

**BACHELORARBEIT**

# **KI im Studium**

## **Untersuchung zur**

### **Effektivität von Digital Humans als**

#### **virtuelle Assistentin für Studierende**

---

vorgelegt am 26. November 2023  
Stefan Kulow, [REDACTED]

Erstprüfer: Prof. Dr. Eike Langbehn  
Zweitprüferin: Prof. Dr. Larissa Putzar

---

**HOCHSCHULE FÜR ANGEWANDTE  
WISSENSCHAFTEN HAMBURG**  
Department Medientechnik  
Finkenau 35  
22081 Hamburg

# Zusammenfassung

Schlagwörter wie KI, Digital Humans, oder Metaverse sind in unserer heutigen Zeit allgegenwärtig. In vielen verschiedenen Bereichen werden bereits entsprechende Anwendungen, wie beispielsweise OpenAI's ChatGPT, zur Unterstützung menschlicher Tätigkeiten verwendet.

In dieser Arbeit geht es darum zu untersuchen, wie solche Technologien genutzt werden können, um Studierende im Alltag zu begleiten und dabei zu helfen, konkrete Fragen bezüglich des Studiums zu beantworten. Mit Hilfe eines Embodied Conversational Agents soll hierbei die Möglichkeit geschaffen werden, mit einer virtuellen Person interagieren und auf Basis eines Gespräches Fragen klären zu können.

Im Rahmen dieser Untersuchung soll evaluiert werden, wie gut sich eine solche Anwendung, unter Betrachtung verschiedener Aspekte, in der Praxis eignet und ob es sich dabei um eine sinnvolle Erweiterung für den Studienalltag handelt.

## Abstract

Buzzwords like AI, digital humans, or the metaverse are omnipresent nowadays. In many different areas such applications, like openai's ChatGPT are already used to support human work.

This work is about analyzing how such technologies can be used to support students in their daily life and help them to answer certain questions about the study itself. With the help of embodied conversational agents there shall be the possibility created to interact with a virtual person to clarify questions on the basis of a conversation.

In the context of this analysis there shall be evaluated how well such an application under consideration of multiple aspects is suitable for the practical use and if its a reasonable addition for the students daily life.

# Inhaltsverzeichnis

<b>Abbildungsverzeichnis</b>	<b>III</b>
<b>Tabellenverzeichnis</b>	<b>IV</b>
<b>1 Einleitung</b>	<b>1</b>
1.1 Begriffserklärung . . . . .	1
1.2 Problemstellung . . . . .	2
1.2.1 Ausgangssituation . . . . .	2
1.2.2 Motivation . . . . .	2
1.2.3 Zielsetzung . . . . .	2
<b>2 Verwandte Arbeiten</b>	<b>3</b>
<b>3 Projektdurchführung</b>	<b>7</b>
3.1 Rahmenbedingungen . . . . .	7
3.1.1 Software . . . . .	7
3.1.2 Hardware . . . . .	8
3.2 Prototyp . . . . .	9
3.2.1 Anforderungen . . . . .	9
3.2.2 Lösungskonzept / Design . . . . .	10
3.2.3 Proof of Concept . . . . .	11
3.2.4 Stand des Prototypen . . . . .	13
3.2.5 Weiterentwicklungsmöglichkeiten / Zukunftsaussicht . . . . .	15
3.3 Methodik . . . . .	16
3.4 Statistiken . . . . .	17
3.5 Probleme . . . . .	20
3.5.1 Halluzinationen . . . . .	21
3.5.2 Uncanny Valley Effekt . . . . .	22
3.5.3 Human Computer Interface . . . . .	24
<b>4 Ergebnisse</b>	<b>25</b>
<b>5 Fazit</b>	<b>29</b>

**Literatur**

**30**

**Anhang**

**32**

# Abbildungsverzeichnis

2.1	Darstellung IPSofts Amelia . . . . .	4
2.2	Darstellung Virtuelle Assistenten in AR . . . . .	5
3.1	Abbildung: Überblick des Systemaufbaus . . . . .	12
3.2	Abbildung: Aussehen des Prototyps . . . . .	14
3.3	Abbildung: Timer Funktion der Spracherkennung . . . . .	15
3.4	Statistik: Anzahl Nutzer*innen von digitalen Assistenten . . . . .	18
3.5	Statistik: Bevorzugte Stimme . . . . .	18
3.6	Statistik: Erwartungen zu Sprachassistenten . . . . .	19
3.7	Statistik: Gründe zur Ablehnung von Sprachassistenten . . . . .	20
3.8	Statistik: Nutzung zur Informationsausgabe . . . . .	20
3.9	Grafik: Umgang mit KI Tools . . . . .	22
3.10	Grafik: Uncanny Valley . . . . .	23
4.1	Grafik: Einordnung SUS Score . . . . .	28
1	Abbildung: Aufbau Fragebogen . . . . .	34

# Tabellenverzeichnis

3.1	Software . . . . .	8
3.2	Hardware . . . . .	9
4.1	Scores der Probanden*/innen . . . . .	25
4.2	Durchschnittliche Punkte pro Aussage . . . . .	26

# 1 Einleitung

## 1.1 Begriffserklärung

Der Begriff “Digital Human” ist in der Literatur weit gefächert, während verschiedene Quellen diesen, oder alternativ auch “Digital Human Modelling” verwenden, sprechen andere Quellen von “Virtual Humans” bzw. “Virtual Human Modelling”, angesichts dessen ist es zunächst erforderlich, eine Definition der Begrifflichkeit für den Kontext dieser Arbeit herauszustellen.

Beide Begriffe werden hierbei häufig als Synonym füreinander verwendet (Richard C. Ward, 2008) und beschreiben die virtuelle Repräsentation bzw. Simulation eines Menschen in Bezug auf Erscheinung, Physiologie, Verhalten, als auch die Möglichkeit zur Interaktion mittels multimodaler Interfaces. Solche Simulationen erfolgen dabei in einem virtuellen Raum, in dem, nach Nadia und Daniel Thalmann, Virtual Humans die Schlüssel Technologie darstellen, um Aufgaben in verschiedenen Rollen, so zum Beispiel als virtuelle Moderatoren, Guides oder Schauspieler, übernehmen zu können (Nadia Magnenat-Thalmann, 2004). Typische Charakteristika für Virtual Humans sind dabei die menschliche bzw. menschenähnliche Verkörperung durch ein Charaktermodell, sowie an Menschen angelehnte Verhaltensmuster wie Sprache, Emotionen, Gesten und Bewegungen von Kopf, Augen und anderen Körperteilen (David Burden, 2019). Ein weiterer Begriff, der in der Fachliteratur häufiger vorkommt und zumeist mit dem technischen Hintergrund dieser Thematik zusammenhängt ist der Term “Embodied Conversational Agent”, welcher die Repräsentation einer multimodalen Schnittstelle durch die virtuelle Abbildung eines Menschen beschreibt.

Für den weiteren Verlauf dieser Arbeit werden der Einfachheit halber die Begriffe Virtual Human bzw. virtueller Assistent verwendet, da diese die für die Forschungsarbeit zugrunde liegende Softwarelösung passend umschreiben und potenzielle Verwirrung durch den Gebrauch zu vieler verschiedener Terme vermieden wird.

## **1.2 Problemstellung**

### **1.2.1 Ausgangssituation**

Der Alltag von Studierenden an Hochschulen und Universitäten ist geprägt von vielen Verpflichtungen und Herausforderungen. Aufgrund zahlreicher und zumeist Studiengangsspezifischer Richtlinien und Anforderungen kann es daher schwierig sein, den Überblick zu behalten. Auch für Interessierende ist es nicht immer leicht, einen Überblick über den jeweiligen Studiengang zu bekommen und sich über die Einzelheiten genau zu informieren.

Häufig ist es schwierig, bestehende Fragen beispielsweise in einem persönlichen Gespräch mit Mitarbeitenden der jeweiligen Institution zu klären und daher nicht selten mit einem hohen zeitlichen Aufwand verbunden, um an benötigte Informationen zu gelangen. Darüber hinaus besteht das Problem, dass sich Menschen mit körperlichen Beeinträchtigungen oft zusätzlichen Herausforderungen für den Erhalt solcher Informationen gegenüber sehen.

### **1.2.2 Motivation**

Um Studierende und solche, die sich für ein Studium interessieren, dabei zu unterstützen, liegt der Fokus dieser Arbeit auf der Beantwortung der Frage, wie effektiv virtuelle Assistenten in der Rolle als Berater sein können. Da diese eine Kommunikation mittels multimodaler Interfaces ermöglichen, würde hiermit ein virtueller Ansprechpartner zur Verfügung stehen, der zeitlich unabhängig auf Fragen und Probleme eingehen und somit eine gute Alternative zu Beratungsgesprächen, oder Webrecherchen sein kann, um es Studierenden zu ermöglichen einfacher und zugänglicher an benötigte Informationen zum Studium zu gelangen.

### **1.2.3 Zielsetzung**

Ziel dieser Arbeit ist es zu untersuchen, wie effektiv der Einsatz eines Virtual Humans als virtueller Assistent im Umfeld eines Studiums sein kann und ob dieser einen Mehrwert für Studierende bietet. Zu diesem Zweck wird ein entsprechender Prototyp entwickelt und die bei der Verwendung auftretenden Reaktionen sowie eingehendes Feedback in Bezug auf dessen Effektivität im Studienalltag ausgewertet.

Dieser virtuelle Assistent soll in der Lage sein, auf verbal gestellte Fragen einzugehen und diese im Rahmen des aktuellen Kontexts beantworten zu können. Um anwendenden Personen das Gefühl zu geben, mit einer realen Person zu sprechen, soll zudem eine Erkennung von Gesten implementiert werden, auf die der Assistent reagieren kann.

## 2 Verwandte Arbeiten

Zur Einordnung dieser Arbeit und der im Zuge dieser zu erstellende Prototypisierung werden in diesem Kapitel vergleichbare Arbeiten vorgestellt und abgegrenzt. Virtuelle Assistenten gibt es in Form von Smart Speakern, wie beispielsweise Apples Siri, oder Amazons Alexa, seit vielen Jahren. Solche reinen Sprach - Interfaces, sowie die dazugehörigen Sprachmodelle wurden im Laufe der Jahre weiterentwickelt und um eine visuelle Ebene erweitert, die nicht nur für die Darstellung, sondern auch für die Möglichkeit einer non-verbale Kommunikation dient. Diese virtuellen Assistenten finden bereits Anwendung in verschiedenen Bereichen, wie beispielsweise dem Kunden Support.

Ein bekanntes Beispiel stellt die von IPSoft entwickelte virtuelle Assistentin Amelia dar, welche im August 2016 zum ersten Mal für die SEB, einer schwedischen Bank, zum Einsatz kam. Aufgrund ihrer Fähigkeiten im Bereich der natürlichen Sprachverarbeitung wurde sie zunächst für den internen IT Service eingesetzt und hierfür sowohl als Web als auch als mobile App zur Verfügung gestellt, die neben einer visuellen Darstellung über einen Chat zur Kommunikation verfügt (Abbildung 2.1). Ihre Zuständigkeit lag dabei zunächst auf Anfragen bzgl. Identity Access Management und Knowledge Management, welche zusammen 15 Prozent der eingehenden Anrufe ausmachte (Lacity et al., 2017). Zu diesen Anfragen zählten unter Anderem das Zurücksetzen des Passwortes, welches Amelia dazu veranlasst den Mitarbeitenden selbständig zu befragen, die Antworten mit System Aufzeichnungen gegen zu prüfen und anschließend das Passwort zurückzusetzen, aber auch das Beantworten von Fragen zu Themen wie bestimmte Systeme bedient, oder Formulare beantragt werden können. Sie war dabei in der Lage, 50 % der Anfragen ohne weitere menschliche Hilfe zu lösen und beschleunigte damit die Bearbeitungszeiten um ein Fünffaches. Durch diese Erfolge steht sie seit Dezember 2016 auch für externe Kunden als Web Anwendung zur Verfügung (Lacity et al., 2017).

