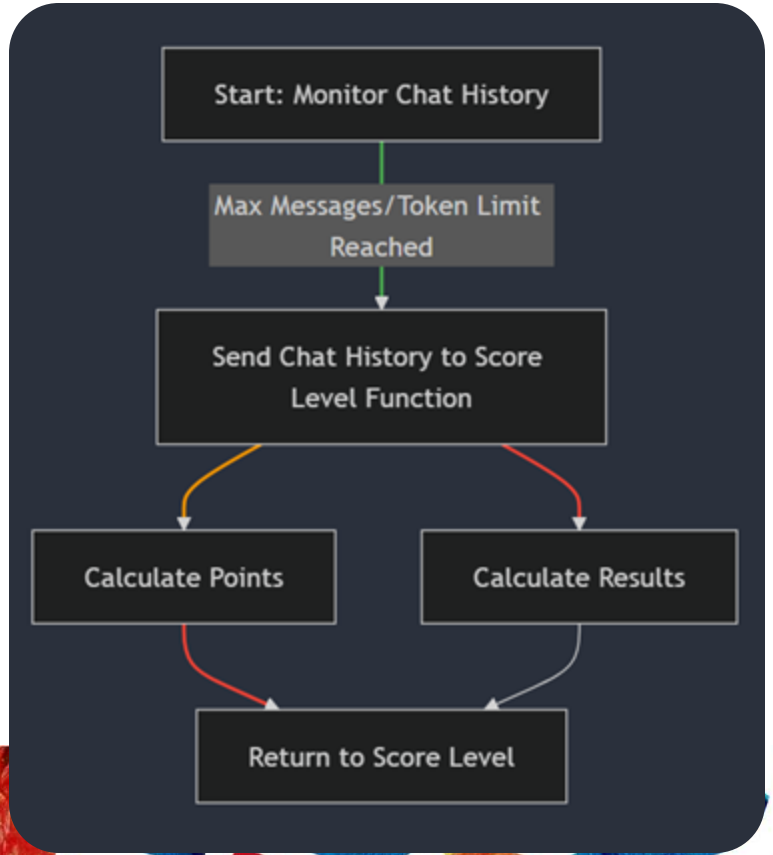
Feature 1: Conversational AI Avatar - Animations

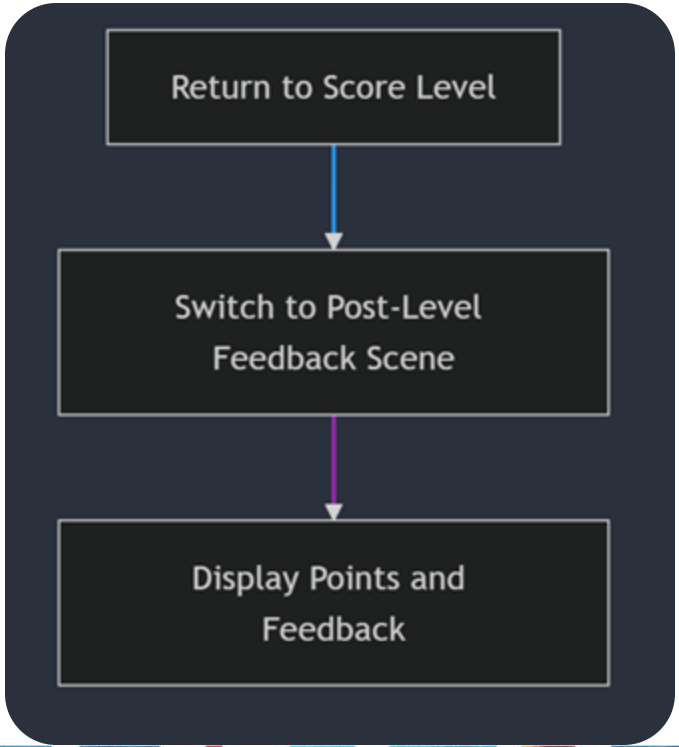
- Mixamo was used to download animations





Feature 2: Grammar Checking AI





Feature 3: Immersive Scene

Goal is to provide a unique immersive experience.

