

# GuudeVR: VR-gestützte Lernszenarien zum Sprachtraining basierend auf generativer KI

Andreas Fuchs <sup>1</sup>, Sven Appel <sup>2</sup> und Paul Grimm <sup>3</sup>

**Abstract:** In diesem Beitrag stellen wir *GuudeVR* vor, eine Virtual Reality Anwendung, die generative Künstliche Intelligenz verwendet, um das Sprachenlernen in immersiven und authentischen Umgebungen zu unterstützen. Es simuliert Szenarien in verschiedenen Situationen, in denen Lernende in einer gewünschten Sprache interagieren und so ihre Sprachkenntnisse üben können. Das System nutzt Spracherkennung, natürliche Sprachverarbeitung und Computervision, um adaptive und personalisierte Dialoge zu erstellen. Wir präsentieren das Design und die Umsetzung, um den Spracherwerb zu fördern.

**Keywords:** Virtual Reality, Language Learning, Generative Künstliche Intelligenz, Speech-to-Text, Text-to-Speech

## 1 Einführung


Eine der Herausforderungen beim Sprachenlernen besteht darin, Lernenden authentische und immersive Situationen bereitzustellen, die reale Umgebungen und Interaktionen simulieren sowie eine personalisierte Lernerfahrung bieten. Virtual Reality (VR) ist eine Technologie, die solche Situationen erfahrbar machen kann, indem Lernende in eine dreidimensionale Umgebung eintauchen, die auf ihre Aktionen und Eingaben reagiert [Dö16]. Die Erstellung von VR-Anwendungen ist jedoch aufwendig und folgt meist statischen Szenarien. Vor diesem Hintergrund ist *GuudeVR* eine VR-Anwendung, die generative Künstliche Intelligenz (KI) verwendet, um das Sprachenlernen zu unterstützen und zu personalisieren, ohne in manueller Arbeit für neue Situationen jeweils neue Übungsbeispiele erstellen zu müssen. Dies ermöglicht es, individuelle Lerninhalte schnell und kostengünstig bereitzustellen, die an den Wissensstand der Lernenden angepasst sind und verschiedene Sprachen abbilden können.


## 2 Verwandte Arbeiten

Die Integration von KI und VR in das Sprachenlernen zeigt vielversprechende Ergebnisse. In “The 2 Sigma Problem” von Bloom aus 1984, wurde festgestellt, dass personalisierter

---

1 Hochschule Darmstadt, Media, Max-Planck-Straße 2, 64807 Dieburg, Germany, andreas.fuchs@h-da.de,  <https://orcid.org/0000-0003-1477-9556>

2 Hochschule Darmstadt, Informatik, Birkenweg 7, 64295 Darmstadt, Germany, sven.appel@h-da.de,  <https://orcid.org/0009-0003-7796-6970>

3 Hochschule Darmstadt, Media, Max-Planck-Straße 2, 64807 Dieburg, Germany, paul.grimm@h-da.de,  <https://orcid.org/0000-0003-4189-2642>

Einzelunterricht zu signifikant besseren Lernergebnissen führt [Bl84]. KI bietet hier durch maschinelles Lernen die Möglichkeit, personalisierten Unterricht skalierbar zu machen und somit die Vorteile von eins zu eins Lernen breiter zugänglich zu machen [VJ19]. Die Rolle von VR im Sprachunterricht wurde in einer Übersicht von 26 wissenschaftlichen Arbeiten analysiert. VR schafft immersive Lernumgebungen, die das Sprachlernen interaktiver und ansprechender gestalten, birgt jedoch Herausforderungen hinsichtlich technischer und pädagogischer Aspekte [Pa23]. Empirische Studien zu mobilen Text-to-Speech (TTS) und automatischer Spracherkennung (ASR) zeigen, dass diese Technologien die Aussprache und Sprachproduktion in einer mobilen Lernumgebung verbessern. Sie bieten Lernenden besseren Zugang zu Sprachinput und fördern ihre Autonomie [LCL17]. KI und Extended Reality (XR) im Fremdsprachenunterricht bieten immersive Lernmethoden ähnlich internationalen Immersionsprogrammen. Die Cognitive Immersive Language Learning Environment (CILLE) ermöglicht natürliche, multi-modale Konversationsinteraktionen. Eine siebenwöchige Studie mit Universitätsstudierenden zeigt signifikante Verbesserungen im Chinesischen als Fremdsprache, was die Effektivität von CILLE als Lernwerkzeug unterstreicht [Di22]. Der Einsatz von KI in der Fehlerkorrektur beim Zweitspracherwerb, wie in der Software “Intelligent Tutor”, zeigte eine signifikante Reduktion der Fehlerrate bei Schreibaufgaben. Dies unterstreicht die Effektivität von KI-gestützter Fehlerkorrektur im Sprachunterricht [Do07a; Do07b].

