

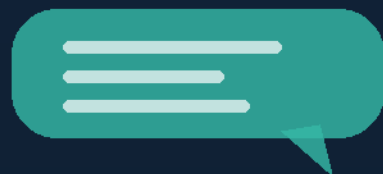
Machbarkeitsstudie zur KI-gestützten Erweiterung

mit Diversitätsbezug

Drei unabhängige Lösungsansätze unterschiedlicher Komplexität
für zukünftige Projektgruppen

Fokus

KI-gestützte Sprache, kontrollierte
Dialoginteraktion und virtueller Patient
als Erweiterungen des bestehenden
Unity-/VR-Lernspiels.



1. EINORDNUNG

Ziel und Rahmen der Machbarkeitsstudie

Die Studie untersucht, wie KI-gestützte Sprach- und Dialogfunktionen den Diversitätsbezug von Room of Horror sinnvoll erweitern können. Sie beschreibt bewusst keine fertige Implementierung, sondern technisch nachvollziehbare Lösungswege für eine zukünftige Projektgruppe.

Auszug aus der Zielvereinbarung

Prüfung und ggf. prototypische Integration KI-gestützter Funktionen, beispielsweise Sprachfunktion oder virtueller Patient, mit Bezug zu Diversitätsaspekten. Als Erfolgskriterium ist eine kurze Machbarkeitsstudie mit umrissenen Anwendungsbereichen vorgesehen.

Ausgangslage des Projekts

Room of Horror 2.0 erweitert das bestehende VR-Lernspiel bereits um diverse Patientenmodelle und patientenspezifische Fehler. Die Auswahl des Patienten erfolgt im React-Frontend; Unity aktiviert anschließend die passenden Modelle und Fehlerobjekte. Für KI-Funktionen sind vor allem die bestehende Unity-Anwendung, das TypeScript-/Express-Backend und die VR-Hardware relevant.

Framing der Studie

- Die drei Ansätze sind unabhängig voneinander umsetzbar und bauen technisch nicht aufeinander auf.
- Die Reihenfolge zeigt ausschließlich eine steigende technische Komplexität und einen zunehmenden Anteil von KI zur Laufzeit.
- Die Szenarien sind fachliche Beispiele. Vor einer Umsetzung müssen sie gemeinsam mit den Ansprechpartner:innen des Bundeswehrkrankenhauses geprüft und konkretisiert werden.
- OpenAI wird als mögliches API-Beispiel genannt; die Architektur kann grundsätzlich auch mit vergleichbaren Sprach- und KI-Diensten umgesetzt werden.

Zentrale Leitfrage

Wo erzeugt KI einen echten didaktischen Mehrwert gegenüber klassischen Popups, festen Antwortlisten oder rein visuellen Fehlerobjekten?

2. ÜBERBLICK

Drei unabhängige Lösungsansätze

Alle Ansätze adressieren Sprache und Kommunikation, unterscheiden sich jedoch deutlich darin, wann KI eingesetzt wird, wie offen die Interaktion ist und welche Unity-/Backend-Komponenten benötigt werden.

01

KI-generierte Tonspuren

NIEDRIGE KOMPLEXITÄT

KI erzeugt vorab Audioinhalte. Im Spiel werden feste Dateien abgespielt.

02

Kontrollierte Sprachinteraktion

MITTLERE KOMPLEXITÄT

Freie Spracheingabe wird erkannt und einer vordefinierten Reaktion zugeordnet.

03

Dynamischer virtueller Patient

HOHE KOMPLEXITÄT

Ein KI-Modell führt zur Laufzeit ein freies, rollenbasiertes Patientengespräch.

Vergleich auf einen Blick

Lösungsansatz	KI-Einsatz	Interaktion	Zentrale Unity-Konzepte	Komplexität
KI-generierte Tonspuren	vor der Laufzeit	feste Audio- und Auswahlsequenz	AudioClip, AudioSource, Trigger	niedrig
Kontrollierte Sprachinteraktion	zur Laufzeit für Erkennung	freie Sprache, feste Reaktionsklassen	Microphone, UnityWebRequest, UI/State	mittel
Dynamischer virtueller Patient	zur Laufzeit für Dialog	freies Gespräch	Streaming, Conversation Manager, Audio Playback	hoch

Wichtig

Die Ansätze können separat umgesetzt werden. Ansatz 3 verwendet beispielsweise nicht automatisch die in Ansatz 1 erzeugten Audiodateien.