Ein weiteres Beispiel für den Einsatz von virtuellen Assistenten findet sich in der Arbeit von Hartholt et al., 2019 in der es darum geht, solche für Übungszwecke im Rahmen einer Augmented Reality Anwendung zu integrieren. Der Fokus dieser Arbeit liegt dabei auf der praktischen Übung von Einstellungsgesprächen, bei der ein virtueller Assistent mittels AR in das Sichtfeld der anwendenden Person projiziert wird und hierbei die Rolle des Fragestellers übernimmt (Abbildung 2.2). Das besondere hierbei ist, dass der virtuelle Assistent auf eine Ebene Fläche in einer sitzenden Pose projiziert wird und dabei durch das Ausrichten des Kopfes

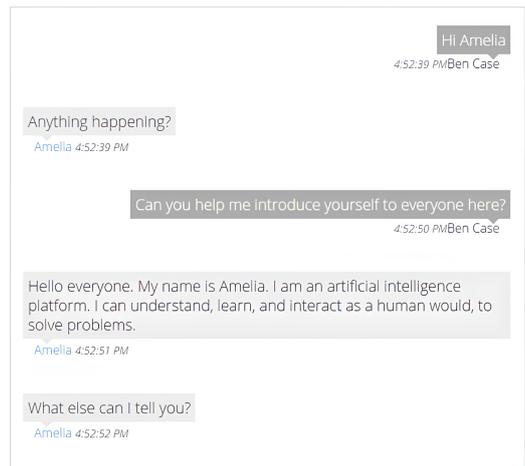


Abbildung 2.1: Ansicht der von IPSoft entwickelten virtuellen Assistentin Amelia (Quelle: Medium, 2018)

nach Blickkontakt mit den anwendenden Personen sucht, um somit einen höheren Grad an Immersion zu erzielen und den Trainingseffekt zu verstärken. Anwendende Personen können verbal mit dem virtuellen Assistenten sprechen, welcher dabei Fragen aus einer vorgefertigten Liste stellt. Für die visuelle Darstellung des Assistenten stehen dabei sechs verschiedene Modelle sowie drei verschiedene Konversationsmodi zur Auswahl. Diese Modi geben die Möglichkeiten, das Verhalten und Auftreten des Fragestellers zwischen freundlich, neutral und unfreundlich wechseln zu können. Das Ziel hierbei ist es, die Anwendung effektiver zu gestalten, indem unterschiedliche Situationen simuliert werden können. Erste Tests zeigten, dass die Probanden\*/innen dieser Technologie zugeneigt waren und die Möglichkeit zum Üben wichtiger Kompetenzen innerhalb einer erweiterten realen Umgebung begrüßt wurde. Durch die positive Resonanz und die daraus abzuleitende Effektivität soll die Anwendung noch weiterentwickelt und um neue Inhalte erweitert werden.

Des Weiteren spielen Virtual Humans auch im Bereich der Unterhaltung und medialen Interesse eine zunehmende Rolle. So genannte Virtual Influencer erfreuen sich in den letzten Jahren zunehmender Beliebtheit (Lim & Lee, 2023). Dabei handelt es sich um Virtual Humans, die als Menschen oder menschenähnliche Avatare in sozialen Medien präsent und hierbei in der Lage sind, als eigenständige Content Creator aufzutreten. Sie sind dabei keine Repräsentation von echten Nutzern, sondern von Experten bzw. Agenturen entworfene Systeme, die verschiedene Technologien wie Web 3.0, NFT, Chatbots, CGI oder Künstliche Intelligenz verwenden und zudem über ihre eigene Persönlichkeit sowie Hintergrundgeschichte verfügen (Koles et al., 2024). Häufig werden diese von größeren Firmen bzw. Marken eingesetzt, um Meinungen oder Verhalten ihres Publikums im Zuge einer interaktiven Marketing Strategie zu beeinflussen, indem sie beispielsweise für entsprechende Produkte werben (Arsenyan & Mirowska, 2021; Audrezet & Koles, 2023). Ähnlich wie bei menschlichen Influencern profitieren diese von den



Abbildung 2.2: Ansicht der AR Lernsimulation zum Üben von Einstellungsgesprächen  
(Quelle: Hartholt et al., 2019)

Vorteilen sozialer Medien im Hinblick auf die Möglichkeit zur Interaktion zwischen Content Creator und den Zuschauenden (Arsenyan & Mirowska, 2021). Zu den bekanntesten Virtual Influencern zählen Lil Miquela, eine auf Instagram aktive Sängerin und Shudu, ein ebenfalls auf Instagram vertretendes Model.

Auch wenn diese virtuellen Influencer einem anderen Verwendungszweck dienen, so verwenden sie dennoch die gleiche Grundfunktionalität bzw. Technologien zur Interaktion mit Menschen, wie die virtuellen Assistenten, die im Fokus dieser Arbeit stehen.

Virtual Humans kommen neben der Tätigkeit als Influencer, oder Assistenten auch in anderen Gebieten bereits zum Einsatz. So werden sie beispielsweise im Rahmen von Lernsimulationen eingesetzt, um den Anwendenden in verschiedenen Situationen für die jeweilige Interaktion zur Verfügung zu stehen. Gerade in Bereichen wie Medizin, Notfallübungen oder Psychotherapien können glaubwürdige Virtual Humans einen Mehrwert für das jeweilige Training ausmachen (Nadia Magnenat-Thalmann, 2004). Vereinigungen wie das Institute for Creative Technologies arbeiten dabei an der Umsetzung von Virtual Humans, die für diesen Zweck als virtuelle Trainer ausgelegt sind und dabei verschiedene Rollen, so zum Beispiel als Guides, Mentoren, Kontrahenten oder Teammitglieder übernehmen können und neben physischen auch zwischenmenschliche Fertigkeiten trainieren (Kenny et al., 2007). Auch hierbei spielt die Fähigkeit von Virtual Humans zur Interaktion mit den Anwendenden eine große Rolle, da

Virtual Humans im Gegensatz zu bisheriger Lernsoftware ihr Verhalten situationsabhängig anpassen und somit den Trainingseffekt verstärken können (Kenny et al., 2007).

Im Bereich des Studiums und somit der Unterstützung von Studierenden durch virtuelle Assistenten gibt es bereits wissenschaftliche Ausarbeitungen, die sich ebenfalls mit der Rolle von virtuellen Assistenten an Hochschulen und Universitäten beschäftigen (Gubareva & Lopes, 2020). Im Gegensatz zu diesen Ausarbeitungen grenzt sich die hier vorgenommene Untersuchung dahingehend ab, dass neben einem grundlegenden Verständnis dieser Technologie untersucht wird, wie effektiv diese für die Nutzung durch Studierende ist und wie eine entsprechende Softwarelösung als Assistent aussehen könnte.

Obwohl die zuvor aufgezeigten Beispiele verdeutlichen, dass es bereits eine Vielzahl von Lösungen in verschiedenen Bereichen, sowie Firmen gibt, die auf die Konzeption und Realisierung solcher Virtual Humans spezialisiert sind, existiert noch keine konkrete Lösung für den Einsatz im Rahmen eines Studiums. Da der für die Prototypisierung angestrebte Verwendungszweck voraussetzt, dass spezifische Fragen zum entsprechenden Studiengang gestellt werden können, ist es somit erforderlich, ein System bzw. ein Sprachmodell einzusetzen, das über die dafür notwendige Wissensbasis verfügt. Dadurch ist die Verwendung solcher bereits existierenden Lösungen für einen derart spezifischen Fall nicht ohne Weiteres möglich. Um der Frage nach der Effektivität virtueller Assistenten im Studium auf den Grund zu gehen, ist es somit erforderlich, eine für diesen Zweck ausgerichtete Softwarelösung zu entwickeln.

# 3 Projektdurchführung

Im folgenden Abschnitt wird die zur Beantwortung der Forschungsfrage durchgeführte Prototypisierung, sowie die dabei zugrunde liegende Methodik, näher vorgestellt. Der erstellte Prototyp dient als Ausgangspunkt für die weitere Untersuchung in Bezug auf Zugänglichkeit, intuitive Bedienung, Akzeptanz und Wirksamkeit im Vergleich mit anderen Medien, um auf Basis des Feedbacks ein konkretes Bild über die Effektivität eines Digital Humans im Rahmen des Studiums abbilden zu können.

## 3.1 Rahmenbedingungen

Im Folgenden wird die für dieses Projekt verwendete Hard- und Software aufgelistet, um hierbei eine Übersicht über die eingesetzten Softwarepakete und die benötigte Peripherie zu geben.

### 3.1.1 Software

Die Übersicht in [3.1](#) gibt die für die Entwicklung verwendete Software mit deren Bezeichnung, sowie eine kurze Beschreibung in Hinblick auf die Verwendung in diesem Projekt an. Die Tabelle stellt dabei auch die Voraussetzung für weitere Entwicklungsarbeiten, sowie die grundsätzliche Ausführung des Prototypen dar. Die detaillierte Beschreibung der verwendeten Software wird im Rahmen der Machbarkeitsstudie in [Abschnitt 3.2.3](#) aufgeführt.

Tabelle 3.1: Software

Software	Beschreibung
Windows 10 / 11	benötigtes Betriebssystem
Unity	3D Engine
Ready Player Me	Unity Plugin zur Erstellung / Importierung von 3D Charakter Modellen
Adobe Mixamo	Webapplikation zur Bereitstellung von Animationen
Oculus Lipsync	Unity Package zur automatisierten Animation der Lippenbewegungen eines 3D Charakter Modells
Flask	Python Framework zur Erstellung des Webservices
OpenAi Whisper	Sprachmodell zur Transkription von gesprochenen Wörtern zu Text
OpenAi ChatGPT	Sprachmodell zur Verarbeitung und Reaktion auf Textbasierte Nutzer Eingaben
Gtts	Python Package zur Umwandlung von Text zu gesprochenen Worten
Langchain	Python Framework zur vereinfachten Handhabung von Sprachmodellen

### 3.1.2 Hardware

Die Tabelle unter [3.2](#) dient zur Übersicht der erforderlichen Hardware, welche zur Inbetriebnahme der Anwendung benötigt wird. Es werden hierbei sowohl die Komponenten benannt, als auch deren Verwendungszweck kurz beschrieben. Die hierbei aufgelisteten Komponenten sind für den später angedachten Aufbau zur Durchführung der Nutzeranalyse und somit die Evaluation notwendig. Neben der grundlegenden Peripherie von Computer, Monitor, Maus und Tastatur zur Inbetriebnahme des virtuellen Assistenten werden Kamera und Mikrofon, sowie Lautsprecher für die Verwendung der angedachten multimodalen Schnittstellen benötigt. Da Datenmodelle über eine Web API verwendet werden sollen, wird zudem ein Internetzugang benötigt.

Tabelle 3.2: Hardware

Hardware	Beschreibung
PC	Computer zur Ausführung der Anwendung
Monitor	Full HD Monitor zur Darstellung der Visualisierung
Kamera	Webcam zur Erfassung der Anwender*innen
Mikrofon	Aufnahme von Sprachbefehlen
Lautsprecher	Wiedergabe der Antwort der virtuellen Assistentin
HDMI Kabel	Verbindung von PC und Monitor
Netzkabel	Stromversorgung von PC und Monitor
Internetzugang	Verbindung zur OpenAi API
Maus und Tastatur	Zum initialen Starten der Anwendung

## 3.2 Prototyp

In diesem Kapitel werden die Konzeption und Entwicklung des für die Beantwortung der Forschungsfrage relevanten Prototyps, sowie die damit verbundenen Designentscheidungen herausgestellt.

