

Bachelorarbeit  
im Bachelorstudiengang  
**Wirtschaftspsychologie**  
an der Hochschule für angewandte Wissenschaften Neu-Ulm

**Der Einfluss des Geschlechts von KI-generierten Lehrpersonen auf Lerner-  
folg und Kompetenzwahrnehmung**

Erstkorrektor/-in: Prof. Dr. Michaela Eßbach  
Zweitkorrektor/-in: Prof. Dr. Johannes Basch

Verfasser/-in: Alexia Roxana Schneider (Matrikel-Nr.: 315597)

Thema erhalten: 11.12.2025  
Arbeit abgegeben: 12.03.2026

## Abstract

Mit der zunehmenden Integration KI-generierter Lehrpersonen in digitalen Lernumgebungen stellt sich die Frage, inwiefern soziale Kategorien wie Geschlecht deren Wahrnehmung und Wirksamkeit beeinflussen. Die vorliegende Studie untersucht, ob das Geschlecht einer KI-generierten Lehrperson die wahrgenommene didaktische Kompetenz sowie den Lernerfolg der Teilnehmenden beeinflusst. Auf Grundlage der Role Congruity Theory wurde angenommen, dass kongruente Kombinationen aus Geschlecht der KI-generierten Lehrperson und geschlechtlich konnotierter Fachdomäne zu positiveren Bewertungen und höherer Lernleistung führen. In einem experimentellen Design wurden eine männliche und eine weibliche KI-generierte Lehrperson in zwei Fachdomänen, Informatik und Arbeitsgesundheit, eingesetzt. Entgegen den Annahmen der Role Congruity Theory wurde die weibliche KI-generierte Lehrperson bei Kontrolle des wahrgenommenen Realismus kompetenter bewertet. Darüber hinaus erzielten Teilnehmende in rolleninkongruenten Bedingungen höhere Lernleistungen als in rollenkongruenten Bedingungen. Die Befunde werden vor dem Hintergrund bestehender theoretischer Ansätze eingeordnet und hinsichtlich ihrer praktischen Implikationen diskutiert.

# Inhaltsverzeichnis

<b>ABBILDUNGSVERZEICHNIS</b>	<b>4</b>
<b>TABELLENVERZEICHNIS</b>	<b>5</b>
<b>1 EINLEITUNG</b>	<b>6</b>
<b>2 THEORETISCHER UND EMPIRISCHER HINTERGRUND</b>	<b>8</b>
2.1 Geschlechterstereotype und Fachdomänen .....	8
2.2 Role Congruity Theory .....	9
2.3 Digitale Avatare im Bildungskontext .....	11
2.4 Übersicht der Studie und Hypothesen .....	12
<b>3 METHODE</b>	<b>14</b>
3.1 Versuchsplanung .....	14
3.2 Stichprobe .....	15
3.3 Versuchsmaterialien .....	16
3.4 Versuchsablauf .....	19
<b>4 ERGEBNISSE</b>	<b>21</b>
4.1 Manipulationskontrolle .....	21
4.2 Befunde zu den Hypothesen .....	22
4.3 Weiterführende Analysen .....	24
<b>5 DISKUSSION</b>	<b>28</b>
<b>LITERATUR</b>	<b>31</b>
<b>ÜBERSICHT VERWENDETER HILFSMITTEL</b>	<b>37</b>
<b>ANHANG</b>	<b>38</b>

## Abbildungsverzeichnis

Abbildung 1 <i>Titelbilder der Lehrvideos in den vier experimentellen Bedingungen. Dargestellt sind die kongruenten und inkongruenten Kombinationen der Variablen Fachbereich (Informatik vs. Arbeitsgesundheit) und Geschlecht der Lehrperson (männlich vs. weiblich).....</i>	17
Abbildung 2 <i>Mittelwerte der didaktischen Kompetenz nach Avatar-Geschlecht und Fachdomäne (Fehlerbalken = 95%-Konfidenzintervalle).....</i>	22
Abbildung 3 <i>Lernerfolg (Erreichte Punktzahlen im Wissenstest) nach Avatar-Geschlecht und Fachdomäne (Fehlerbalken = 95%-Konfidenzintervalle).....</i>	23