Key elements include:

- Unique verbal conversation
- High quality environments
- Responsive animated avatar
- Background noise

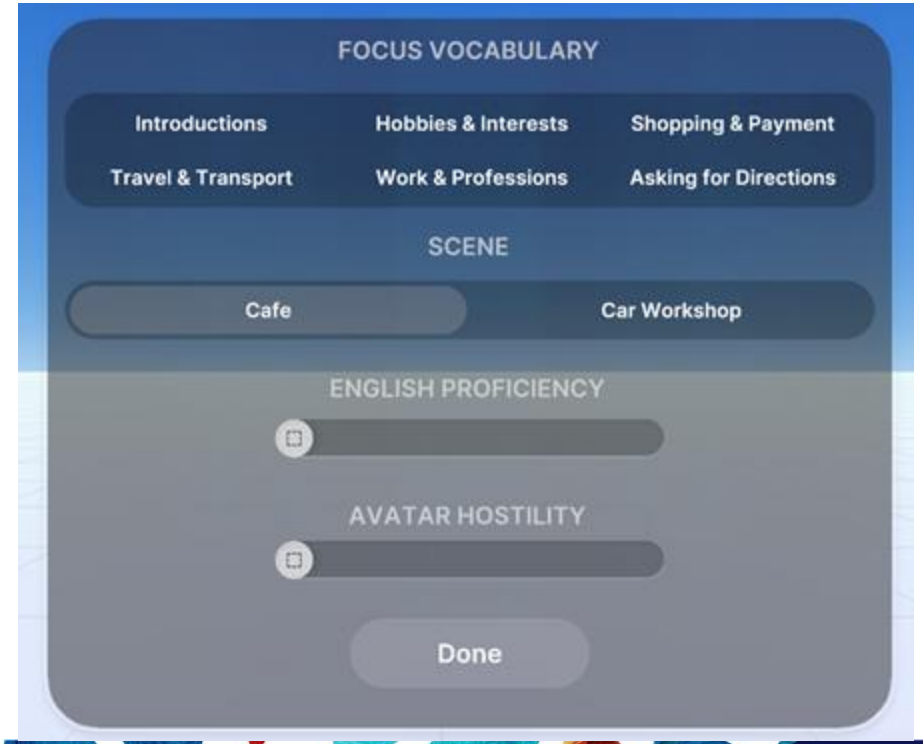
Context rich ‘Cafe’ scene (designed by Uday)

Feature 4: Menu and Scene Selection

The user wants to personalize the experience so it suits their learning needs:

We enable this by providing options for:

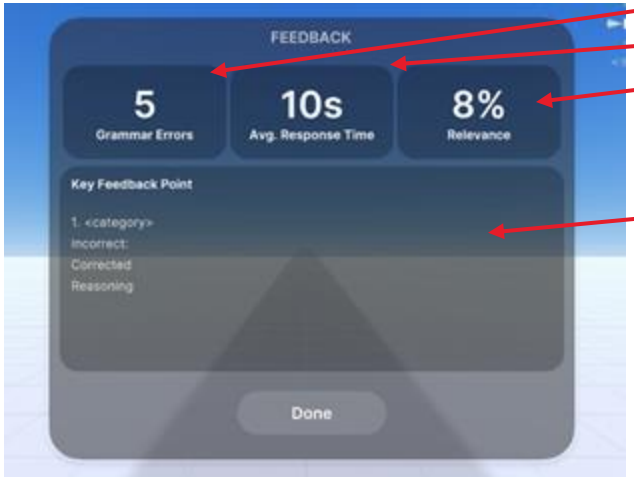
- **Focus vocabulary**
- **Scene**
- **English proficiency**
- **Avatar hostility**



Feature 5: Gamification

1. **Point based system:** Simple metric to drive gamified experience
2. **Feedback:** Educational points the user can learn from