### 3.2.1 Anforderungen

Für die Erstellung des Prototypen ist es zunächst erforderlich, die genauen Anforderungen hinsichtlich des angestrebten Verwendungszwecks aufzustellen. Das Ziel der weiteren Untersuchung sieht hierbei vor, dass sich die Anwendung an verschiedenen Kriterien in Bezug auf Zugänglichkeit, intuitive Bedienung, Akzeptanz und Wirksamkeit im Vergleich mit anderen Medien misst, um die Effektivität besser einordnen und beurteilen zu können. Wie in den vorherigen Kapiteln bereits ausgeführt, gibt es bereits vergleichbare virtuelle Assistenten in anderen Fachbereichen, oder Ausrichtungen, an denen sich auch dieser Prototyp orientiert. Nach Kenny et al., 2007 stehen für die Effektivität eines Virtual Humans allen voran die Aspekte Glaubwürdigkeit, Reaktionsfähigkeit und Interpretierbarkeit im Fokus.

Angefangen bei einer der wohl wichtigsten Anforderungen für einen solchen virtuellen Assistenten, soll der Prototyp über ein multimodales Interface verfügen und somit in der Lage sein, sowohl auf verbale, als auch nonverbale Aktionen der anwendenden Person reagieren zu können. Mit Bezug auf die Reaktionsfähigkeit soll es hierbei für anwendende Personen möglich sein, direkt mit dem Prototypen zu interagieren, ohne dass bestimmte Aktionen vorangestellt werden müssen. Der Prototyp soll eigenständig in der Lage sein, die Interaktion zu erkennen und darauf entsprechend zu reagieren.

Der virtuelle Assistent soll über ein 3D Modell dargestellt werden, welches über verschiedene Animationen verfügt, um das Verhalten bzw. das Erscheinungsbild eines realen Menschen zu simulieren. Damit dieser eine möglichst hohe Glaubwürdigkeit erzielt und somit eine ansprechende Wirkung auf die anwendende Person hat, sollen sowohl Modell, als auch Animationen nach Möglichkeit menschenähnlich sein, da eine anthropomorphe Gestalt ähnlich wie eine gänzlich fehlende zu einer geringeren Akzeptanz führen (Luger & Sellen, 2016). Die zu implementierenden Animationen sollen dabei eine Idle-Animation, sowie eine Eye-Blink-Animation enthalten, damit der Eindruck entsteht, dass es sich um ein lebendiges Wesen handelt. Des Weiteren sollen einfache Gesten angefangen bei einem Winken erkannt und mit einer entsprechenden Animation darauf reagiert werden.

Damit gestellte Fragen hinsichtlich ihrer Korrektheit geprüft und vom Assistenten gegebene Antworten wiederholt nachvollzogen werden können, soll ein Chatfenster den Gesprächsverlauf wiedergeben. Gerade in Hinblick auf die Interpretierbarkeit sollen sowohl die getroffenen Aussagen als auch das Verhalten des virtuellen Assistenten für die anwendenden Personen nachvollziehbar sein. Als weitere Anforderung soll ein Indikator die Aufnahmefähigkeit des virtuellen Assistenten implizieren, damit Anwendende wissen, wann sie ihre Frage stellen können und somit für eine intuitivere Bedienung der Anwendung gesorgt wird.

### **3.2.2 Lösungskonzept / Design**

Für die Umsetzung des Prototypen unter Berücksichtigung der aufgestellten Anforderungen soll zunächst ein entsprechendes Konzept erstellt werden. Da ein virtueller Assistent aus mehreren verschiedenen Funktionalitäten besteht, werden diese als einzelne Module betrachtet, welche im Zusammenhang das Gesamtsystem ergeben. Betrachtet man die zuvor aufgestellten Anforderungen, so ergeben sich folgende Module: Spracherkennung, Textverarbeitung mittels GPT und Text to Speech. Jedes Modul soll dabei in einer eigenen Klasse abgebildet werden, die ihre jeweilige Funktionalität nacheinander ausführen und somit eine Prozesskette bilden. Da dieses Projekt verschiedene Sprachmodelle zur Verarbeitung von Eingaben verwenden soll und diese zumeist in Python geschrieben sind, ist die Bereitstellung eines zusätzlichen Services zur Gewährleistung der Modul Funktionalitäten erforderlich.

Das Modul der Spracherkennung stellt den Beginn der Prozesskette dar. Es dient dazu, auditive Benutzereingaben wie Fragen oder Aufforderungen zu erfassen und in entsprechenden Text umzuwandeln, welcher an das nachfolgende Modul der Prozesskette weitergereicht werden soll. Das Modul selbst soll dabei aus zwei Teilen bestehen. Im ersten Teil soll eine Verbindung mit dem benötigten Mikrofon aufgebaut und eingehende Sprache im Dateiformat MP3 zwischengespeichert werden, welche im zweiten Teil vom Sprachmodell verarbeitet werden soll. Zu diesem Zweck soll die erzeugte MP3 Datei mittels Webrequest an den Service übergeben werden, welcher diese wiederum durch ein Sprachmodell in Text umwandelt und

an das Modul zurückgibt. Der so erhaltene Text soll zur Kontrolle im Chatfenster des UIs ausgegeben werden. Im nächsten Schritt der Prozesskette soll dieser Text weiterverarbeitet werden. Hierzu wird der Text an das zweite Modul übergeben, welches diesen erneut an den Webservice schickt. Mit Hilfe eines auf maschinellem Lernen basierten GPT Algorithmus soll der Text verarbeitet und eine entsprechende Antwort, ebenfalls in textueller Form, erstellt werden. Diese Antwort soll an das Modul zurückgegeben und anschließend dem Chatfenster hinzugefügt werden. Im dritten und somit letzten Schritt der Prozesskette, dem Text to Speech Modul, soll der zuvor erhaltene Antwort Text ein letztes Mal via Webrequest an den Service übergeben und dort zu einem MP3 umgewandelt werden. Im Anschluss soll diese MP3-Datei vom Service zurück an das Modul übergeben und von diesem abgespielt werden.

Für alle Module ist es somit essentiell, dass ein entsprechender Webservice zur Verfügung steht, der die spezifischen Anfragen verarbeiten kann. Dieser Service soll parallel zur Anwendung im Hintergrund, oder alternativ auf einem externen System wie einem Server laufen und für jede benötigte Funktion ein entsprechendes Routing aufweisen. Jede dieser Routen soll somit von dem dazugehörigen Modul des virtuellen Assistenten aufgerufen werden können.

Hinsichtlich des Designs kommt neben der visuellen Darstellung noch ein Modul für die Gestenerkennung hinzu. Dieses Modul soll als Schnittstelle für einen weiteren Background Service dienen, der über eine Verbindung zu einer angeschlossenen Kamera verfügt und permanent die Bewegungen von Personen im Sichtfeld erfasst. Sobald ein Bewegungsmuster als Geste erkannt wird, soll diese an das Modul gemeldet werden, welches daraufhin eine passende Animation beim virtuellen Assistenten als Reaktion einleitet.

### 3.2.3 Proof of Concept

Hinsichtlich der Machbarkeit des Projektes ergaben die ersten Recherchen, dass es für die Umsetzung der einzelnen Module, sowie Services bereits entsprechende Technologien in Form von Schnittstellen, Frameworks und Packages gibt, die hier in Kombination verwendet werden können. Der folgende Abschnitt befasst sich somit mit der Durchführbarkeit unter Berücksichtigung des zuvor aufgestellten Konzeptes und zeigt, welche dieser Technologien für die Implementierung des Prototyps zum Einsatz kommen sollen.

Angefangen beim zugrunde liegenden System stellt die 3D Engine *Unity* die benötigten Tools und Möglichkeiten zur Realisierung des Prototypen dar, da es sich hierbei um eine Anwendung mit audiovisuellen Komponenten handelt, ist die darauf ausgelegte Engine für diesen Zweck sehr gut geeignet. Als zusätzliche Packages für Unity stehen mit *Ready Player Me*, einem Package für die Erstellung und dem Importieren von 3D Charakter Modellen und *Oculus LipSync*, einem Plugin zur Synchronisation von Lippenbewegungen mit den dazugehörigen Sprachaufnahmen, zwei zusätzliche Tools für die Erstellung der visuellen Repräsentation des virtuellen Assistenten zur Verfügung. Mit Hilfe der von Adobe zur Verfügung gestellten

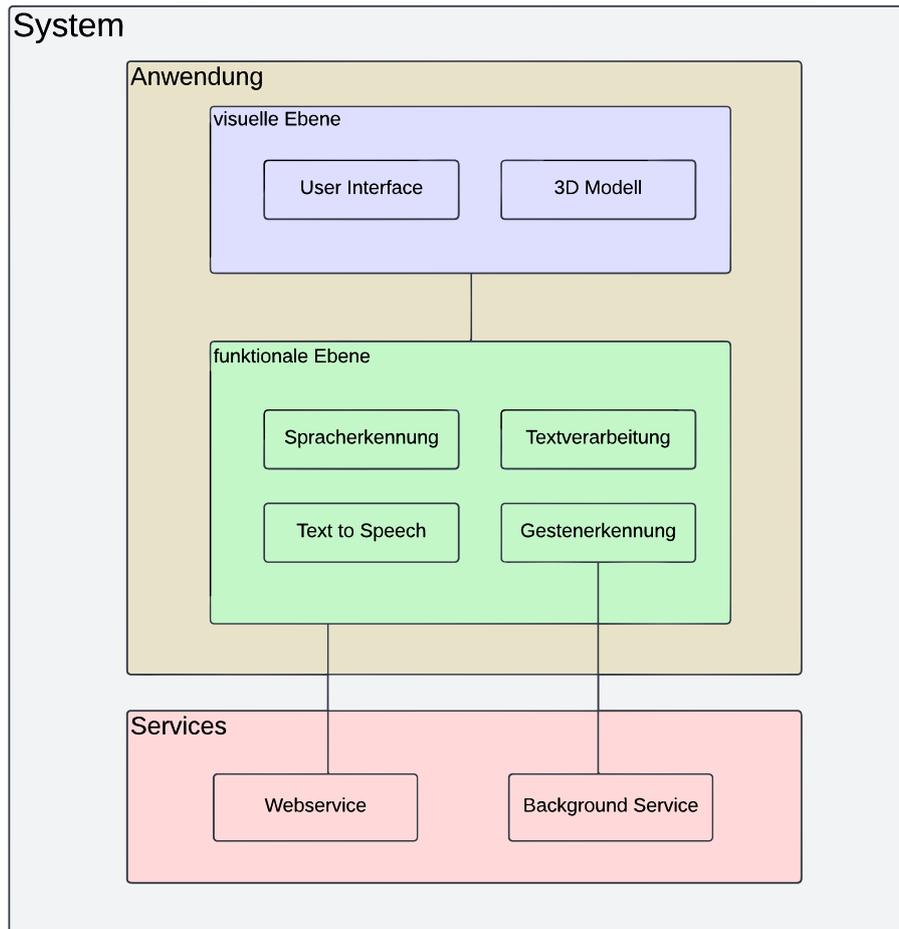


Abbildung 3.1: Grafische Übersicht zur Darstellung des grundlegenden Systemaufbaus  
(Quelle: Eigene Abbildung)

Webapplikation *Mixamo* lassen sich zudem die benötigten Animationen, neben denen von Ready Player Me integrierten Animationen für die Bewegungen von Augen und Lippen, in Unity importieren. Auch bietet Unity integrierte UI Elemente zum Aufbau des angedachten Chatfensters, sowie für den Indikator, somit werden alle visuellen Komponenten von diesen Technologien abgedeckt.