## Tabellenverzeichnis

Tabelle 1 <i>Stichprobenverteilung in den experimentellen Bedingungen</i> .....	15
Tabelle 2 <i>Interne Konsistenz der Variable Wahrnehmung didaktischer Kompetenz nach experimenteller Gruppe</i> .....	18
Tabelle 3 <i>Dreifaktorielle Varianzanalyse für die wahrgenommene didaktische Kompetenz unter Einbezug des Teilnehmergeschlechts</i> .....	24
Tabelle 4 <i>Dreifaktorielle Varianzanalyse für den Lernerfolg unter Einbezug des Teilnehmergeschlechts</i> .....	24
Tabelle 5 <i>Kovarianzanalyse (ANCOVA) zum Einfluss von Fachdomäne und Avatar-Geschlecht auf die wahrgenommene didaktische Kompetenz unter Kontrolle des Alters</i> .....	25
Tabelle 6 <i>Kovarianzanalyse (ANCOVA) zum Einfluss von Fachdomäne und Avatar-Geschlecht auf den Lernerfolg unter Kontrolle des Alters</i> .....	25
Tabelle 7 <i>Varianzanalyse für den wahrgenommenen Realismus nach experimenteller Gruppe</i> .....	26
Tabelle 8 <i>Lineare Regressionsanalyse zur Vorhersage der wahrgenommenen didaktischen Kompetenz durch das Avatar-Geschlecht</i> .....	27
Tabelle 9 <i>Multiple Regressionsanalyse zur Vorhersage der didaktischen Kompetenz durch das Avatar-Geschlecht unter Kontrolle des wahrgenommenen Realismus</i> .....	27

# 1 Einleitung

KI-Avatare haben sich von einer Nischentechnologie zu einer zunehmend relevanten IT-Lösung für unterschiedliche Unternehmensbranchen und -bereiche entwickelt (Wadhvani, P. & Ambekar, A., 2024). Insbesondere in der Bildungsbranche sowie in der Aus- und Weiterbildung von Mitarbeitenden werden sie vermehrt genutzt (*E-Learning-Studie 2025*, 2025; Wadhvani, P. & Ambekar, A., 2024). KI-Avatare können dabei als virtuell erzeugte Lehrpersonen fungieren, die Lerninhalte in Form von Videos menschenähnlich präsentieren oder auch interaktiv mit Lernenden kommunizieren (Fink et al., 2024). Aktuelle Fallstudien von Rizvi (2024) zeigen anhand etablierter Bildungsorganisationen wie der American Psychological Association oder Berlitz, dass der Einsatz solcher KI-generierten Lehrpersonen unter anderem mit verkürzten Produktionszeiten für Lehrvideos, einem geringeren Bedarf an personellen Ressourcen sowie einer erhöhten Flexibilität bei der Aktualisierung von Lehrmaterialien einhergeht.

Mit der zunehmenden Verbreitung von KI-Avataren in digitalen Lernumgebungen rückt jedoch nicht nur deren technische Leistungsfähigkeit in den Fokus, sondern auch ihre soziale Wahrnehmung durch Lernende. Lehrsituationen sind nicht ausschließlich durch die vermittelten Inhalte geprägt, sondern werden auch von den Eigenschaften der digitalen Lehrperson beeinflusst (Armando et al., 2022). In der Personenwahrnehmung spielt das Geschlecht eine zentrale Rolle und ist auch im Bildungskontext häufig mit stereotypen Annahmen und Erwartungen an Lehrpersonen verbunden (Aragón et al., 2023).

Frühere Forschung zu digitalen Avataren weist darauf hin, dass das Geschlecht des Avatars die Bewertung der digitalen Lehrperson als auch den Lernerfolg der lernenden Person beeinflussen kann, wenngleich die bisherigen Befunde nicht eindeutig sind (Armando et al., 2022). Ob sich diese Effekte auch auf KI-generierte Lehrpersonen übertragen lassen, ist bislang nicht abschließend geklärt und gewinnt angesichts der zunehmenden Nutzung entsprechender Technologien in Unternehmen an Bedeutung. Vor diesem Hintergrund untersucht die vorliegende Arbeit, ob das Geschlecht einer KI-generierten Lehrperson die wahrgenommene didaktische Kompetenz des Avatars sowie den Lernerfolg der Teilnehmenden beeinflusst.

Durch die empirische Untersuchung dieser Zusammenhänge leistet die Arbeit einen Beitrag zur Genderforschung, zur Bildungspsychologie sowie zur Forschung im Bereich der Mensch-Maschine-Interaktion.