3. LÖSUNGSANSATZ 1

KI-generierte Tonspuren

Die KI wird als Werkzeug zur Produktion von Sprachinhalten eingesetzt. Die erzeugten Audiodateien werden anschließend wie normale Unity-Assets in das Spiel integriert.

KOMPLEXITÄT: NIEDRIG

KI NICHT ZUR LAUFZEIT

OFFLINE SPIELBAR

Beispielszenario: Patient spricht kein Deutsch

Ein Patient äußert sein Anliegen ausschließlich in einer anderen Sprache. Eine vorab durch Text-to-Speech erzeugte Tonspur wird im Zimmer abgespielt. Die lernende Person erlebt die Sprachbarriere unmittelbar und muss anschließend eine geeignete Maßnahme auswählen. Die fachlich richtige Reaktion ist die Organisation eines professionellen Dolmetschers.

- Reinigungskraft rufen
- Kollegin oder Kollegen zur Übersetzung rufen
- Professionellen Dolmetscher organisieren

Diversity-Mehrwert

Die Sprachbarriere wird nicht nur als Text in der Patientenakte erwähnt, sondern hörbar und damit im VR-Szenario erfahrbar. Unterschiedliche Sprachen, Stimmen und Sprechweisen lassen sich mit vergleichsweise geringem Produktionsaufwand abbilden.

Konzeptionelle Unity-Roadmap

- 1 Fachliches Szenario und genaue Patientenaussage gemeinsam mit den Ansprechpartner:innen festlegen.
- 2 Audio über einen Text-to-Speech-Dienst erzeugen und als WAV, MP3 oder vergleichbares Format speichern.
- 3 Audiodatei als AudioClip in das Unity-Projekt importieren und einem AudioSource-Component am Patienten-Prefab zuordnen.
- 4 Abspielauslöser definieren, beispielsweise Szenariostart, Annäherung oder XR-Interaktion mit dem Patienten.
- 5 Popup oder vorhandene Fehlerinteraktion mit den Antwortmöglichkeiten verknüpfen.
- 6 Bei korrekter Auswahl den Fehlerstatus aktualisieren; bei falscher Auswahl den bestehenden Fehlversuch-Mechanismus nutzen.
- 7 Optional mehrere Sprachvarianten als konfigurierbare Assets hinterlegen, beispielsweise über ScriptableObjects oder eine JSON-/Backend-Zuordnung.

Benötigte Konzepte

- Text-to-Speech zur Erzeugung der Audiodateien
- Unity AudioClip und AudioSource für Import und Wiedergabe
- XR-Trigger beziehungsweise Collider/Interaktionsobjekt
- Anbindung an die vorhandene Fehler- und Fehlversuchlogik

Abgrenzung: Die KI ist bei diesem Ansatz nicht Bestandteil des laufenden Spiels. Sie unterstützt die Erstellung der Lerninhalte; zur Laufzeit spielt Unity vorbereitete Dateien ab.

4. LÖSUNGSANSATZ 2

Kontrollierte Sprachinteraktion

Die lernende Person formuliert frei. KI-gestützte Spracherkennung und Klassifikation ordnen die Aussage einer vorher definierten Kategorie zu; die Patientenantwort bleibt fachlich kontrolliert und vorbereitet.

KOMPLEXITÄT: MITTEL

KI ZUR AUSWERTUNG

KONTROLLIERTE ANTWORTEN

Beispielszenario: Verständliche Kommunikation bei eingeschränkten Deutschkenntnissen

Der Patient versteht grundsätzlich etwas Deutsch, hat jedoch Schwierigkeiten mit langen Sätzen und medizinischen Fachbegriffen. Damit die Sprachbarriere im Szenario nachvollziehbar ist, spricht er bewusst in einfachen, teilweise grammatikalisch unvollständigen Sätzen und verwendet nur einen begrenzten Wortschatz. Die lernende Person muss die bevorstehende Handlung frei, respektvoll und in verständlicher Sprache erklären - beispielsweise eine Blutdruckmessung.