Zur Umsetzung der einzelnen Module eignen sich verschiedene Python Packages, welche mit Hilfe des Frameworks *Flask* zu dem im Konzept vorgestellten Web Service verbunden werden können. Dazu zählt das von Open AI bereitgestellte Package zur Implementierung von *Whisper AI* und *Chat GPT*. Bei *Whisper AI* handelt es sich um ein Sprachmodell für die Erkennung von Sprachmustern in Audiodateien und die Umwandlung in zugehörige Texte. Zudem lässt es sich lokal einrichten und somit ohne zusätzliche Latenzen verwenden, was es im Hinblick auf Leistung und Performance ideal für den Einsatz zur Spracherkennung macht.

Für die Implementierung der Funktionalität zum Verarbeiten und Antworten von gestellten Fragen eignet sich die Kombination aus der im Package mitgelieferten Schnittstelle zu Chat GPT und dem Framework *Langchain*, welches dazu dient eine spezifische Wissensbasis mit der Anfrage an Chat GPT zu verknüpfen, damit der Algorithmus in der Lage ist auf Fragen zum Studium zu antworten für die dieser die nötigen Informationen normalerweise nicht vorliegen hat. Für das verbleibende Modul zur Wiedergabe der erhaltenen Antwort steht mit *Google Text to Speech* eine Python Bibliothek zur Verfügung, die wie die zuvor genannten Packages einfach über Pythons Package Manager Pip installiert werden kann und die die benötigte Funktion für die Umwandlung der Antwort zu einer Audiodatei ermöglicht.

Auf Basis der somit aufgestellten Vorlage ist eine Implementierung des Prototyps möglich und wird im nächsten Schritt dieser Arbeit durchgeführt.

### 3.2.4 Stand des Prototypen

Der Prototyp zur Evaluation der Effektivität von Virtual Humans als Assistenten im Studiengang konnte gemäß dem aufgestellten Konzept umgesetzt werden. Die so entstandene virtuelle Assistentin ist somit in der Lage, auf verbale Fragen bzgl. des Studienganges Media Systems zu antworten und gezielt Informationen aus dem Modulhandbuch und der Website wiedergeben zu können. Durch die Implementierung von Open AI's Api zur Einbindung von Chat GPT ist hierbei auch eine Konversation über andere Themen und Problemstellungen auf Basis der von Open AI verwendeten Wissensbasis möglich, was allerdings zu Latenzen von bis zu Zwanzig Sekunden führt, da die Anfrage zunächst an den entsprechenden Server geschickt, verarbeitet und zurückgesendet werden muss. Die für den Prototyp verwendete Spracherkennung funktioniert ähnlich wie bei anderen Sprachassistenten, indem der Name Ava (Ableitung von "Artificial Virtual Assistant") als Schlüsselwort genannt wird, um nachfolgend eine Frage zu stellen. Anwendende haben hierbei ein Zeitfenster von Acht Sekunden, um ihre Frage oder Aufforderung zu formulieren.

Aufgrund der zeitlichen Vorgabe erfolgt die visuelle Repräsentation der Assistentin über ein 3D Modell in Cartoon ähnlicher statt der zuvor angedachten fotorealistischen Optik. Trotzdem verfügt das Modell über Animationen, die es menschenähnlich erscheinen lassen sollen. Dazu gehört neben einer Idle-Animation, welche das Modell kontinuierliche Bewegungen wie Atmen, oder das Verändern der Körperhaltung ausführen lässt, auch das Bewegen bzw. Blinzeln der Augen. Zudem werden die Lippen passend zu den gesprochenen Worten im Falle einer Konversation animiert. Zusätzlich ist die Anwendung in der Lage, Gesten zu erkennen. Im Augenblick kann hierbei erkannt werden, ob anwendende Personen im Sichtfeld der Assistentin dieser zu winken, was dazu führt, dass diese ebenfalls eine Wink-Geste als Reaktion darauf ausführt. Dadurch ist der Prototyp in der Lage, menschliches Auftreten und Verhalten weitestgehend zu simulieren.

Bezüglich der UI wurden zwei Elemente für eine verständlichere Bedienung bzw. Interaktion mit der Anwendung implementiert. Zum einen umfassen diese das im Konzept angedachte Chatfenster zum Mitlesen der Konversation. Dies dient vorrangig der Kontrolle der von der Spracherkennung erfassten Inhalte, aber auch der Nachverfolgung der gesamten Konversation, was es somit auch Leuten mit eingeschränktem Hörvermögen ermöglicht, die Dienste der virtuellen Assistentin in Anspruch zu nehmen. Zum Anderen erscheint oberhalb des Chatfensters ein Icon, sobald die zuvor angesprochene Spracherkennung mittels Schlüsselwortes gestartet wurde. Dies dient dazu, den Zeitraum zu visualisieren, in dem Fragen gestellt werden können und ist daher zu diesem Zweck mit einer kreisförmigen Umrandung ausgestattet, die mit jeder vergangenen Sekunde abnimmt (Abbildung ). Ein zusätzlicher Hilfstext oberhalb des Chatfenster soll Anwendenden zudem die Funktionsweise des Schlüsselwortes und somit die generelle Verwendung des Prototypen vermitteln.

Im aktuellen Stand ist der Prototyp für die Verwendung auf einem Computer mit Windows 10 / 11 Betriebssystem ausgelegt und lässt sich aufgrund von Abhängigkeiten zu Microsoft Paketen nicht ohne entsprechende Anpassungen auf anderen Systemen einsetzen.



Abbildung 3.2: Screenshot des umgesetzten Prototyps (Quelle: Eigene Abbildung)



Abbildung 3.3: Screenshot des eingesetzten Timers für das Stellen einer Frage (Quelle: Eigene Abbildung)

### 3.2.5 Weiterentwicklungsmöglichkeiten / Zukunftsaussicht

Auch wenn der Prototyp alle erforderlichen Anforderungen für die weiteren Untersuchungen erfüllt, so gibt es dennoch Möglichkeiten zur Weiterentwicklung. Allen voran bestünde die Möglichkeit, den Prototypen als Webanwendung oder auch nativ für andere Plattformen wie beispielsweise mobilen Endgeräten zur Verfügung zu stellen. Auf diese Weise würde sich die Verfügbarkeit deutlich verbessern, da die Hilfe der virtuellen Assistentin somit auch unterwegs in Anspruch genommen werden könnte.

Ein weiterer Punkt für eine mögliche Überarbeitung bietet die visuelle Präsentation durch 3D Modell und Animationen. Eine Verbesserung in Hinsicht auf eine realistischere Darstellung könnte hierbei durch den Austausch des 3D Modells, sowie die Hinzunahme weiterer Animationen erreicht werden. Durch verschiedene Idle, oder Gesprächs Animationen, die in zufälliger Reihenfolge abgespielt werden, könnte somit der Eindruck eines menschlicheren Verhaltens der virtuellen Assistentin erzielt werden.

Um den Anschein einer realen Gesprächssituation zu verstärken, könnte darüber hinaus eine zusätzliche Funktion implementiert werden, mit der die virtuelle Assistentin in der Lage wäre, den Augenkontakt zum Gesprächspartner / zur Gesprächspartnerin zu suchen und zu halten. Hinsichtlich der Performance wäre eine weitere Möglichkeit zur Verbesserung der Anwendung die Schnittstelle zu Open AI's Chat GPT durch ein für den Verwendungszweck als Studienberaterin spezifiziertes Datenmodell im Webservice zu ersetzen, um die Abhängigkeit von Open AI Servern aufzulösen und somit die Latenz zu verbessern. Mit einer Erweiterung der zur Verfügung stehenden Wissensbasis, sowie Anpassungen an den Sprach Settings der

einzelnen Modelle wäre es zudem möglich, weitere Studiengänge in zusätzlich verschiedenen Sprachen anzubieten. Modelle wie das verwendete Whisper AI verfügen hierbei bereits über eine Funktion zur Erkennung der Sprache, was es der virtuellen Assistentin somit ermöglichen würde, in derselben Sprache zu antworten, in der die zuvor gestellte Frage formuliert wurde, ohne dass diese manuell ausgewählt wurde, was somit auch Austauschstudierenden und Anwendenden anderer Nationalitäten zu Gute kommen würde. Auch die Gestenerkennung bietet Spielraum für weitere Verbesserungen, so könnte das Datenmodell zur Erkennung von verschiedenen Gesten dahingehend erweitert, oder alternativ ausgetauscht werden, um zusätzlich Gebärdensprache erkennen und auswerten zu können. Somit könnte die Inklusion und damit auch die Zugänglichkeit zur Anwendung verbessert werden. Mittels Data Mining stünde zusätzlich noch eine Option zur Verfügung, um die Daten innerhalb der Wissensbasis autonom auf dem aktuellen Stand zu halten.

In Anbetracht der vielen Möglichkeiten zur Erweiterung bzw. Verbesserung bietet der Prototyp eine Grundlage für weitere Projektarbeiten und damit verbundene Untersuchungen.

### **3.3 Methodik**

Für die Untersuchung der Forschungsfrage nach der Effektivität von KI im Studiengang ist die Wahl der Methodik auf eine Nutzeranalyse mittels System Usability Scale (SUS) gefallen bei der die verschiedenen Aspekte der Effektivität mittels spezifischer Fragen an Probanden\*/innen bewertet und eingeordnet werden sollen, nachdem diese den Prototypen begutachtet und getestet haben.

Zur Durchführung dieser Studie wird zunächst ein virtueller Assistent als Prototyp entwickelt und darauf ausgelegt, grundlegende Fragen zum Studiengang beantworten zu können, um somit eine Basis für die spätere Analyse seitens der Probanden\*/innen zu schaffen. Im Anschluss wird ein Fragebogen erstellt, der aus zehn Fragen zu verschiedenen Facetten des Prototyps besteht, damit eine möglichst große Bandbreite an Faktoren, die über dessen Effektivität entscheiden, abgedeckt wird. Zu diesen Faktoren zählen die Bedienbarkeit, die Zuverlässigkeit, die Akzeptanz in Bezug auf visuelle Repräsentation und auditive Wiedergabe der Anwendung. Nach dem SUS werden die dazugehörigen Fragen in Aussagen verpackt, die abwechselnd positiv und negativ konnotiert sind und jeweils mit Punkten zwischen Eins und Fünf bewertet werden können. Dabei stehen fünf Punkte für "trifft komplett zu" und ein Punkt für "trifft überhaupt nicht zu", während die Punkte dazwischen als Abstufungen dienen. Die Verteilung der Punkte erfolgt über das Ankreuzen der zu einer Aussage gehörenden Felder, bei denen pro Aussage nur ein Kreuz gesetzt werden darf. Der so für diese Studie erstellte Fragebogen kann im Anhang dieser Arbeit eingesehen werden (Siehe Abschnitt 5).

Da sich der Verwendungszweck um den Einsatz im Rahmen des Studiums dreht, werden für diese Studie vorrangig Studierende der Fakultät Design, Medien und Information an der HAW Hamburg als Probanden\*innen eingeladen. Die Einladung erfolgt eine Woche vor Durchführung der Nutzeranalyse über eine Rundmail an alle Studierenden. Hierfür werden Zeiträume von je drei Stunden an zwei verschiedenen Tagen eingerichtet, an denen jeweils ein Raum mit zwei Teststationen zur Verfügung steht, an denen der Prototyp zuvor eingerichtet wurde. Um eine gegenseitige Beeinträchtigung der Probanden\*/innen bei Verwendung des Sprach Interfaces zu vermeiden, werden die Teststationen mit Headsets ausgestattet, die jeweils über ein eigenes Mikrofon verfügen.