## 2 Theoretischer und empirischer Hintergrund

Die vorliegende Arbeit befasst sich mit den Auswirkungen des Geschlechts von KI-generierten Lehrpersonen in digitalen Lernumgebungen. Zur theoretischen Fundierung werden zunächst Geschlechterstereotype und deren Bedeutung für Fachdomänen erläutert (Abschnitt 2.1). Anschließend wird die Role Congruity Theory als zentraler Erklärungsansatz für geschlechtsbezogene Bewertungsunterschiede vorgestellt (Abschnitt 2.2). Darauf aufbauend werden KI-Avatare als soziale Akteure im Bildungskontext eingeordnet und bisherige Befunde zu Geschlechtereffekten digitaler Lehrpersonen zusammengefasst (Abschnitt 2.3). Abschließend werden die Fragestellung der vorliegenden Studie sowie die daraus abgeleiteten Hypothesen dargestellt (Abschnitt 2.4).

### 2.1 Geschlechterstereotype und Fachdomänen

Geschlechterstereotype sind sozial geteilte Annahmen über typische Eigenschaften und Verhaltensweisen von Frauen und Männern (Boring, 2017). Sie strukturieren soziale Wahrnehmung und beeinflussen Erwartungen gegenüber Mitgliedern einer Geschlechtergruppe (Heilman et al., 2024). Zur Erklärung der Entstehung solcher Stereotype wird häufig die Social Role Theory herangezogen (Koenig & Eagly, 2014). Nach der Social Role Theory entstehen Geschlechterstereotypen aus historisch gewachsenen Arbeitsteilungen, insbesondere aus der Verteilung von Erwerbsarbeit, Führungspositionen und Sorgearbeit (Lange et al., 2012). Beobachtende schließen von den Tätigkeiten, die Frauen und Männer häufig ausüben, auf entsprechende persönliche Eigenschaften (Lange et al., 2012). Vor diesem Hintergrund werden Frauen häufiger mit kommunalen Eigenschaften wie Fürsorglichkeit, Hilfsbereitschaft und sozialer Sensibilität assoziiert, während agentische Eigenschaften wie Durchsetzungsfähigkeit, Selbstsicherheit und Führungsorientierung stärker Männern zugeschrieben werden (Eagly & Karau, 2002). Trotz gesellschaftlicher Veränderungen in der Erwerbs- und Rollenverteilung zeigen sich nach wie vor deutliche geschlechtsspezifische Muster in unterschiedlichen Fach- und Berufsfeldern (Suhr, 2023). Eine Längsschnittanalyse aus den USA verdeutlicht zudem, dass

agentische Merkmale weiterhin eher Männern zugeschrieben werden und kommunale Eigenschaften verstärkt mit Frauen assoziiert werden (Eagly et al., 2020). Neuere Befunde ergänzen die Social Role Theory und zeigen, dass nicht nur die Art der Tätigkeit stereotype Zuschreibungen beeinflusst, sondern auch die numerische Geschlechterverteilung in Berufen und Fachdomänen. Tätigkeitsfelder, in denen Männer überrepräsentiert sind, werden häufig als männlich konnotiert wahrgenommen, im Vergleich zu Bereichen mit höherem Frauenanteil oder ausgeglichener Geschlechterverteilung (Cheryan et al., 2017; Heilman et al., 2024). Umgekehrt werden Domänen mit hohem Frauenanteil, wie die Pflege oder die Kindheitspädagogik, häufiger mit kommunalen Eigenschaften verbunden (Boring, 2017). Ähnliche Befunde sind im Bildungskontext wiederzufinden. Je stärker ein Fachbereich durch eine geschlechtsspezifische Verteilung geprägt ist, desto eher werden entsprechende Rollenerwartungen an Lehrpersonen herangetragen (Aragón et al., 2023). Aus diesen Befunden lässt sich ableiten, dass Geschlechterstereotype kontextabhängig sind. Insbesondere die Fachdomäne könnte beeinflussen, in welchem Ausmaß Lehrpersonen geschlechtsbezogenen Zuschreibungen und Bewertungsmustern unterliegen.