Point Screen



Feedback Screen

Number of errors

Average response time

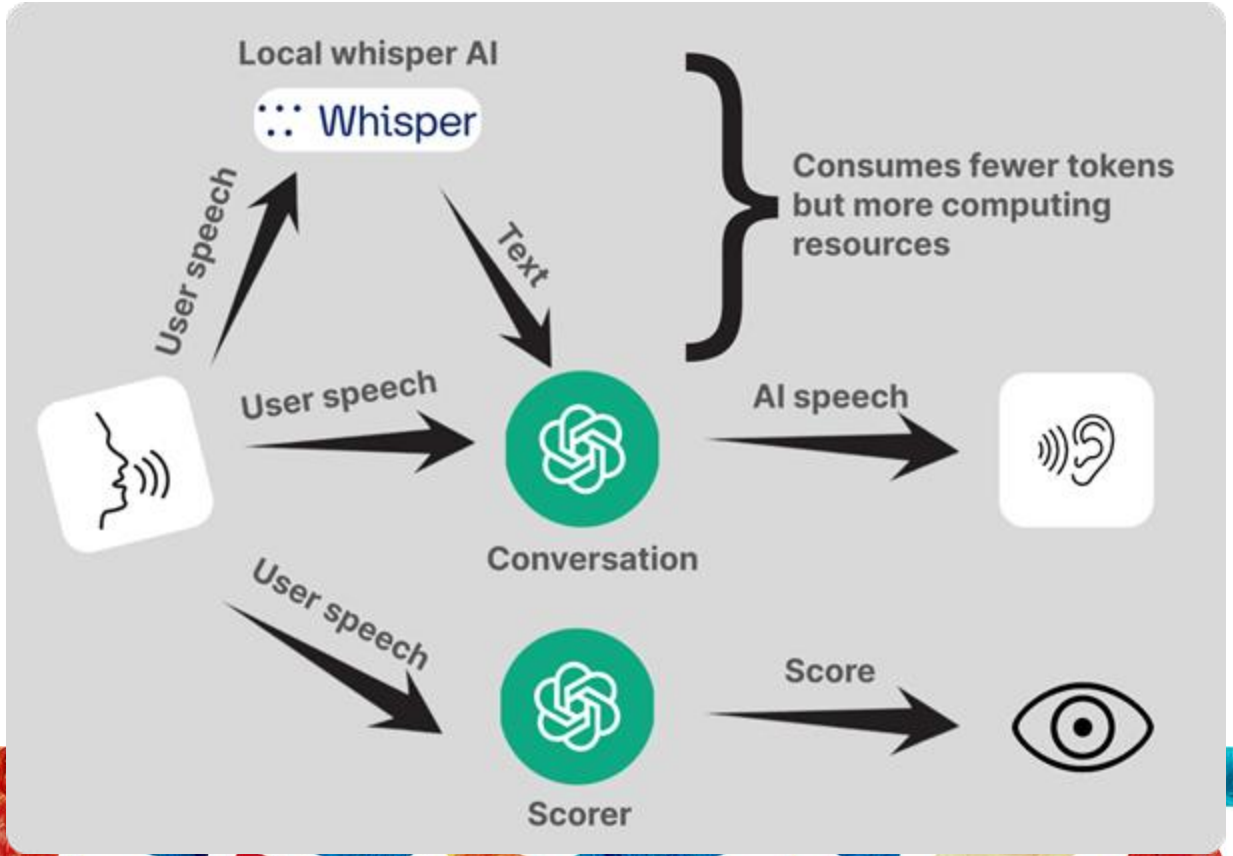
Relevance score