Sprachliche Darstellung des Patienten

Die Patientenaussage sollte nicht in vollständig sicherem Standarddeutsch vertont werden. Der Text kann bewusst kurz und teilweise grammatikalisch unvollständig formuliert werden. Je nach verwendetem Text-to-Speech-Dienst lassen sich zusätzlich langsames Sprechen, Unsicherheit, Pausen oder ein fremdsprachiger Akzent anweisen. Alternativ können einzelne Wörter aus der Muttersprache ergänzt werden. Die erzeugten Varianten sollten vor der Integration angehört und fachlich geprüft werden, damit die Darstellung glaubwürdig, respektvoll und nicht überzeichnet wirkt.

Patient	„Entschuldigung, Deutsch nur bisschen. Blutdruck - was bedeutet? Was Sie machen?“
Gute Antwort	„Ich lege Ihnen eine Manschette um den Arm. Damit messe ich Ihren Blutdruck. Es tut nicht weh.“
Mögliche Reaktion	„Ja, jetzt verstehe ich. Sie messen am Arm.“

Diversity-Mehrwert

Im Unterschied zu Multiple Choice muss die lernende Person selbst eine verständliche Formulierung entwickeln. Trainiert werden einfache Sprache, respektvolle Ansprache, Vermeidung unnötiger Fachbegriffe und das Überprüfen, ob der Patient die Erklärung verstanden hat.

Was die KI bewertet

- Wurde die bevorstehende Handlung erklärt?
- Ist die Formulierung kurz und verständlich?
- Wurden unnötige medizinische Fachbegriffe vermieden?
- Wurde gegebenenfalls das Verständnis des Patienten überprüft, zum Beispiel durch eine Teach-back-Frage?

Konzeptionelle Unity-Roadmap

- 1 Push-to-Talk- oder Start-/Stopp-Interaktion in VR vorsehen und die Sprache über Unitys Microphone-API als AudioClip aufnehmen.
- 2 Aufnahme in ein für die API geeignetes Audioformat überführen und über UnityWebRequest an einen neuen Endpoint des bestehenden Express-Backends senden.
- 3 Im Backend die Aufnahme an einen Speech-to-Text-Dienst weiterleiten und das Transkript empfangen.
- 4 Das Transkript durch ein Sprachmodell oder eine kontrollierte Klassifikationslogik in definierte Kategorien einordnen, beispielsweise „verständlich erklärt“, „zu viel Fachsprache“ oder „Handlung nicht erklärt“.
- 5 Das Backend liefert ein strukturiertes Ergebnis an Unity zurück, idealerweise als klar begrenztes JSON-Objekt.

- 6 Unity wählt anhand der Kategorie eine vorbereitete Patientenantwort aus und spielt den zugehörigen AudioClip ab.
- 7 Gesprächsstatus und Lernerfolg in einem einfachen Zustandsmodell speichern und mit der vorhandenen Fehlerlogik verbinden.

Sicherheitsprinzip

API-Schlüssel gehören nicht in den Unity-Build oder auf die VR-Brille. Unity kommuniziert mit dem vorhandenen Backend; erst das Backend ruft den externen KI-Dienst auf.

5. LÖSUNGSANSATZ 3

Dynamischer virtueller Patient

Ein KI-Modell übernimmt zur Laufzeit eine klar definierte Patientenrolle und reagiert flexibel auf freie Fragen. Spracheingabe, Dialoggenerierung und Sprachausgabe bilden einen zusammenhängenden Gesprächsablauf.

KOMPLEXITÄT: HOCH

KI IM FREIEN DIALOG

ECHTZEIT-VERARBEITUNG

Beispielszenario: Asiatische Patientin spricht Beschwerden weniger direkt an

Der virtuelle Patient wird mit einem fachlich festgelegten Profil konfiguriert: Eine asiatische Patientin hat tatsächlich starke Schmerzen, äußert diese auf allgemeine Fragen jedoch weniger direkt. Erst durch gezieltes Nachfragen und die Verwendung einer strukturierten Schmerzskala wird die tatsächliche Schmerzstärke sichtbar.

Lernende Person	„Wie geht es Ihnen?“
Virtuelle Patientin	„Es geht schon. Es ist nicht so schlimm.“
Lernende Person	„Wenn null keine Schmerzen und zehn die stärksten vorstellbaren Schmerzen sind: Welche Zahl würden Sie nennen?“
Virtuelle Patientin	„Vielleicht eine sieben.“

Diversity-Mehrwert

Das Szenario trainiert nicht nur das Erkennen einer richtigen Auswahl, sondern das aktive Führen eines diversitätssensiblen Gesprächs: allgemeine Antworten hinterfragen, gezielt nach Beschwerden fragen und standardisierte Instrumente wie eine Schmerzskala einsetzen.