## 2.2 Role Congruity Theory

Die Role Congruity Theory von Eagly und Karau (2002) bietet einen zentralen theoretischen Rahmen zur Erklärung geschlechtsbezogener Vorurteile. Geschlechterrollen bezeichnen gesellschaftlich geteilte Erwartungen darüber, welche Eigenschaften, Verhaltensweisen und Kompetenzen Frauen und Männern typischerweise zugeschrieben oder von ihnen erwartet werden. Eagly und Karau (2002) gehen davon aus, dass negative Bewertungen dann entstehen, wenn die wahrgenommenen Eigenschaften einer Person nicht mit den als erforderlich betrachteten Eigenschaften einer sozialen Rolle übereinstimmen. Eine solche Inkongruenz zwischen Geschlecht und Rollenanforderungen kann zwei Formen von Bewertungsnachteilen nach sich ziehen. Erstens können Personen als weniger geeignet oder kompetent für eine Position eingeschätzt werden, wenn ihr Geschlecht nicht mit stereotypen Rollenerwartungen übereinstimmt. Zweitens kann identisches Verhalten unterschiedlich bewertet werden, je nachdem, ob es von einer als rollenkongruent oder rolleninkongruent wahrgenommenen Person gezeigt wird (Eagly & Karau, 2002).

Empirische Befunde stützen diese Annahmen. In männlich dominierten Bereichen erfahren Frauen häufig Benachteiligungen in Form geringerer Kompetenzzuschreibungen, schlechterer Leistungsbewertungen oder eingeschränkter Aufstiegschancen (Heilman et al., 2024). Umgekehrt können auch Männer in weiblich dominierten Berufsfeldern negative Bewertungen oder geringere soziale Akzeptanz erleben (Croft et al., 2015). Auch im Hochschulkontext lassen sich entsprechende Muster beobachten. Studien zu Lehrveranstaltungsevaluationen zeigen, dass weibliche Lehrpersonen im Durchschnitt schlechtere Bewertungen als ihre männlichen Kollegen erhalten, wenn in den Fachbereichen Männer stärker vertreten sind (McPherson et al., 2009; Boring, 2017). Aragón et al. (2023) untersuchten universitäre Lehrveranstaltungsevaluationen konkret im Kontext der Role Congruity Theory. In einem ersten Schritt analysierten sie Archivdaten aus sieben Universitäten in den USA. Die Ergebnisse zeigten, dass Lehrpersonen tendenziell positiver bewertet wurden, wenn ihr Geschlecht der Geschlechtermehrheit des jeweiligen Fachbereichs entsprach. In einer anschließenden experimentellen Laborstudie wurde die Geschlechterverteilung innerhalb einer Fachdomäne gezielt manipuliert. Studierende sollten sich vorstellen, an einem Kurs teilzunehmen, und diesen im Anschluss evaluieren. Dabei wurden sie explizit über die Geschlechterzusammensetzung des jeweiligen Fachbereichs informiert. Die Ergebnisse zeigten, dass diese Manipulation signifikant wirksam war: Die wahrgenommene Geschlechterverteilung beeinflusste systematisch, ob weibliche oder männliche Lehrpersonen positiver bewertet wurden. Dies zeigte sich ebenso für Fachdomänen, die nicht eindeutig stereotypisch konnotiert waren. Diese Befunde sprechen dafür, dass Rollenkongruenz auch im Bildungskontext einen Einfluss auf die Bewertung der Lehrperson hat.

Vor dem Hintergrund dieser Befunde erscheint es plausibel, dass die Role Congruity Theory auch auf digitale Lehrpersonen übertragbar sein könnte. Ein weiblicher Avatar könnte demnach negativer bewertet werden, wenn das Lehrvideo eine Fachdomäne behandelt, die stereotypisch als männlich konnotiert gilt; entsprechend wäre auch für männliche Avatare in stereotyp weiblich konnotierten Domänen mit Bewertungsnachteilen zu rechnen. Für die vorliegende Studie wurden zwei Fachdomänen ausgewählt, die sich deutlich hinsichtlich ihrer Geschlechterverteilung unterscheiden. Als stereotypisch männlich konnotierte Domäne wurde Informatik gewählt, da der Frauenanteil in diesem Bereich im Jahr 2023 bei rund 18 % lag

(*Frauenanteil in der technischen Forschung und Entwicklung binnen zehn Jahren von 11 % auf 18 % gestiegen*, 2024). Als stereotypisch weiblich konnotierte Domäne wurde hingegen ein arbeitsgesundheitliches Thema herangezogen, das dem Gesundheits- und Sozialwesen zuzuordnen ist. In diesem Bereich sind Frauen mit etwa 77 % deutlich überrepräsentiert (Urmersbach B., 2025). Diese Gegenüberstellung könnte eine systematische Operationalisierung von Rollenkongruenz im Kontext digitaler Lehrpersonen ermöglichen.