Elaboration on one error

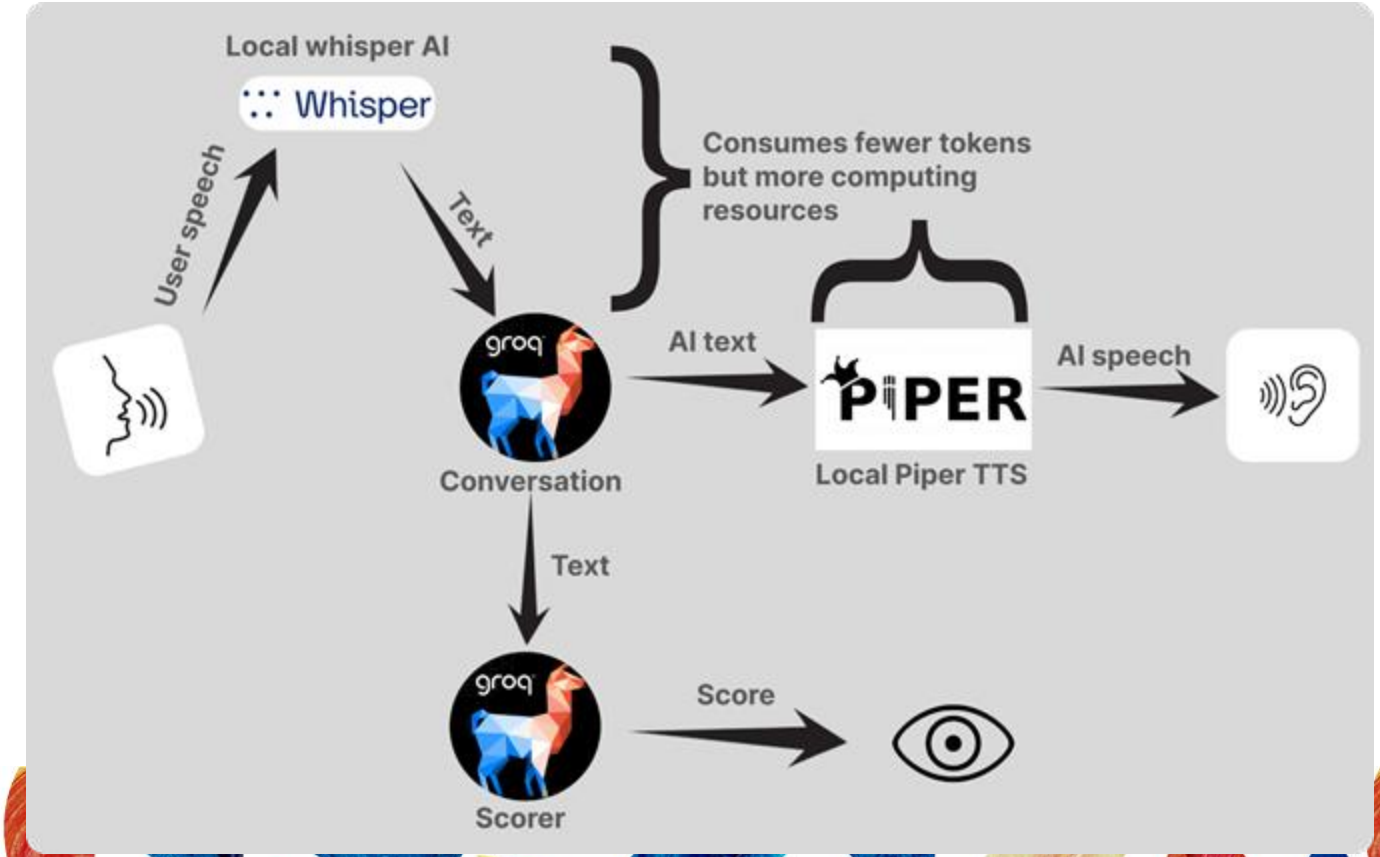
We have a system for converting feedback points into a flat score