Ziel dieser Studie ist es zu untersuchen, welche Aspekte sich wie auf die Effektivität auswirken und ob Studierende überhaupt das Gefühl eines Mehrwerts bei der Verwendung eines virtuellen Assistenten haben, oder herkömmliche Methoden zum Erhalt von Informationen, wie beispielsweise das persönliche Gespräch mit Dozenten oder Mitarbeitern der jeweiligen Fakultät, oder alternativ die Recherche über Webseiten vorziehen. Aufgrund der Tatsache, dass das SUS Verfahren auf die Einschätzung der vorgegebenen Aussagen beschränkt ist und sich somit nicht unbedingt die genauen Hintergründe zu getroffenen Wertungen ableiten lassen, verfügt der Fragebogen über ein optionales Feld für Anmerkungen um zusätzliches Feedback und Eindrücke in die Bewertung der Ergebnisse einfließen zu lassen und Kritikpunkte zu präzisieren.

### **3.4 Statistiken**

Bezüglich der Relevanz belegen Statistiken, wie die in Abbildung 3.4 dargestellt einen zunehmenden Anstieg im Erwerb und der Verwendung von virtuellen Assistenten in verschiedenen Bereichen. Durch diese Entwicklung bekommen virtuelle Assistenten eine immer größere Bedeutung im Alltag, was sich auch im Hinblick auf Diversität der Verwendungszwecke, Umfang und Reichweite bemerkbar macht (Vergleich Abschnitt 2).

Rückblickend betrachtet wurden einige Entscheidungen, wie die Darstellung des Prototypen, auf Basis von entsprechenden Statistiken getroffen. So wurden Stimme, als auch das Erscheinungsbild des 3D Modells mit Bezug auf die im Jahr 2022 von Statista veröffentlichte Umfrage zur Akzeptanz von verschiedenen Geschlechtern im Feld der sprachgesteuerten Assistenten gewählt (Siehe 3.5). Wie dieser Statistik zu entnehmen ist, bevorzugen 60% der Befragten Personen eine weibliche gegenüber einer männlichen oder neutralen Stimme, während sich 19% hierbei gänzlich enthalten. Auf Basis dieser Umfrage und somit zur potentiellen Steigerung der Akzeptanz seitens der Anwendenden verwendet der Prototyp sowohl eine Stimme als auch eine Erscheinung mit femininer Ausrichtung. Gerade die Stimme sowie die Aussprache spielen für die regelmäßige Nutzung eine wichtige Rolle. Wie die Abbildung

### Anzahl der Nutzer virtueller digitaler Assistenten weltweit in den Jahren von 2015 bis 2021 (in Millionen)

Prognose zur Nutzung von virtuellen digitalen Assistenten weltweit bis 2021

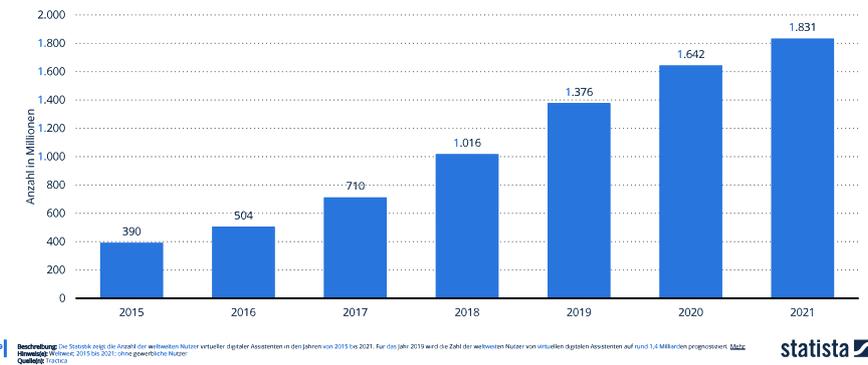


Abbildung 3.4: Statistik zur prognostizierten Nutzung von virtuellen Assistenten (Quelle: Statista.com, 2023)

3.6 belegt, sollten diese so aufgestellt sein, dass sie nicht als befremdlich oder unangenehm aufgefasst werden. Die im Rahmen dieser Arbeit durchgeführte Nutzeranalyse fokussiert sich somit auch auf die Frage, wann die stimmliche Wiedergabe eines virtuellen Assistenten als befremdlich oder unangenehm wahrgenommen wird.

### Würden Sie lieber mit einem digitalen Sprachassistenten mit einer weiblichen oder männlichen Stimme kommunizieren?

Bevorzugung von weiblichen oder männlichen Stimmen bei Sprachassistenten 2022

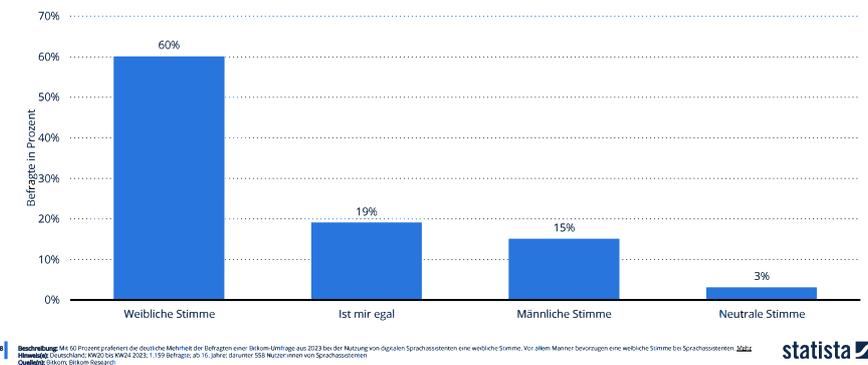


Abbildung 3.5: Umfrage zur Präferenz der geschlechtlichen Stimmausrichtung von virtuellen Assistenten (Quelle: Statista.com, 2023)

Weitere für die Untersuchung relevante Informationen bilden Umfragedaten zu potenziellen Gründen hinsichtlich der Ablehnung von Sprachassistenten bei denen es vor allem um die Ängste geht, dass Daten abgefangen und Anwendende abgehört werden könnten (Vergleich

## Welche Aussagen treffen auf Ihre Erwartungen bezüglich Sprachassistenten zu?

Umfrage: Erwartungen an Sprachassistenten 2022

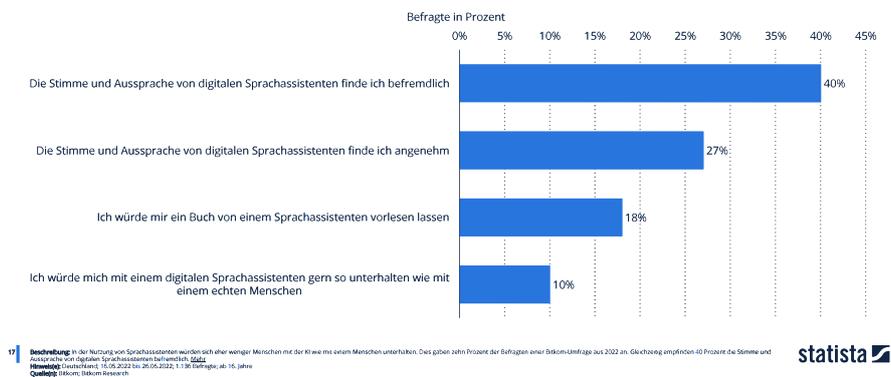


Abbildung 3.6: Umfrage zur Erwartungshaltung bzgl. Sprachassistenten (Quelle: Statista.com, 2023)

Abbildung 3.7), was gleichermaßen auch einen Einfluss auf die Effektivität von virtuellen Assistenten haben könnte. Zu diesem Zweck werden die zu verarbeitenden Sprachaufnahmen erst durch ein Schlüsselwort gestartet und von dem Prototypen auch nicht gespeichert. Einen ebenfalls problematischen Umstand, welcher in Verbindung zu den bereits vorgestellten Statistiken und den daraus abgeleiteten Motiven, stellt das Ergebnis einer Umfrage aus dem Jahr 2022 dar (Vergleich 3.8), derer zufolge lediglich 47% der Nutzer und Nutzerinnen eines Sprachassistenten diesen zum Abfragen von Informationen verwenden. Da dies der Hauptverwendungszweck des Prototypen darstellt, ist es somit erforderlich, diesen Umstand in der Nutzeranalyse näher zu untersuchen und für die Evaluation der Frage nach der Effektivität im Nutzungsrahmen von Studierenden aufzugreifen.

## Umfrage zu den Gründen der Ablehnung von Sprachassistenten in Deutschland im Jahr 2022

Umfrage zur Ablehnung von Sprachassistenten 2022

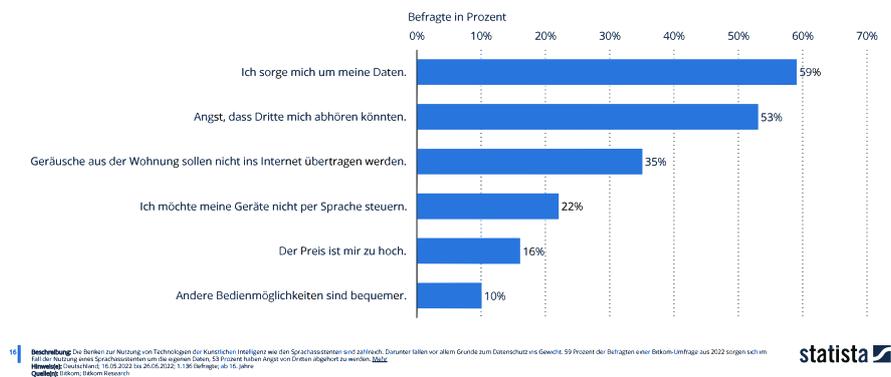


Abbildung 3.7: Umfrage zu potenziellen Gründen für die Ablehnung von Sprachassistenten in Deutschland (Quelle: Statista.com, 2023)

## Nutzen Sie die Möglichkeit, per Sprache Informationen abzufragen und Geräte zu steuern?

Umfrage: Nutzung von Sprachassistenten im Jahr 2022

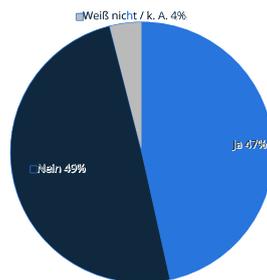


Abbildung 3.8: Umfrage zur Nutzung von virtuellen Assistenten zur Informationsbeschaffung (Quelle: Statista.com, 2023)

## 3.5 Probleme

Bei Betrachtung aktueller Entwicklungen und Erkenntnisse in den für diese Arbeit zugrunde liegenden Technologiefeldern stellt sich heraus, dass es derweilen noch verschiedene Probleme in Form und Funktion gibt, die sich auch auf die Bewertung des Prototypen auswirken und in diesem Kapitel näher beleuchtet und thematisiert werden.

### 3.5.1 Halluzinationen

Natural Language Processing (NLP) und Large Language Models (LLM) aus dem Teilgebiet der Künstlichen Intelligenz, wie das für die Umsetzung des Prototypen verwendete Chat GPT, weisen trotz der stetigen Fortschritte in den letzten Jahren noch immer Probleme in der Nutzung auf. Neben der Performance im Hinblick auf die Zeit, die zwischen der Eingabe des Prompts und dem Erhalt der dazugehörigen Reaktion vergeht, steht dabei vor allem die inhaltliche Korrektheit im Fokus. Hierbei kann es vorkommen, dass vom LLM kommende Antworten im gegebenen Kontext keinen Sinn ergeben, oder sich selbst widersprechen. Dieses Verhalten wird als Halluzination bezeichnet, ein Begriff, welcher schon länger im Bereich des NLP verwendet wird und vorrangig in drei verschiedene Kategorien von fehlerhaften Reaktionen unterteilt werden kann (Zhang et al., 2023).