### **3 Konzept von GuudeVR**

*GuudeVR* nutzt die Leistungsfähigkeit der natürlichen Sprachgenerierung und des natürlichen Sprachverständnisses, um Lernenden variabel Sprachinhalte und Feedback bereitzustellen. *GuudeVR* ermöglicht es, Lernende in vier verschiedenen Szenarien ihre Sprachkenntnisse zu fördern und zu prüfen. Um dies so realitätsnah wie möglich zu gestalten, wird dabei auf die Aspekte Hören, Sprechen und Lesen eingegangen. Die Szenarien umfassen u.a. ein interaktives Kochstudio, das Beschreiben von Bildern in einer Bildergalerie, Chorales Wiederholen, sowie Szenarien für das freie Reden. Durch die Verwendung von KI löst sich *GuudeVR* von einer linearen und unflexiblen Lernumgebung und ermöglicht dennoch eine weitestgehend deterministische Erfahrung für das Lernen von Sprachen.

#### **3.1 Stärken des Ansatzes**

Der Ansatz von *GuudeVR* optimiert das Sprachenlernen durch eine VR-Umgebung, die Lernende aktiv in interaktive Szenarien einbindet. Ein zentrales Merkmal ist die Möglichkeit, dynamisch Inhalte zu ergänzen, um die Anwendung adaptiv anzupassen. Generische Lernszenarien sprechen verschiedene Sprachniveaus und Methoden an, wie Hören, Nachsprechen, freies Unterhalten sowie Lesen und Interagieren, was Lernende zum Üben anregt und somit das Sprachverständnis fördern soll. KI-Komponenten bieten direktes Feedback, helfen, Fehler schnell zu erkennen und zu korrigieren, und unterstützen spontane Gespräche, wodurch

das Selbstvertrauen gestärkt wird. Visuelle, auditive und haptische Elemente verbessern die Lerneffektivität, während spielerische Elemente wie Zielerreichungen die Motivation und das Engagement steigern. Schließlich passt die KI Lerninhalte und Sprachniveaus individuell an, um eine individuelle Lernerfahrung zu bieten.

### 3.2 Lernszenario 1: Interagieren

Im Lernmodul *Interagieren* wird “gemeinsam gekocht”. Der Fokus liegt auf dem Hören, Lesen und Handeln, um ein praxisnahes Sprachlernerlebnis zu bieten. Ein Avatar agiert als Kochlehrer:in und gibt detaillierte Anweisungen zur Zubereitung eines Rezepts (Abbildung 1 (links)). Lernende müssen die Anweisungen verstehen und die Zutaten in der richtigen Reihenfolge kombinieren, um das Rezept erfolgreich abzuschließen. Dabei erfassen sie die Zutaten durch aktives Zuhören, finden diese auf einem Tisch durch Sehen und Lesen der Zutatenbezeichnungen und geben diese anschließend durch Handeln in eine Rührschüssel. Das Lernszenario simuliert typische Alltagssituationen und nutzt visuelle, auditive und haptische Elemente zur Vermittlung der Lerninhalte. Spielerische Elemente motivieren die Lernenden und steigern ihr Engagement. Der Erfolg beim Kochen dient als direktes Feedback über den Lernfortschritt und stärkt das Selbstvertrauen.



Abb. 1: Lernszenario 1 *Interagieren*: Zutaten müssen hier für das Backen eines Kuchens in der richtigen Reihenfolge zusammengefügt werden (links) sowie Lernszenario 2 *Beschreiben*: Durch kreative und interaktive Bildbeschreibungen wird das Sprachgefühl gefördert (rechts).