Demo

M1- AI Architecture



M3 (Now) - AI Architecture

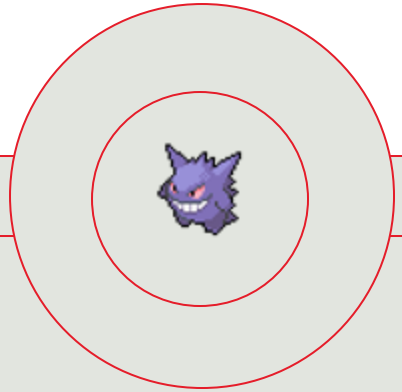


Summary

- Voxii is a VR language learning app designed to help users **improve conversational skills through immersive, real-world scenarios**.
- The app addresses literacy and language proficiency challenges by offering **dynamic, interactive experiences tailored to specific scenes and contexts**
- The project demonstrates **the potential of mixed reality and AI** to revolutionize educational tools

Pokemon TCG AR™

Role evolution



Josh Piscioneri - AI Dev

Owen Atkinson - Game/AR Dev

Archie Dorward - Lead Game Dev

Harrison Orosz - Game/AR Dev

Leo Barnes - IO Specialist

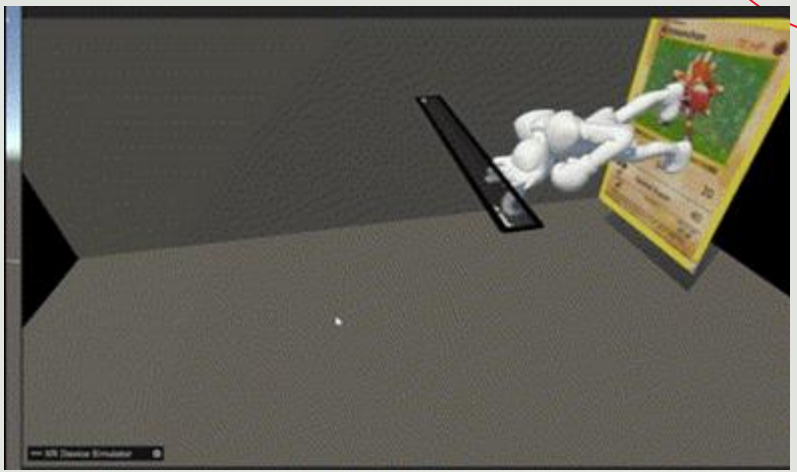
AR tracking

Concept:

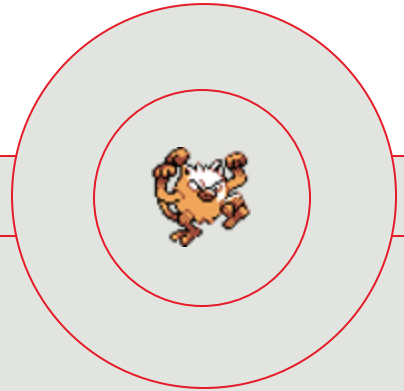
- Cards track individually
- Gamestate is determined by colliders placed on AR cards

Changes:

- Pivoted to a singled tracked marker
- Gamestate determines where virtual Pokemon show up