### **2.3 Digitale Avatare im Bildungskontext**

KI-Avatare sind KI-gesteuerte digitale Figuren, die häufig auf Large Language Models basieren und durch die Kombination von Sprache und Mimik soziale Interaktionen simulieren können (Fink et al., 2024). Dass Menschen mit digitalen Systemen sozial interagieren, lässt sich durch das Konzept des Anthropomorphismus erklären. Anthropomorphismus beschreibt die Tendenz, nicht-menschlichen Entitäten menschliche Eigenschaften zuzuschreiben (Craiu & Iancu, 2022). In der Interaktion mit digitalen Agenten wenden Menschen häufig ähnliche soziale Bewertungsmechanismen an wie im zwischenmenschlichen Kontext (Guzman & Lewis, 2019). Dadurch können auch soziale Kategorien wie Geschlecht auf KI-Avatare übertragen werden (Craiu & Iancu, 2022). Forschungen zu technologischen Agenten zeigen, dass Geschlechterstereotype nicht auf menschliche Interaktionspartner beschränkt sind (Craiu & Iancu, 2022). Beispielsweise zeigt sich, dass geschlechtlich gestaltete Chatbots abhängig vom Aufgabenbereich unterschiedlich wahrgenommen werden. So berichten McDonnell & Baxter (2019), dass Nutzende insbesondere dann auf Geschlechterstereotype zurückgreifen, wenn ein Chatbot in einer geschlechtskonnotierten Domäne, etwa im Bereich Mechanik, eingesetzt wird oder das dargestellte Geschlecht nicht den erwarteten Stereotypen entspricht.

Wie sich solche Geschlechterstereotype konkret auf KI-Avatare im Bildungskontext auswirken, ist bislang jedoch unzureichend erforscht. Hinweise geben Befunde zu sogenannten pädagogischen Agenten, die als Vorläufer moderner KI-Avatare im Lehrkontext verstanden werden können (Fink et al., 2024). Eine Übersicht früherer Studien zeigt, dass die Effekte des Geschlechts dieser Agenten auf Lernende uneinheitlich ausfallen und stark von Studiendesign, Aufgabenart und Messmethoden abhängen (Armando et al., 2022). Moreno et al. (2002) berichten beispielsweise

höhere Leistungswerte in einem Multiple-Choice-Test, wenn ein männlicher Agent ein Lehrvideo zum Thema Blutdruck präsentierte. Auch Kim et al. (2007) untersuchten in zwei Experimenten den Einfluss des Geschlechts pädagogischer Agenten. Im ersten Experiment zeigte sich in einem Informatikkurs ein höheres Interesse an einem männlichen Agenten. Zudem wurden männliche Agenten als engagierter, unterstützender und menschenähnlicher wahrgenommen. Im zweiten Experiment zeigte sich darüber hinaus eine höhere Lernleistung, wenn ein männlicher Agent im Kontext der Entwicklung eines E-Learning-Kurses eingesetzt wurde. Shiban et al. (2015) analysierten den Einfluss des Geschlechts anhand eines digitalen Statistik-Lehrvideos. Die Ergebnisse zeigten, dass weibliche Agenten stärker Interesse und Motivation förderten, während männliche Agenten als glaubwürdiger und kompetenter wahrgenommen wurden und tendenziell mit höheren Leistungswerten einhergingen.