Konzeptionelle Unity-Roadmap

- 1 Patientenrolle, Lernziele, bekannte Fakten, Kommunikationsstil und unzulässige Aussagen als klar begrenztes Szenarioprofil definieren.
- 2 In Unity eine Gesprächsinteraktion und einen Conversation Manager vorsehen, der Start, Ende, aktuellen Dialogzustand und Abbruch/Fallback verwaltet.
- 3 Mikrofon-Audio als Stream oder in kurzen Segmenten erfassen und über eine WebSocket-/WebRTC-Verbindung beziehungsweise über das Backend an einen Realtime-Sprachdienst übertragen.
- 4 Das Backend erstellt und kontrolliert die KI-Sitzung, übergibt das Szenarioprofil und schützt Zugangsdaten sowie serverseitige Regeln.
- 5 Die KI erzeugt abhängig vom bisherigen Gespräch eine Patientenantwort als Text oder direkt als Audio.
- 6 Unity empfängt die Antwort, spielt sie möglichst verzögerungsarm über den AudioSource des Patienten ab und synchronisiert optional einfache Gesichts-/Sprechanimationen.
- 7 Eine separate Auswertungslogik prüft definierte Lernziele, beispielsweise gezieltes Nachfragen oder Nutzung der Schmerzskala, statt den Lernerfolg ausschließlich aus dem freien Dialog abzuleiten.

- 8 Fallbacks für Verbindungsabbrüche, unpassende Antworten und Szenario-Reset vorsehen.

Zentrale technische Anforderungen

- Dauerhafte oder gestreamte Audioverbindung mit niedriger Latenz
- Gesprächszustand und Rollenbeschreibung über mehrere Äußerungen hinweg
- Begrenzung des Modells auf das validierte Szenario
- Trennung von freiem Patientendialog und nachvollziehbarer Lernzielbewertung

Der Ansatz bietet die flexibelste und realistischste Interaktion, benötigt aber die stärkste technische und fachliche Absicherung. Ein vollständig freier medizinischer Chatbot ohne klar begrenztes Szenario ist für ein Lernspiel nicht empfehlenswert.

6. ANWENDUNGSBEREICHE

Wo KI einen Diversitäts-Mehrwert erzeugen kann

Die drei Lösungsansätze lassen sich auf verschiedene klinische Kommunikationssituationen übertragen. Entscheidend ist, dass die KI eine konkrete Lernhandlung ermöglicht und nicht nur als technischer Zusatz eingesetzt wird.

Sprachbarrieren

Patientenaussagen in einer anderen Sprache hörbar machen, Dolmetscherbedarf erkennen und professionelle Sprachmittlung thematisieren.

Einfache Sprache und Gesundheitskompetenz

Medizinische Abläufe verständlich erklären, Fachbegriffe vermeiden und das Verständnis des Patienten überprüfen.

Kulturell geprägte Kommunikation

Zurückhaltende oder indirekte Beschwerdeäußerungen erkennen und durch gezielte, strukturierte Rückfragen ergänzen.

Religiöse und persönliche Bedürfnisse

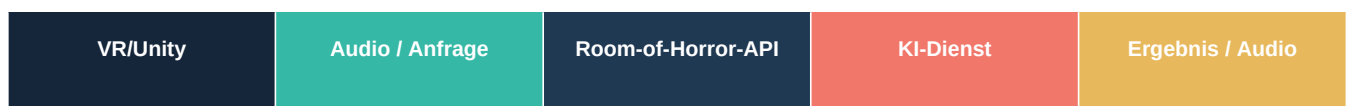
Gesprächssituationen zu Gebetszeiten, Privatsphäre, Ernährung oder geschlechtsspezifischen Präferenzen simulieren.

Kommunikation in belastenden Situationen

Unruhe, Angst oder Scham in einem kontrollierten Szenario abbilden und eine respektvolle Reaktion trainieren.

Gemeinsames Architekturprinzip

Für die beiden laufzeitbasierten Ansätze sollte das vorhandene Backend als kontrollierte Schnittstelle zwischen Unity und einem externen KI-Dienst dienen. Dadurch bleiben API-Zugangsdaten außerhalb des Unity-Builds und Szenarioregeln können zentral verwaltet werden.