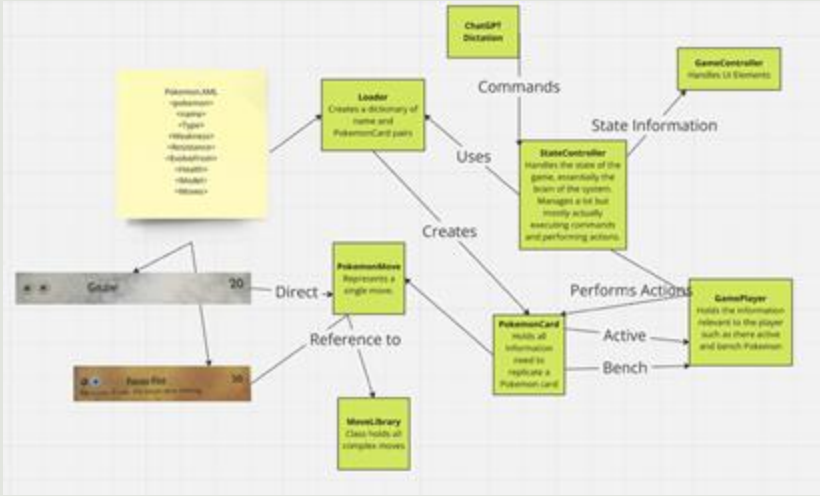


The Backend

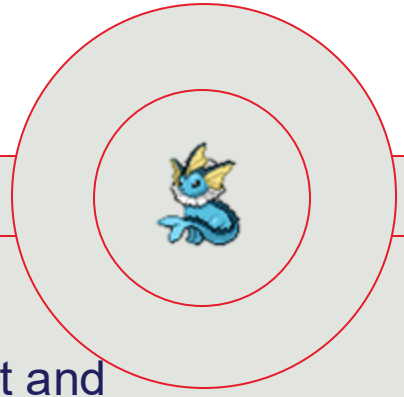


The backend was a very challenging to create;

- It ended up being quite large and complex.
- Had to be dependable.
- Keep track of even the smallest rules from the card game.
- Be memory efficient.
- Generic, robust methods that work with any pokemon and style of play.
- Making sure no data gets out of sync.
- Differences between running in the editor and the build



AI Dictation and processing



Dictation

- Originally whisper, pivoted to windows dictation
  - **Native support in MRTK**
  - **Less accurate**

Processing

- Dictation result and context given to chat GPT
  - **Chat GPT outputs the action it thinks the user wants**
  - **To make up for less accurate dictation, a dynamic prompt is used**

Prompt Generation

Dynamic Generation:

- Even with a lot of context in the prompt, Chat GPT was still confused. Dictation just wasn't good enough.
- Solution: give it as much of the correct answer as possible. This included giving it all possible commands that we know.

```
context += "Rival Player Active Pokemon: " + rival.active.name + "\n";
context += "Possible Commands for current player:\n";
foreach (var move in current.active.moves) {
    context += "{Attack " + move.name + ", " + current.active.name + ", " + rival.active.name + "}\n";
}
context += "{Energy, Fighting, " + current.active.name + "}";
context += "{Energy, Grass, " + current.active.name + "}";
context += "{Energy, Water, " + current.active.name + "}";
context += "{Energy, Fire, " + current.active.name + "}";
context += "{Energy, Psycic, " + current.active.name + "}";
context += "{Energy, Metal, " + current.active.name + "}";
context += "{Energy, Lightning, " + current.active.name + "}";
context += "{Energy, Darkness, " + current.active.name + "}";
context += "{Energy, Fairy, " + current.active.name + "}";

context += "{Retreat, " + current.active.name + ", <PokemonCard>}";
context += "{Summon, <PokemonCard>}";

se {
context += "Possible Commands:\n";
context += "{Energy, Fighting, " + current.active.name + "}";
context += "{Energy, Grass, " + current.active.name + "}";
context += "{Energy, Water, " + current.active.name + "}";
context += "{Energy, Fire, " + current.active.name + "}";
context += "{Energy, Psycic, " + current.active.name + "}";
context += "{Energy, Metal, " + current.active.name + "}";
context += "{Energy, Lightning, " + current.active.name + "}";
context += "{Energy, Darkness, " + current.active.name + "}";
context += "{Energy, Fairy, " + current.active.name + "}";
context += "{Summon, <PokemonCard>}";
```


Haunter

Other projects my team is currently looking at..

- AI Agent to Automate Asynchronous XR Tasks:
  - In this project, we aim to embed LLM-based AI agents into an asynchronous MR collaboration system to investigate how this approach can support remote collaborators in completing tasks together across time and space.
- Virtual Reality-Based Emotional Intelligence (EI) Intervention for Children with Autism Spectrum Disorder
  - The primary focus is on assessing improvements in emotion recognition, social interaction, and empathy. The VR training will feature scenarios simulating real-life social situations, enabling participants to practice emotional recognition, social interaction, and empathy in an AI+XR environment.
- Using Motion-free ML-trained Physiological Data as Superpower Input in Mixed Reality

What's Next

Tamil will focus on a lot of hardcore technical content in the afternoon sessions to get you onboard. Please start brainstorming you project ideas, scope, roles, and team members :P