Insgesamt deuten die bisherigen Befunde darauf hin, dass das wahrgenommene Geschlecht digitaler Lehrpersonen sowohl Kompetenzzuschreibungen als auch lernrelevante Variablen beeinflussen kann. Die Studien zu pädagogischen Agenten basieren jedoch nicht explizit auf der Role Congruity Theory und leiten ihre Hypothesen nicht systematisch aus geschlechtsbezogenen Rollenerwartungen ab. Entsprechend bleibt unklar, in welchen fachlichen Kontexten Effekte auf den Lernerfolg auftreten und ob diese konsistent durch Rollenkongruenz zwischen Geschlecht und Fachdomäne erklärbar sind. Zwar finden sich Hinweise auf potenziell kongruente Muster, wie etwa höhere Lernleistungen bei einem männlichen Agenten in einem Statistik-Lehrvideo bei Shiban et al. (2015), jedoch wurde dieser Zusammenhang bislang nicht theoriegeleitet im Sinne der Role Congruity Theory überprüft. Vor diesem Hintergrund erscheint es sinnvoll, neben Bewertungsmaßen auch den Lernerfolg als abhängige Variable systematisch zu untersuchen, um mögliche rollenkongruente Effekte nicht nur auf Wahrnehmungsebene, sondern auch auf Leistungsebene zu erfassen.

## 2.4 Übersicht der Studie und Hypothesen

Zusammenfassend zeigen theoretische und empirische Befunde, dass Geschlechterstereotype Bewertungen in Abhängigkeit vom situativen Kontext beeinflussen

können. Die Role Congruity Theory legt nahe, dass Inkongruenzen zwischen Geschlecht und Rollenanforderungen zu geringeren Kompetenzzuschreibungen führen können. Gleichzeitig zeigen Studien zu pädagogischen Agenten, dass Geschlecht auch im digitalen Lehrkontext eine wahrnehmbare und bewertungsrelevante Kategorie darstellt. Die vorliegende Studie untersucht daher, ob das Geschlecht einer KI-generierten Lehrperson, in Abhängigkeit von der Fachdomäne, die wahrgenommene didaktische Kompetenz sowie den Lernerfolg beeinflusst. Dabei wird angenommen, dass Effekte insbesondere zwischen kongruenten und inkongruenten Bedingungen bestehen. Es werden folgende Hypothesen geprüft:

**H1:** Lehrpersonen werden als didaktisch kompetenter wahrgenommen, wenn ihr Geschlecht stereotypisch kongruent mit der jeweiligen Fachdomäne ist, im Vergleich zu inkongruenten Bedingungen.

**H1a:** In einem Informatik-Lehrvideo wird die Lehrperson höher in didaktischer Kompetenz eingeschätzt, wenn sie männlich ist (kongruent), im Vergleich zu einer weiblichen Lehrperson (inkongruent).

**H1b:** In einem arbeitsgesundheitlichen Lehrvideo wird die Lehrperson höher in didaktischer Kompetenz eingeschätzt, wenn sie weiblich ist (kongruent), im Vergleich zu einer männlichen Lehrperson (inkongruent).

**H2:** Lernende erzielen einen höheren Lernerfolg, wenn das Geschlecht der Lehrperson mit den gesellschaftlich stereotypisierten Erwartungen an die jeweilige Fachdomäne übereinstimmt, im Vergleich zu inkongruenten Bedingungen.

**H2a:** In einem Informatik-Lehrvideo erzielen Teilnehmende eine höhere Lernleistung, wenn die Lehrperson männlich ist (kongruent), im Vergleich zu einer weiblichen Lehrperson (inkongruent).

**H2b:** In einem arbeitsgesundheitlichen Lehrvideo erzielen Teilnehmende eine höhere Lernleistung, wenn die Lehrperson weiblich ist (kongruent), im Vergleich zu einer männlichen Lehrperson (inkongruent).

## 3 Methode

### 3.1 Versuchsplanung

Die Studie wurde als 2 x 2 Between-Subjects-Design konzipiert. Die unabhängigen Variablen waren das Geschlecht der KI-generierten Lehrperson mit den Ausprägungen männlich und weiblich sowie die Fachdomäne des Lehrvideos mit den Ausprägungen Informatik und Arbeitsgesundheit. Daraus ergaben sich vier experimentelle Bedingungen. Die Teilnehmenden wurden per Zufallszuweisung einer der vier Bedingungen zugeordnet.

Als abhängige Variablen wurden der Lernerfolg der Teilnehmenden sowie die wahrgenommene didaktische Kompetenz der KI-generierten Lehrperson erhoben. Der Lernerfolg wurde anhand eines Wissenstests zum jeweiligen Lehrvideo erfasst. Die wahrgenommene didaktische Kompetenz wurde mittels mehrerer Items auf einer fünfstufigen Likert-Skala gemessen. Eine nähere Beschreibung zur Messung der abhängigen Variablen findet sich im Abschnitt 3.3 Versuchsmaterialien. Die Auswertung der Hypothesen H1 und H2 erfolgte mittels zweifaktorieller ANOVA.