### 3.3 Lernszenario 2: Beschreiben

Im Lernmodul *Beschreiben* befinden sich die Lernenden in einer “Bildergalerie”. Der Fokus liegt auf dem Sehen und freien Beschreiben des Gesehenen. Lernende beschreiben Bilder in verschiedenen Detailgraden (je nach Schwierigkeit) und entwickeln dabei eigene Ideen, wie sie das Gesehene mit ihrem Wortschatz ausdrücken können. Ein Avatar gibt Feedback, welche Elemente gut erkannt und beschrieben wurden, und bietet Hinweise, wie bestimmte Dinge sprachlich detaillierter und bildhafter beschrieben werden können. Je nach Schwierigkeitsstufe müssen die Lernenden ihre Beschreibungen wiederholen und verbessern

oder können mit dem nächsten Bild fortfahren. Die virtuelle Bildergalerie (Abbildung 1 (rechts)) fördert präzise und detaillierte Beschreibungen, erweitert den Wortschatz und verbessert die sprachliche Flexibilität. Visuelle Reize unterstützen die Wahrnehmung, während das mündliche Beschreiben und das Feedback die auditive Verarbeitung fördern. Direktes Feedback hilft bei der Fehlerkorrektur und steigert die sprachliche Präzision. Spielerische Elemente motivieren die Lernenden und fördern die aktive Auseinandersetzung mit der Sprache, was zu einem tieferen Verständnis und einer verbesserten Beherrschung führt.

### **3.4 Lernszenario 3: Unterhalten**

Im Lernmodul *Unterhalten* müssen sich die Lernenden zu einem Kinobesuch verabreden. Der Fokus liegt auf dem Führen eines freien Dialoges mit einem Avatar, um sich zu einem gemeinsamen Kinobesuch zu verabreden. Der Avatar eröffnet das Gespräch und fordert die Lernenden zum Antworten auf. Die KI wird so geprimed, dass sie keine geschlossenen Fragen stellt, sodass Lernende in ganzen Sätzen antworten müssen, ausformulierte Antworten geben und eigene Entscheidungen treffen. Um den Dialog komplexer zu gestalten, kann es vorkommen, dass der Avatar Vorschläge bezüglich eines Films ablehnt oder um einen anderen Tag oder eine andere Uhrzeit bittet, da er ansonsten nicht teilnehmen kann. Diese nicht linearen Gesprächsverläufe trainieren das situative Reagieren und Adaptieren der Lernenden. Außerdem erhöht dies die Wiederholbarkeit der Lernszenarien, da jeder Gesprächsverlauf anders verlaufen kann. Diese Methode fördert die sprachliche Flexibilität und das freie Sprechen in realistischen Situationen. Lernende müssen spontan reagieren, Entscheidungen treffen und ihre Sprache anpassen, was zu einem tieferen Verständnis und einer verbesserten Beherrschung der Sprache führt.

### **3.5 Lernszenario 4: Nachsprechen**

Im Lernmodul *Nachsprechen* werden die Lernenden jeweils dazu aufgefordert, einen vom Avatar vorgesprochenen Satz korrekt nachzusprechen. Thematisch ist dies in der Situation eines Cafés angesiedelt, sodass übliche Sätze wie beispielsweise zum Bestellen eines Kaffees geübt werden. Lernende erhalten ein direktes Feedback und müssen einen gewissen Prozentsatz an Übereinstimmung erreichen, um mit der nächsten Satz wiederholung fortfahren zu können. Dieses Modul trainiert das Hören und Nachsprechen, ohne dass über eigene Formulierungen nachgedacht werden muss.

## **4 Umsetzung**

Die technische Umsetzung von *GuudeVR* basiert auf der Integration mehrerer Technologien, um eine dynamische und adaptive Sprachlern-Anwendung zu schaffen. Die folgenden