Die erste Kategorie von Halluzinationen bilden die *Input-Conflicting Hallucinations*, bei denen das LLM Informationen zurückgibt, die von der eigentlichen Anfrage abweichen, oder mit dieser im Widerspruch stehen. Ein solcher Fehler basiert auf einer Fehlinterpretation der zu verarbeitenden Anfrage und der darauf beruhenden Intention des Anwendenden. Ein einfaches Beispiel hierfür könnte sein, dass nach einer Empfehlung für ein Abendessen gefragt wird, während die Antwort der LLM dabei auf ein Mittagessen eingeht. Der ursprüngliche Gedanke von der Frage nach einer Essensempfehlung bleibt zwar erhalten, jedoch weicht die Antwort so weit davon ab, dass das eigentliche Ziel der Frage nicht erreicht wird.

Bei der zweiten Kategorie, den *Context-Conflicting Hallucinations*, werden Informationen wiedergegeben, die gar nicht angefragt wurden und somit keinen Sinn im aktuellen Kontext ergeben. Diese Form der Halluzination tritt insbesondere dann auf, wenn das LLM den Bezug zum aktuellen Kontext bzw. der vorangegangenen Konversation verliert. Dieser Verlust in der Beständigkeit der Antworten wird zum Einen in der Limitierung des Langzeitgedächtnisses und zum Anderen in der mangelhaften Erfassung des aktuell relevanten Kontextes gesehen. Beispielhaft könnten anwendende Personen nach einer Empfehlung für ein Buch fragen und als Antwort der Zauberer von Oz mit der Anekdote bekommen, dass sie viel Spaß beim Besuch der Kinovorstellung haben sollen, obwohl das im aktuellen Kontext keinen Sinn ergibt und aufzeigt, dass das LLM hierbei den Bezug zur eigentlichen Kernaussage verloren hat.

*Fact-Conflicting Hallucinations* stellen die dritte Kategorie dar und bezeichnen Halluzinationen, bei denen das LLM eine gänzlich falsche Antwort auf eine Frage zurückgibt. Auch angegebene Bibliographien mit bekannten Autoren und schlüssigen Titeln können dabei komplett erfunden sein (Salvagno et al., 2023). Die Gefahr bei dieser Art von Halluzination liegt darin, dass das LLM die Antwort mit der Überzeugung wiedergibt, dass diese richtig ist, was gerade für anwendende Personen, welche die richtige Antwort nicht kennen, zu Fehlleitungen und falschen Annahmen führen kann. Laut Zhang et al., 2023 können die Gründe, die zu diesem Fehlverhalten führen, vielfältig und in unterschiedlichen Abschnitten

des LLM Lebenszyklus verortet sein. Als simples Beispiel könnte man einem LLM die Frage nach der Farbe des Himmels stellen und würde im Falle einer solchen Halluzination die Antwort grün erhalten, was faktisch falsch wäre.

Da ein virtueller Assistent im Bereich der Informationsbeschaffung diese möglichst korrekt und akkurat wiedergeben soll, stellen solche Halluzinationen ein Problem dar, welches die Zuverlässigkeit der Anwendung untergräbt. Nach Salvagno et al., 2023 liegt die Verantwortung für Korrekt- bzw. Akkuratheit somit bei den Nutzer\*Innen selbst, da KI basierte Tools, wie virtuelle Assistenten als solche verstanden und mit Bedacht verwendet werden sollen. Wie die Abbildung 3.9 aufzeigt, liegt die richtige Verwendung von KI-Tools in einer Mischung aus Vertrauen und Bedacht sowie dem Wissen um die richtige Bedienung. Solange man die Informationen, die der virtuelle Assistent bereitstellt, hinterfragt und im Zweifelsfall weitere Quellen hinzuzieht, kann dieser eine Unterstützung in Bereichen wie dem Studium darstellen.

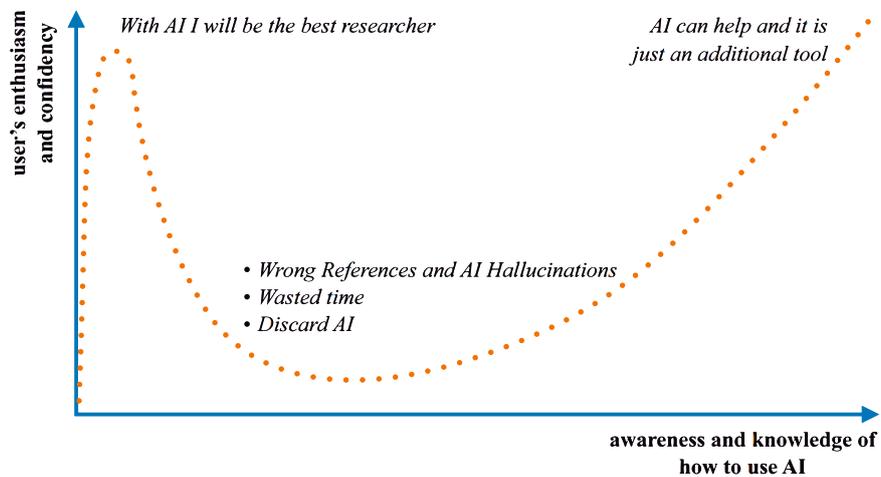


Abbildung 3.9: Grafik zur Veranschaulichung der bedachten Nutzung von KI Tools (Quelle: Salvagno et al., 2023)

### 3.5.2 Uncanny Valley Effekt

Auch in der visuellen Darstellung von virtuellen Assistenten können Probleme auftreten, die sich auf die Akzeptanz und somit die Effektivität des Prototypen auswirken. Allen voran stellt der sogenannte Uncanny Valley Effekt, welcher erstmals 1970 von Masahiro Mori, einem japanischen Ingenieur für Robotik, beschrieben wurde, einen wichtigen Faktor für die Repräsentation eines menschlichen oder menschenähnlichen Charakters dar. Dabei bezieht sich dieser Effekt auf ein starkes Gefühl des Unwohlseins oder auch der Angst, welche durch das Betrachten eines detailliert komplexen Replikats eines Menschen, wie sie beispielsweise in der Robotik, aber vor allem auch bei virtuellen Charakter Modellen vorkommen (Stein &

Ohler, 2016). Ausgelöst wird dieses Gefühl durch leichte Abweichungen in der Darstellung, die zumeist erkennen lassen, dass es sich dabei um ein künstlich geschaffenes Abbild eines Menschen handelt. Neben dem bloßen Erscheinungsbild können auch Bewegungen, die als unnatürlich wahrgenommen werden, zum Auftreten des Uncanny Valley Effektes beitragen.

In der Vergangenheit wurden zahlreiche Studien zu diesem Effekt durchgeführt, um die Zusammenhänge zwischen einer solchen visuell künstlichen Erscheinung und der daraus resultierenden kognitiven Dissonanz zu ergründen. Auch wenn es dabei verschiedene Ansätze zu den Gründen für das Auftreten des Uncanny Valley Effektes gibt, so haben diese Studien gemein, dass der Auslöser für das Gefühl des Unwohlseins in der Betrachtung von visuellen Abweichungen gegenüber realen Menschen besteht und dass es hinsichtlich des Grundes um eine Form des Selbstschutzes, also einem Instinkt, der vor potenziellen Gefahren für das eigene Leben schützen soll, handelt.

Die folgende Abbildung (3.10) demonstriert dabei, wie sich der Uncanny Valley Effekt in Bezug auf die menschliche Wahrnehmung in Form eines Funktionsgraphen einordnet. Wie der Name dieses Effektes bereits andeutet, erkennt man in der grafischen Darstellung, dass es sich hierbei um ein Tal handelt, in dem der Grad der Vertrautheit bei zunehmender menschlicher Darstellung eines Objektes abfällt, bevor im Anschluss die Kurve bei weiterhin zunehmender menschlicher Darstellung wieder stetig ansteigt und in die Wahrnehmung eines gesunden Menschen münden kann, aber nicht zwingend muss. Auffällig am Verlauf des Graphen ist dabei, dass dieser nicht gleichmäßig ansteigt, aber dafür rapide in das Tal abfällt, was nach Mori et al., 2012 vor allem bei Objekten der Fall ist, die sich ähnlich wie Menschen bewegen und verhalten sollen.

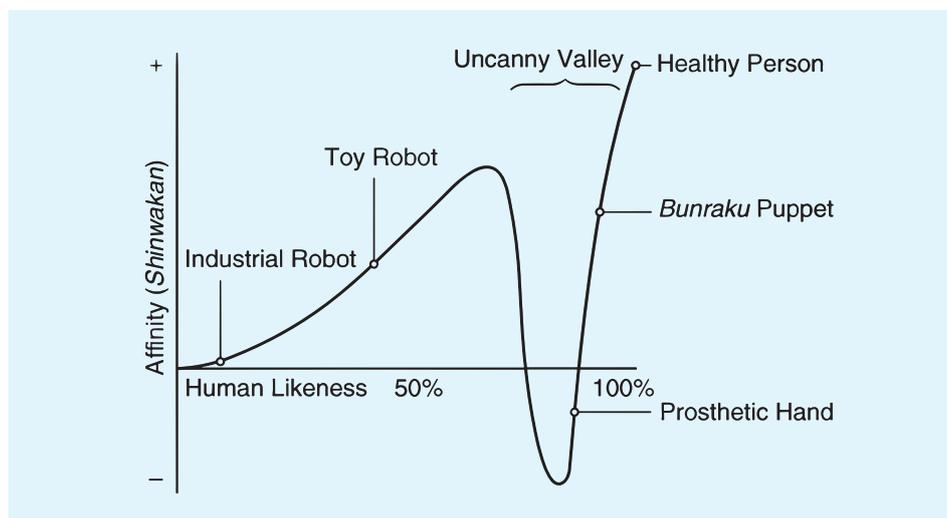


Abbildung 3.10: Grafik zur Darstellung des Uncanny Valley Effektes (Quelle: Mori et al., 2012)

Wie im Abschnitt 3.2.4 erläutert kommt für den Prototypen aus zeitlichen Gründen ein stilisiertes Charakter Modell mit Ausrichtung zu einer Cartoon haften Optik zum Einsatz, welches dabei als Ausgangspunkt für die Untersuchung hinsichtlich des Einflusses auf die Forschungsfrage dient und mittels der Nutzeranalyse evaluiert werden soll, ob diese Darstellung sich für den Einsatz eines virtuellen Assistenten im Studienalltag eignet oder sich hierbei Auswirkungen des Uncanny Valley Effektes zeigen. Da das Modell somit von der Darstellung eines realen Menschen abweicht und noch dazu menschliche Bewegungen und Verhalten mittels Animationen nachahmt, besteht hierbei die Gefahr, dass die Akzeptanz seitens der Nutzenden negativ beeinflusst wird.

### 3.5.3 Human Computer Interface

Als ein vorrangiges Problem bei der Verwendung von virtuellen Assistenten steht die Schnittstelle zwischen Mensch und Computer im Fokus dieses Abschnittes. Virtuelle Assistenten, so auch der entwickelte Prototyp, verwenden für die Interaktion mit den Anwendenden ein multimodales Interface, also eine Kombination aus mehreren verschiedenen Schnittstellen, wie in diesem Fall Sprache und Video. Mahmud et al., 2020 definiert den Zweck einer solchen Schnittstelle als ein Mittel für die robustere und natürliche Interaktion zwischen Menschen und Computern. Gerade in Hinblick auf die Benutzerfreundlichkeit, einen wichtigen Teilaspekt der Effektivität virtueller Assistenten, spielen die verwendeten Schnittstellen somit eine wichtige Rolle. Wie Sharma et al., 1998 beschreibt, ist die Interaktionen von Menschen mit ihrer Umgebung von Natur aus multimodal, somit erschwert sich die Bedienung einer Anwendung durch den Einsatz einer Einzigen unimodalen Schnittstelle und fühlt sich dabei auch deutlich unbefriedigender an, vor allem wenn Schnittstellen, wie Maus oder Tastatur, deren Gebrauch unnatürliche Bewegungen verlangen, zum Einsatz kommen.