Zusätzlich wurden weitere Variablen erhoben, die nicht Teil der Hypothesenprüfung waren. Die subjektive Stereotypisierung der Fachdomäne diente als Manipulationskontrolle zur Validierung der experimentellen Stimuli. Hierbei gaben die Teilnehmenden an, inwieweit sie die beiden Fachdomänen Informatik mit Männern und Arbeitsgesundheit mit Frauen assoziierten. Zudem wurde das selbstberichtete Vorwissen in der jeweiligen Fachdomäne als Kontrollvariable erfasst, um potenzielle Verzerrungen des Lernerfolgs durch unterschiedliche Wissensvoraussetzungen zu berücksichtigen. Darüber hinaus wurden demografische Angaben zum Alter und Geschlecht der Teilnehmenden erhoben. Explorativ wurde gemessen, wie realistisch der KI-Avatar wahrgenommen wurde. Das Studiendesign sowie die geplanten Auswertungsverfahren wurden vor Beginn der Datenerhebung präregistriert.

## 3.2 Stichprobe

An der Studie nahmen insgesamt  $N = 129$  Personen teil. Das Alter der Teilnehmenden lag zwischen 18 und 66 Jahren ( $M = 24$ ,  $SD = 8.2$ ). Von den Teilnehmenden gaben 100 Personen an, weiblich zu sein (77,5%), 28 Personen männlich zu sein (21,7%) und 1 Person divers zu sein (0,9%). Zwischen den experimentellen Gruppen zeigten sich deskriptiv keine großen Unterschiede im Alter und in der Geschlechterzusammensetzung. Tabelle 1 gibt einen Überblick über die Verteilung der Teilnehmenden auf die vier experimentellen Bedingungen des  $2 \times 2$ -Designs.

**Tabelle 1**

*Stichprobenverteilung in den experimentellen Bedingungen*

UV Avatar Geschlecht	UV Fachdomäne	
	Informatik	Arbeitsgesundheit
Männlich	$N = 29$ (kongruent)	$N = 29$ (inkongruent)
Weiblich	$N = 34$ (inkongruent)	$N = 37$ (kongruent)

Es wurde eine Gelegenheitsstichprobe erhoben. Primär richtete sich die Studie an Studierende der Hochschule Neu-Ulm sowie an Psychologiestudierende der Universität Ulm. Diese wurden über interne Verteiler und der Plattform SONA zur Teilnahme eingeladen. Zusätzlich wurde der Studienlink über soziale Netzwerke (z. B. LinkedIn, private Kontakte) weitergegeben.

Für die Teilnahme erhielten die Teilnehmenden, die über die SONA-Plattform teilgenommen haben oder ihre SONA-ID angaben, 0.25 Versuchspersonenstunden. Außerdem hatten alle Teilnehmenden die Möglichkeit einen 20€ Amazon Gutschein zu gewinnen. Teilnehmende konnten freiwillig an der Verlosung teilnehmen, wenn sie nach Abschluss der Datenerhebung ihre E-Mail-Adresse angaben.

Da die Teilnehmenden die Möglichkeit hatten, ihre E-Mail-Adressen sowie ihre SONA-ID anzugeben, war die Studie formal nicht anonym. Diese personenbezogenen Angaben wurden jedoch beim Export der Antwortdaten nicht berücksichtigt. Stattdessen wurden E-Mail-Adressen und SONA-IDs separat exportiert und ausschließlich für die Durchführung der Gewinnspielverlosung und Vergabe der Ver-

suchspersonenstunden verwendet. Ein Rückschluss von den Antwortdaten auf einzelne E-Mail-Adressen war somit nicht möglich, sodass die Anonymität der erhobenen Daten gewährleistet wurde.

Als Einschlusskriterien für die Teilnahme galten ein Mindestalter von 18 Jahren sowie ausreichende Deutschkenntnisse. Verpflichtend zu beantworten waren alle inhaltlichen Fragen des Fragebogens; freiwillig waren lediglich die Angabe der E-Mail-Adresse sowie der SONA-ID. Datensätze von Personen, die nicht alle verpflichtenden Fragen beantwortet hatten, wurden von der Analyse ausgeschlossen. Dies betraf  $N = 39$  Datensätze. Insgesamt begannen 168 Personen die Studie.