Technologien werden verwendet: **ChatGPT** [Op24] wird eingesetzt, um interaktive Dialoge und realistische Konversationsszenarien zu erstellen. Es generiert Antworten basierend auf den (Sprach-) Eingaben der Lernenden und hilft, den Lernprozess durch sofortiges Feedback zu unterstützen. Die Antworten von ChatGPT werden im JSON-Format generiert, was es ermöglicht, den Fortschritt einzelner Übungen zu verfolgen und schrittweise abzuarbeiten. Die generierten JSON-Dateien enthalten zusätzliche Informationen wie Bewertungskriterien und Sprachfeedback, die innerhalb der Anwendung ausgewertet und verarbeitet werden können. **Whisper** [Ra22] wird verwendet, um gesprochene Sprache in Text zu transkribieren. Dies ermöglicht es der Anwendung, die gesprochenen Eingaben der Lernenden zu verstehen und zu verarbeiten. Die transkribierten Texte werden dann mit entsprechenden Anweisungen an ChatGPT weitergeleitet, um gezielte Antworten und Auswertungen zu generieren und den Lernprozess voranzutreiben. **TTS** (Text-to-Speech) [Wa23] von OpenAI generiert die Sprachwiedergabe basierend auf den Texten, die situativ von der Anwendung zur Verfügung gestellt werden. Dies stellt sicher, dass Lernende eine natürliche und verständliche Sprachausgabe erhalten, die ihnen hilft, ihre Aussprache und ihr Hörverständnis zu verbessern.

Die Technologien arbeiten nahtlos zusammen, um eine interaktive und immersive Lernumgebung zu schaffen. Ein beispielhafter Ablauf sieht folgendermaßen aus:

1. Lernende sprechen einen Satz, der von *Whisper* in Text transkribiert wird.
2. Der transkribierte Text wird an *ChatGPT* gesendet, das eine passende Antwort generiert.
3. Die generierte Antwort wird durch die *TTS*-Technologie in gesprochene Sprache umgewandelt und den Lernenden zurückgegeben.
4. Das gesamte Interaktionsprotokoll wird im JSON-Format gespeichert, um den Fortschritt und die Bewertungen zu verfolgen.

Von den Entwickelnden beschriebene JSON-Dateien können dynamisch geladen werden, um Lernszenarien individuell anzupassen und mit neuen, bedarfsgerechten Inhalten zu füllen. Dies ermöglicht es Lehrpersonen, Lerninhalte nachträglich anzupassen, zu erweitern oder zu ergänzen, ohne die Anwendung selbst zu verändern. Die JSON-Konfigurationsdatei steuert den Ablauf und die Inhalte der Szenarien und enthält die erforderlichen Rezepte, Anweisungen und Dialoge für verschiedene Lernkontexte wie gemeinsames Kochen, Bildergalerie, Café-Szenario und freies Sprechen. Diese dynamische und adaptive Struktur macht *GuudeVR* zu einem flexiblen und effektiven Werkzeug für das Sprachenlernen, das sich kontinuierlich an die Bedürfnisse und Fortschritte der Lernenden anpasst.

Visuelles Feedback in *GuudeVR* wird durch verschiedene Methoden gegeben, um die Lernerfahrung zu verbessern. Dazu gehören Fortschrittsanzeigen (Abbildung 2 (links)) im Lernszenario und farbliches Feedback nach dem Ampelsystem (rot, orange, grün). Dieses non-verbale Feedback hilft den Lernenden, ihren Fortschritt und ihre Leistung schnell und intuitiv zu verstehen. Zusätzlich wird über dem Avatar eine Ladeanzeige (Abbildung 2

(rechts)) dargestellt, um Verzögerungen, die durch das Analysieren und Generieren von Sprach- und Bildinhalten entstehen, abzubilden. Diese visuelle Rückmeldung gibt den Lernenden die Sicherheit, dass die Anwendung korrekt funktioniert und vermeidet den Eindruck, dass die Anwendung nicht reagiert. Zukünftige Verbesserungen zielen darauf ab, diese Ladezeiten zu reduzieren. Dies wird durch das Streamen der Audioinhalte und den Einsatz schnellerer KI-Modelle ermöglicht, was die Lernerfahrung weiter optimiert. So kann die Anwendung effizienter und verzögerungsfreier arbeiten, was den Lernenden eine angenehmere Nutzung ermöglicht.



Abb. 2: Übungsfortschritt Indikator, welcher den Anwendenden den aktuellen Stand der Übung anzeigt (links) sowie Ladeindikator, welcher die Sprachverarbeitung anzeigt (rechts).