- Ansatz 1 kann vollständig ohne Laufzeit-API auskommen.
- Ansatz 2 benötigt mindestens Aufnahme, Upload, Transkription und strukturierte Klassifikation.
- Ansatz 3 benötigt zusätzlich Echtzeitkommunikation, Gesprächszustand und dynamische Audioausgabe.

7. OFFENE FACHLICHE KLÄRUNG

Was vor einer Umsetzung abgestimmt werden muss

Die technischen Lösungswege sind grundsätzlich beschreibbar. Die medizinisch-pflegerische Qualität hängt jedoch von klar definierten und validierten Szenarien ab.

Lernziel

Welche konkrete Handlung oder Kommunikationskompetenz soll die lernende Person am Ende zeigen?

Patientenrolle

Welche Informationen kennt der Patient, was sagt er spontan und was erst nach gezielter Nachfrage?

Fachlich korrekte Reaktion

Welche Antworten oder Handlungen gelten als korrekt, teilweise korrekt oder falsch?

Diversity-Framing

Wie wird der kulturelle, sprachliche oder religiöse Bezug fachlich präzise und ohne missverständliche Verallgemeinerungen formuliert?

Bewertung

Soll das System nur Feedback geben, einen Fehlerstatus setzen oder eine nachvollziehbare Punktbewertung erzeugen?

Datenschutz und Betrieb

Dürfen Sprachaufnahmen an einen externen Dienst übertragen werden, müssen sie sofort gelöscht werden und ist für den Einsatz eine Internetverbindung verfügbar?

Fazit

KI kann Room of Horror vor allem dort sinnvoll erweitern, wo Kommunikation, Sprache und individuell unterschiedliche Reaktionen nicht ausreichend durch statische Fehlerobjekte oder Multiple-Choice-Popups abgebildet werden. Die drei untersuchten Lösungsansätze sind unabhängig voneinander umsetzbar und decken unterschiedliche Komplexitätsniveaus ab:

- KI-generierte Tonspuren ermöglichen einen vergleichsweise einfachen Einstieg und machen Sprachbarrieren hörbar.
- Kontrollierte Sprachinteraktion trainiert die freie, verständliche Kommunikation bei weiterhin reproduzierbaren Patientenantworten.
- Ein dynamischer virtueller Patient ermöglicht flexible Gespräche und den größten didaktischen Spielraum, erfordert aber die umfangreichste technische und fachliche Kontrolle.

Gesamteinschätzung

Alle drei Ansätze sind als Erweiterungen des vorhandenen Unity-/Backend-Systems grundsätzlich realisierbar. Welcher Ansatz umgesetzt wird, sollte anhand des gewünschten Lernziels, der verfügbaren Entwicklungszeit und der notwendigen fachlichen Kontrolle entschieden werden.

8. TECHNISCHE REFERENZEN

Verwendete technische Grundlagen

Die folgenden offiziellen Dokumentationen stützen die beschriebenen Architektur- und Unity-Konzepte. Produktnamen dienen als Beispiele; die Lösungsansätze sind nicht auf einen einzelnen Anbieter beschränkt.

OpenAI Audio- und Sprachfunktionen

Speech generation / Text to speech

<https://developers.openai.com/api/docs/guides/text-to-speech>

Speech to text

<https://developers.openai.com/api/docs/guides/speech-to-text>

Realtime and audio

<https://developers.openai.com/api/docs/guides/realtime>

Best practices for API key safety

<https://help.openai.com/en/articles/5112595-best-practices-for-api-key-safety>

Unity Audio- und Netzwerkfunktionen

Microphone API

<https://docs.unity3d.com/6000.3/Documentation/ScriptReference/Microphone.html>

AudioClip

<https://docs.unity3d.com/ScriptReference/AudioClip.html>

UnityWebRequestMultimedia

<https://docs.unity3d.com/6000.3/Documentation/ScriptReference/Networking.UnityWebRequestMultimedia.html>

Hinweis zum Stand

APIs, Modelle und Unity-Versionen können sich ändern. Vor einer späteren Implementierung sollten die jeweils aktuellen offiziellen Dokumentationen und Datenschutzerfordernungen erneut geprüft werden.