Eine natürliche Interaktion über eine Kombination aus Sprache und Gesten ermöglicht zwar eine einfachere und zufriedenstellende Lösung, bringt aber auch Komplikationen mit sich. Im Falle eines virtuellen Assistenten, dessen Hauptaufgabe das Beantworten von Fragen darstellt, beinhalten diese vor allem den Umstand, dass gerade in belebten Umgebungen viele akustische und visuelle Einflüsse vorkommen, die es der Anwendung somit erschweren die relevanten Spracheingaben und Gesten zu erkennen und entsprechend darauf reagieren zu können. Um dieser Problematik entgegenzuwirken, verwendet der Prototyp eine Schlüsselwort Erkennung, wie sie auch in gängigen Sprach-Interfaces in der Industrie zum Einsatz kommt. Dabei soll verhindert werden, dass auf eingehende Sprache reagiert wird, die nicht an die virtuelle Assistentin gerichtet wird und somit wie ein Filter funktioniert. Die anschließende Studie soll hierbei Aufschluss darüber geben, ob diese Herangehensweise an das Problem wirklich zuträglich für die Effektivität virtueller Assistenten ist, oder sich der Einsatz eines Schlüsselwortes als zu umständlich und somit eher nachteilig auf diese auswirkt.

## 4 Ergebnisse

Nach Durchführung der in in Abschnitt 3.3 beschriebenen Nutzeranalyse werden die erhobenen Daten gemäß dem SUS Verfahren analysiert und für jeden der Probanden\*/innen ein Score als Richtwert zur Bestimmung der Effektivität virtueller Assistenten im Studium ermittelt. Zu diesem Zweck wird das von Brooke, 1995 vorgestellte Rechenverfahren verwendet, bei dem im ersten Schritt die Punkte Skala um Eins verringert wird, somit liegen die möglichen Punkte pro Aussage zwischen Null und Vier. Im nächsten Schritt wird die Punkteskala bei negativ konnotierten Aussagen, also alle mit einer geraden Nummer, umgedreht, sodass diese nun von Vier bis Null reicht. Anschließend werden alle Punkte eines Fragebogens addiert, was eine Zahl zwischen Null und Vierzig ergibt, welche zum Abschluss mit dem Faktor 2.5 multipliziert wird, um den jeweiligen Score zu erhalten.

Tabelle 4.1: Scores der Probanden\*/innen

Proband*/in	Score (absteigend sortiert)
1	85
2	80
3	72,5
4	65
5	60
6	60
7	57,5
8	57,5
9	52,5

An der Studie haben insgesamt 9 Probanden\*/innen der Fakultät Design, Medien und Information teilgenommen, deren jeweilige Scores in Tabelle 4.1 eingesehen werden können. Der daraus gebildete Durchschnitt ergibt einen abschließenden Score von 65,56 (gerundet auf zwei Stellen nach dem Komma). Damit ordnet sich das Ergebnis zwischen den von Bangor et al., 2009 aufgestellten Begriffen “OK” und “Gut” ein, was als durchschnittlich mit positiver Tendenz hinsichtlich der Effektivität gewertet werden kann (Vergleich Abbildung 4.1). Zur Einschätzung möglicher Gründe findet sich in der Tabelle 4.2 die durchschnittlich vergebenen Punkte zu jeder Aussage. Davon ausgehend, dass der Wert Zwei den Mittelwert der Skala

bildet, fällt die Aussage Nummer Sieben im direkten Vergleich mit einem Wert von rund 1,67 als unterdurchschnittlich schlecht bewertet auf. Somit lässt sich die Stimme bzw. die Aussprache der virtuellen Assistentin als einer der Hauptgründe für die geminderte Akzeptanz seitens der Probanden\*/innen unter den von der Studie erfassten Aspekten herausstellen. Damit lassen sich Parallelen zu der im Kapitel Statistik aufgeführten Umfrage zur Akzeptanz von Sprachassistenten 3.6 erkennen, wonach die subjektive Wahrnehmung der anwendenden Person in Bezug auf die Stimme ein wichtiges Kriterium für die Akzeptanz darstellt. Dabei spielt nach Aussage der Probanden\*/innen die Natürlichkeit in Klang und Aussprache eine entscheidende Rolle. Des Weiteren wurde, trotz des positiven Feedbacks in Bezug auf die Wahl einer femininen Stimme bzw. Erscheinung, die Option zur Auswahl unterschiedlich ausgerichteter Modelle, sowie unabhängig von dieser Wahl, unterschiedlicher Stimmen über die Anmerkungen im Fragebogen gewünscht.

Tabelle 4.2: Durchschnittliche Punkte pro Aussage

Nummer	Aussage	vergebene Punkte
1	Ich würde die virtuelle Assistentin bei weiteren Fragen erneut konsultieren.	2,45
2	Ich empfinde das Erscheinungsbild der Assistentin als befremdlich.	2,78
3	Meine Fragen konnten in Inhalt und Umfang vollständig beantwortet werden.	2,56
4	Ohne Hilfe wüsste ich nicht, wie die Anwendung verwendet wird.	3,00
5	Ich finde die verschiedenen Funktionen zur Interaktion sind gut eingebunden.	2,56
6	Ich denke die virtuelle Assistentin ist keine gute Alternative zu Beratungsgesprächen, Webrecherchen etc.	2,45
7	Ich empfinde die Stimme und Aussprache der Assistentin als angenehm.	1,67
8	Ich fände es besser, wenn Aussehen und Bewegungen der Assistentin realistischer wären.	2,67
9	Ich finde die UI-Elemente (Chat, Timer, Infotext) hilfreich	3,22
9	Ich bräuchte eine Eingewöhnungsphase, bevor ich die Anwendung regelmäßig benutze.	2,89

Auch die Ergebnisse der Aussagen Eins und Sechs, die mit rund 2,45 Punkten den zweitniedrigsten Wert erzielt haben, sind für die Auswertung besonders interessant. Da beide Aussagen

auf die zukünftige Verwendung der virtuellen Assistentin und somit grundlegend auf deren Effektivität abzielen, spielen diese für die Betrachtung der Forschungsfrage eine besondere Rolle. Bemerkenswert ist hierbei, dass beide Aussagen auf den gleichen Durchschnittswert bei den zugeteilten Punkten kommen, was für die Integrität der erhobenen Daten spricht. Die Anmerkungen geben für diesen Fall weiteren Aufschluss nach potenziellen Gründen, warum die Probanden\*/innen nur eine leicht positive Tendenz hinsichtlich der regelmäßigen Verwendung der virtuellen Assistentin als Alternative zu etablierten Möglichkeiten der Einholung von Informationen bzw. dem Beantworten von Fragen angaben. So wird eine zu hohe Latenz bei der Verarbeitung der Anfrage und Wiedergabe der daraus generierten Antwort als negativer Kritikpunkt angegeben, welcher innerhalb des Fragebogen nicht explizit behandelt wird, aber bei Drei der Neun Probanden\*/innen als entscheidendes Kriterium für die weitere Nutzung gesehen wird.

Mit einem Wert von 2,56 schneiden die Aussagen zur Qualität der Antworten (Nummer Drei), sowie für das Zusammenspiel der einzelnen Module (Nummer Fünf) ebenfalls mit einer positiven Tendenz ab, allerdings zeigen die Anmerkungen einen Kritikpunkt auf, welcher vermuten lässt, dass diese Wertung deutlich besser hätte ausfallen können. Bei diesem Kritikpunkt handelt es sich um die Tatsache, dass die virtuelle Assistentin noch über eine zu kleine Wissensbasis verfügt, die nicht genügend Informationen zur Beantwortung sämtlicher Studienspezifischer Fragen beinhaltet und infolgedessen ihren Verwendungszweck nicht ausreichend erfüllt, was somit auch die Effektivität negativ beeinflusst.

Anzeichen für das Auftreten des Uncanny Valley Effektes oder einer generellen Abneigung gegenüber der cartoonhaften Darstellung der virtuellen Assistentin konnte nicht festgestellt werden. Die dazugehörigen Aussagen Nummer Zwei, welche sich auf die subjektive Empfindung beim Betrachten der visuellen Erscheinung konzentriert und dabei mit einer Wertung von rund 2,78 deutlich besser abschnitt als die Wertungen zu anderen Aspekten, sowie Nummer Acht, welche negativ konnotiert auf den Bedarf nach einer realistischeren Darstellung abzielt, erhielt mit rund 2,67 Punkten ebenfalls eine höhere Wertung als andere Aspekte. Daraus lässt sich ableiten, dass ein solcher Bedarf nach mehr Realismus nicht besteht bzw. keine negativen Empfindungen bei Nutzung der Anwendung auftraten.

Im Hinblick auf die Benutzerfreundlichkeit schneiden die damit verbundenen Aussagen Nummer Vier, Nummer Neun und Nummer Zehn mit Wertungen von 3,00, rund 3,22 und rund 2,89 von allen anderen Wertungen und den dazugehörigen Aspekten am besten ab. Somit führt dieser Aspekt, in Relation zu dem durchschnittlichen Score, zu einer Steigerung der Effektivität. Gespräche mit den Probanden\*/innen wiesen daraufhin, dass die implementierte Funktionalität zur Verwendung eines Schlüsselwortes keinen negativen Einfluss auf die Benutzerfreundlichkeit hatte, da diese bereits durch andere Sprachassistenten vertraut war.

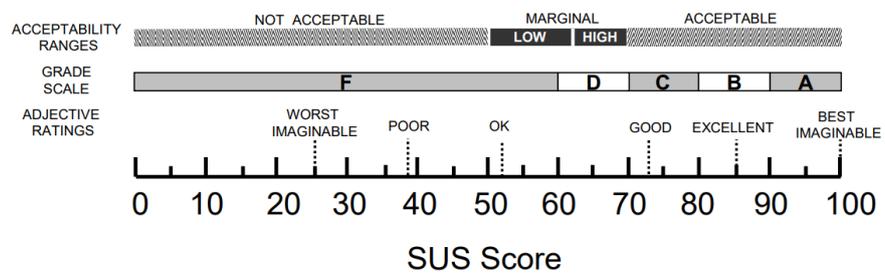


Abbildung 4.1: Grafik zur Einordnung des SUS Scores (Quelle: Bangor et al., 2009)

## 5 Fazit

Virtuelle Assistenten bergen das Potenzial, um als digitale Ansprechpartner für Studierende zur Verfügung zu stehen und diese effektiv unterstützen zu können, allerdings belegen die Ergebnisse der durchgeführten Nutzerstudie, dass es noch große Kritikpunkte gibt, die für eine effektive Nutzung von virtuellen Assistenten im Studium genauer betrachtet und gelöst werden müssen. Vor allem die Punkte zu einer individualisierbaren Repräsentation des Avatars, einer natürlicheren Aussprache und eine verminderte Latenz bei der Beantwortung der Fragen, sowie eine größere und breiter aufgestellte Wissensbasis bilden die Grundlage zu einer möglichen Verbesserung der Effektivität anhand des Prototypen.

Daraus lässt sich ableiten, dass eine höhere Effektivität durch die Weiterentwicklung des Prototypen unter Berücksichtigung der Ergebnisse und des Feedbacks erzielt werden könnte. Für den Moment jedoch lässt sich die Forschungsfrage nach der Effektivität einer virtuellen Assistentin für Studierende dahingehend beantworten, dass die allgemeine Akzeptanz seitens der Studierenden aufgrund der genannten Kritikpunkte noch zu gering ausfällt, damit diese eine gleichwertige Option zu persönlichen Gesprächen, oder Web Recherchen darstellt und auf regelmäßiger Basis konsultiert wird.