## 5 Zusammenfassung

Aktuell verwenden wir *GuudeVR*, um internationalen Studierende beim Spracherwerb zu unterstützen. Es kombiniert VR und KI, um eine immersive und interaktive Sprachlernumgebung zu schaffen. Entwickelt in Unity unter Verwendung des Meta Frameworks für VR, bietet die Anwendung unterschiedliche Szenarien für Interaktionen, für Bildbeschreibungen, für das Hören und Wiedergeben sowie für das freie Sprechen. Während der Verwendung erhalten die Lernenden Anweisungen von einem Avatar und führen interaktive Aufgaben aus, um Hörverständnis und die sprachliche Umsetzung zu verbessern. Hierzu gibt die Anwendung Feedback zur Richtigkeit und Angemessenheit der Sprache. In Zukunft werden diese Ansätze gemeinsam mit internationalen Studierenden evaluiert, um die Anwendung iterativ zu verbessern. Dabei wird die Wirksamkeit von interaktiven Übungen wie das Bestellen von Speisen, das Führen von Gesprächen und das gemeinsame Kochen für die Förderlichkeit von Sprachfertigkeiten und Sprachverstehen betrachtet.

## 6 Danksagung

Diese Arbeit wird gefördert durch die Stiftung für Innovation in der Hochschullehre, Programm Freiraum 23. Wir bedanken uns bei unseren Kollegen und Partnern aus dem Projekt *DaF2L* [LMU24] für ihre Zeit und wertvollen Kommentare.

## Literaturverzeichnis

- [Bl84] Bloom, B.: The 2 Sigma Problem: The Search for Methods of Group Instruction as Effective as One-to-One Tutoring. *Educational Researcher* 13 (6), S. 4–16, 1984.
- [Di22] Divekar, R. R.: Foreign language acquisition via artificial intelligence and extended reality: design and evaluation. *Computer Assisted Language Learning* 35 (9), S. 2332–2360, 2022.
- [Do07a] Dodigovic, M.: Artificial Intelligence and Second Language Learning: An Efficient Approach to Error Remediation. *Language Awareness* 16 (2), S. 99–113, 2007.
- [Do07b] Dodigovic, M.: Artificial Intelligence and Second Language Learning: An Efficient Approach to Error Remediation. *Language Awareness - LANG AWARE* 16, S. 99–113, 2007.
- [Dö16] Dörner, R.; Broll, W.; Grimm, P.; Jung, B.: *Virtual und Augmented Reality (VR/AR): Grundlagen und Methoden der Virtuellen und Augmentierten Realität*. Springer Vieweg Berlin, Heidelberg, 2016.
- [LCL17] Liakin, D.; Cardoso, W.; Liakina, N.: Mobilizing Instruction in a Second-Language Context: Learners' Perceptions of Two Speech Technologies. *Languages* 2 (3), 2017.
- [LMU24] Ludwig-Maximilians-Universität München: DaF2L: DaF lehren - DaF lernen: Qualifikation von Online-Tutorinnen und Tutoren einschließlich der Entwicklung von VR- und XR-Lernwelten, 2024, URL: [https://www.daf.uni-muenchen.de/personen/wiss\\_ma/springer/projekte/daf2l\\_freiraum23/index.html](https://www.daf.uni-muenchen.de/personen/wiss_ma/springer/projekte/daf2l_freiraum23/index.html), Stand: 30.07.2024.
- [Op24] OpenAI et al.: GPT-4 Technical Report, 2024, arXiv: 2303.08774 [cs.CL].
- [Pa23] Parmaxi, A.: Virtual reality in language learning: a systematic review and implications for research and practice. *Interactive Learning Environments* 31 (1), S. 172–184, 2023.
- [Ra22] Radford, A.; Kim, J. W.; Xu, T.; Brockman, G.; McLeavey, C.; Sutskever, I.: Robust Speech Recognition via Large-Scale Weak Supervision, 2022, arXiv: 2212.04356 [eess.AS].
- [VJ19] van der Vorst, T.; Jelcic, N.: Artificial Intelligence in Education: Can AI bring the full potential of personalized learning to education? 2019.
- [Wa23] Wang, C.; Chen, S.; Wu, Y.; Zhang, Z.; Zhou, L.; Liu, S.; Chen, Z.; Liu, Y.; Wang, H.; Li, J.; He, L.; Zhao, S.; Wei, F.: Neural Codec Language Models are Zero-Shot Text to Speech Synthesizers, 2023, arXiv: 2301.02111 [cs.CL].