# Literatur

- Arsenyan, J., & Mirowska, A. (2021). Almost human? A comparative case study on the social media presence of virtual influencers. *International Journal of Human-Computer Studies*, 155. <https://doi.org/10.1016/j.ijhcs.2021.102694>
- Audrezet, A., & Koles, B. (2023). Virtual Influencer as a Brand Avatar in Interactive Marketing. *The Palgrave Handbook of Interactive Marketing*, 155, 353–376. [https://doi.org/10.1007/978-3-031-14961-0\\_16](https://doi.org/10.1007/978-3-031-14961-0_16)
- Bangor, A., Kortum, P., & Miller, J. (2009). Determining what individual SUS scores mean: adding an adjective rating scale. *Journal of Usability Studies*, 4, 114–123. <https://dl.acm.org/doi/10.5555/2835587.2835589>
- Brooke, J. (1995). SUS: A quick and dirty usability scale. *Usability Eval. Ind.*, 189. [https://www.researchgate.net/publication/228593520\\_SUS\\_A\\_quick\\_and\\_dirty\\_usability\\_scale](https://www.researchgate.net/publication/228593520_SUS_A_quick_and_dirty_usability_scale)
- David Burden, M. S.-B. (2019). *Virtual Humans: Today and Tomorrow* (1. Aufl.). Chapman; Hall.
- Gubareva, R., & Lopes, R. (2020). Virtual Assistants for Learning: A Systematic Literature Review. *Proceedings of the 12th International Conference on Computer Supported Education*, 1, 97–103. <https://doi.org/10.5220/0009417600970103>
- Hartholt, A., Mozgai, S., Fast, E., Liewer, M., Reilly, A., Whitcup, W., & Rizzo, A. (2019). Virtual Humans in Augmented Reality: A First Step towards Real-World Embedded Virtual Roleplayers. *Proceedings of the 7th International Conference on Human-Agent Interaction*, 205–207. <https://doi.org/10.1145/3349537.3352766>
- Kenny, P., Hartholt, A., Gratch, J., Swartout, W., Traum, D., Marsella, S., & Piepol, D. (2007). Building Interactive Virtual Humans for Training Environments. *Proceedings of The IEEE - PIIIEE*, 205–207. <https://doi.org/10.1145/3349537.3352766>
- Koles, B., Audrezet, A., Moulard, J. G., Ameen, N., & Brad, M. (2024). The authentic virtual influencer: Authenticity manifestations in the metaverse. *Journal of Business Research*, 170, 205–207. <https://doi.org/10.1016/j.jbusres.2023.114325>
- Lacity, M., Willcocks, L. P., & Craig, A. (2017). Service automation: cognitive virtual agents at SEB Bank. *The Outsourcing Unit Working Research Paper Series*, 205–207. <https://doi.org/10.1145/3349537.3352766>
- Lim, R. E., & Lee, S.-y. (2023). “You are a virtual influencer!”: Understanding the impact of origin disclosure and emotional narratives on parasocial relationships and virtual

- influencer credibility. *Computers in Human Behavior*, 148, 205–207. <https://doi.org/10.1016/j.chb.2023.107897>
- Luger, E., & Sellen, A. (2016). Like Having a Really Bad PA,: The Gulf between User Expectation and Experience of Conversational Agents. *CHI '16: Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, 5286–5297. <https://doi.org/10.1145/2858036.2858288>
- Mahmud, S., Lin, X., & Kim, J.-H. (2020). Interface for Human Machine Interaction for assistant devices: A Review. *2020 10th Annual Computing and Communication Workshop and Conference (CCWC)*, 0768–0773. <https://doi.org/10.1109/CCWC47524.2020.9031244>
- Medium. (2018, 7. Mai). *Amelia chatbot review*. <https://doi.org/10.1145/2858036.2858288>
- Mori, M., MacDorman, K. F., & Kageki, N. (2012). The Uncanny Valley [From the Field]. *IEEE Robotics and Automation Magazine*, 19, 98–100. <https://doi.org/10.1109/MRA.2012.2192811>
- Nadia Magnenat-Thalmann, D. T. (2004, 7. Mai). *Handbook of Virtual Humans* (1. Aufl.). John Wiley; Sons. <https://doi.org/10.1145/2858036.2858288>
- Richard C. Ward, N. B. M. u. S. K. F., Line C. Pouchard. (2008, 7. Mai). *Virtual Human Problem-Solving Environments* (1. Aufl.). Springer. <https://doi.org/10.1145/2858036.2858288>
- Salvagno, M., Taccone, F. S., & Gerli, A. G. (2023). Artificial intelligence hallucinations. *Critical Care*, 5286–5297. <https://doi.org/10.1186/s13054-023-04473-y>
- Sharma, R., Pavlovic, V., & Huang, T. (1998). Toward multimodal human-computer interface. *Proceedings of the IEEE*, 86, 853–869. <https://doi.org/10.1109/5.664275>
- Statista.com. (2023, 7. Mai). *Digitale Sprachassistenten*. <https://doi.org/10.1145/2858036.2858288>
- Stein, J.-P., & Ohler, P. (2016). Venturing into the uncanny valley of mind—The influence of mind attribution on the acceptance of human-like characters in a virtual reality setting. *Cognition*, 160, 43–50. <https://doi.org/10.1016/j.cognition.2016.12.010>
- Zhang, Y., Li, Y., Cui, L., Cai, D., Liu, L., Fu, T., Huang, X., Zhao, E., Zhang, Y., Chen, Y., Wang, L., Luu, A. T., Bi, W., Shi, F., & Shi, S. (2023). Siren’s Song in the AI Ocean: A Survey on Hallucination in Large Language Models. *ArXiv*, 5286–5297. <https://doi.org/10.48550/arXiv.2309.01219>

# Anhang

## Bedienungsanleitung

Sobald die Applikation gestartet wurde, können mittels des Schlüsselwortes „AVA“ Fragen gestellt werden, ähnlich wie bei gängigen Smart Speakern. Ein Icon oberhalb des Chatfensters auf der linken Seite des Bildschirms signalisiert dabei, ob die Aufnahme gestartet wurde. Anschließend haben Anwender\*innen Acht Sekunden Zeit Ihre Frage zu stellen, was durch einen kreisförmigen Timer visualisiert wird. Zur Kontrolle und besseren Übersicht wird die Konversation dabei in dem bereits erwähnten Chatfenster angezeigt.

## Installationsanleitung

Zur Inbetriebnahme der Anwendung ist es notwendig den dazugehörigen Webservice zu starten. Dieser Webservice basiert auf Python Flask und ist so konfiguriert, dass er auf derselben Maschine läuft, wie die Endanwendung. Sollte die Notwendigkeit bestehen, dass dieser auf einem eigenen Server installiert werden soll, so müssen die jeweiligen Urls innerhalb der Unity Skripte angepasst und die Domain von localhost auf die entsprechende Server Domain geändert werden.

Für die initiale Einrichtung und die spätere Ausführung des Webservices wird die Verwendung von Visual Studio Code bzw. dessen Terminal inklusive der Python Extension empfohlen. Der Webservice läuft in einer virtuellen Python Umgebung und benötigt daher zunächst eine Installation von Python 3.10, oder höher, sowie die Rechte zum Ausführen von Skripten. Dies wird durch den folgenden Command ermöglicht:

```
Set-ExecutionPolicy Unrestricted -Scope Process
```

Zum initialisieren und wechseln auf die virtuelle Umgebung:

```
python -m venv api_env  
.\api_env\Scripts\Activate.ps1
```

Zur erstmaligen Einrichtung ist es notwendig alle benötigten Libraries bzw Packages zu installieren. Hierfür wird der Python Packagemanager PIP benötigt. Sofern dieser installiert wurde können die benötigten Packages wie folgt installiert werden:

```
pip install -r .\requirements.txt
```

Die Datei requirements.txt enthält hierbei die Auflistung sämtlicher benötigten Packages. Damit die Module zur Verarbeitung von Sound Dateien fehlerfrei funktionieren ist es bei der Ersteinrichtung erforderlich das multimedia Framework FFmpeg zu installieren. Es wird hierbei empfohlen den Paketmanager Chocolatey über eine Powershell Konsole mit Adminrechten zu verwenden. Die Installation von FFmpeg erfolgt via:

```
choco install ffmpeg
```

Im Anschluss daran empfiehlt es sich den PC einmal neu zu starten, damit die Änderungen übernommen werden. Nach dem Neustart und vor jedem Start des Webservices müssen zunächst wieder die Ausführrechte bestätigt und in die virtuelle Umgebung gewechselt werden. (siehe erste Schritte) Danach kann der Webservice gestartet werden:

```
python api_service.py
```

Da das Projekt eine Anbindung an Open AI's Api vorsieht, wird zudem ein gültiger Api Key benötigt, welcher in der Datei constants.py angegeben sein muss.

Für die Inbetriebnahme der Gestenerkennung erfordert es die ersten Schritte noch einmal für den Projektordner „GestureRecognition“ zu wiederholen:

```
Set-ExecutionPolicy Unrestricted -Scope Process
```

Zum initialisieren und wechseln auf die virtuelle Umgebung:

```
python -m venv python_env
```

```
.\python_env\Scripts\Activate.ps1
```

Auch hier müssen bei der erstmaligen Installation alle benötigten Pakete und Abhängigkeiten installiert werden:

```
pip install -r .\requirements.txt
```

Zum Starten muss anschließend die Anwendung unter .\Unity\BA\_Project\StableBuild\BA\_Project ausgeführt werden.

# Fragebogen

**KI im Studium - Untersuchung zur Effektivität von Digital Humans als virtuelle Assistentin für Studierende**

	Tritt überhaupt nicht zu					Tritt komplett zu
<b>1</b> Ich würde die virtuelle Assistentin bei weiteren Fragen erneut konsultieren.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
<b>2</b> Ich empfinde das Erscheinungsbild der Assistentin als befremdlich.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
<b>3</b> Meine Fragen konnten in Inhalt und Umfang vollständig beantwortet werden.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
<b>4</b> Ohne Hilfe wüsste ich nicht, wie die Anwendung verwendet wird.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
<b>5</b> Ich finde die verschiedenen Funktionen zur Interaktion sind gut eingebunden.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
<b>6</b> Ich denke die virtuelle Assistentin ist keine gute Alternative zu Beratungsgesprächen, Webrecherchen etc.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
<b>7</b> Ich empfinde die Stimme und Aussprache der Assistentin als angenehm.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
<b>8</b> Ich fände es besser, wenn Aussehen und Bewegungen der Assistentin realistischer wären.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
<b>9</b> Ich finde die UI-Elemente (Chat, Timer, Infotext) hilfreich.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
<b>10</b> Ich bräuchte erst eine Eingewöhnungsphase, bevor ich die Anwendung regelmäßig benutze.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Anmerkungen:

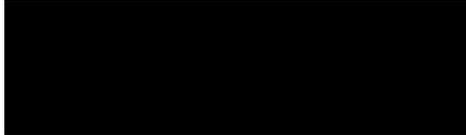
Abbildung 1: Aufbau Fragebogen

# Eigenständigkeitserklärung

Hiermit versichere ich, dass ich die vorliegende Bachelorarbeit mit dem Titel

**KI im Studium - Untersuchung zur Effektivität von Digital Humans als virtuelle Assistent\*innen für Studierende**

selbstständig und nur mit den angegebenen Hilfsmitteln verfasst habe. Alle Passagen, die ich wörtlich aus der Literatur oder aus anderen Quellen wie z. B. Internetseiten übernommen habe, habe ich deutlich als Zitat mit Angabe der Quelle kenntlich gemacht.



Hamburg, 08. Februar 2024