Die Bestimmung der Stichprobengröße erfolgte mittels a-priori Power-Analyse in G\*Power 3.1.9.7. (Faul et al., 2007). Die erstrebte Stichprobe lag bei mindestens  $N = 128$  Personen. Das entspricht der Detektion eines mittleren Effekts ( $f = 0.25$ ) bei einer Teststärke ( $1 - \beta$ ) von 0.80 und einem Signifikanzniveau von  $\alpha = .05$ . Es ist somit zu berücksichtigen, dass mit dieser Stichprobengröße primär Effekte mittlerer Größe zuverlässig identifiziert werden konnten (Cohen, 2013).

### 3.3 Versuchsmaterialien

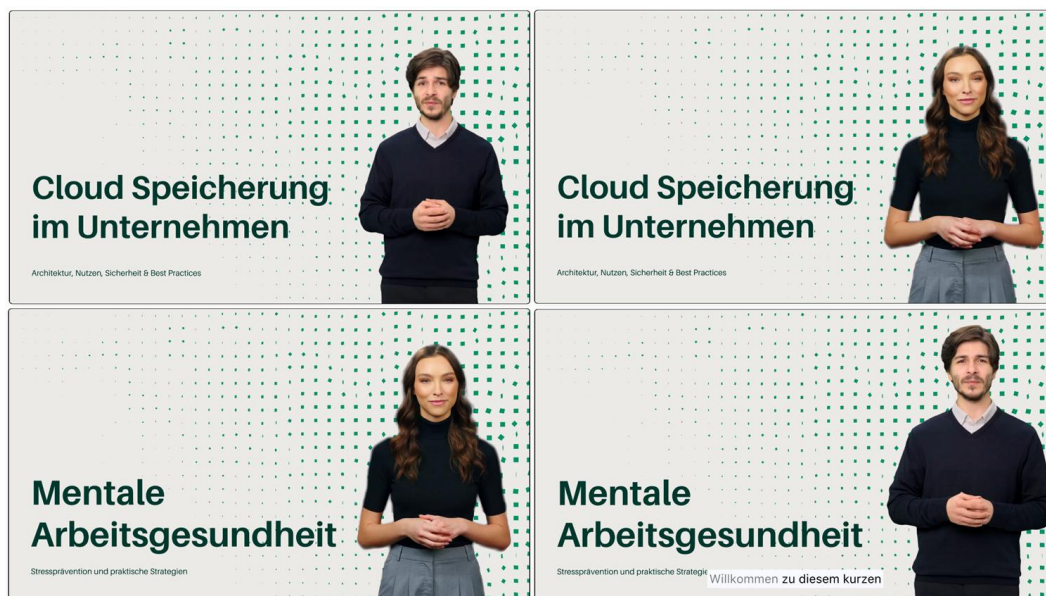
Als Stimulusmaterial dienten die Lehrvideos, die in den experimentellen Gruppen systematisch variiert wurden. Die Erstellung der Videos erfolgte mit der Software HeyGen, in der die Inhalte einer Präsentation von einem KI-generierten Avatar vermittelt wurden. Die inhaltliche Ausarbeitung erfolgte auf Grundlage eigener Recherche. Ziel war es, Lernmaterialien zu entwickeln, die realistischen Schulungssituationen im Unternehmenskontext möglichst nahekommen. Für die Fachdomäne Informatik wurde das Thema Cloud-Speicherung gewählt. Dabei wurden Inhalte aus den Webseiten *Wie Cloud Storage Ihre Infrastruktur voranbringt* (2025) und *14 Best Practices zur Sicherung Ihrer Cloud-Umgebung* (o. J.) berücksichtigt. Für das Lehrvideo zur Arbeitsgesundheit wurde das Thema Stressprävention am Arbeitsplatz gewählt und unter Einbezug von Breucker, Sochert und Mißler (2015), *So erkennen Sie Stresssymptome* (2022), Löhr B. (2022) und Schmidt, J. (2024) konzipiert. Zur Unterstützung bei der Strukturierung der Präsentationsfolien wurde das

Sprachmodell ChatGPT (OpenAI, 2025) eingesetzt. Die generierten Textvorschläge wurden anschließend inhaltlich geprüft und überarbeitet. Der Sprechtext für die KI-generierten Lehrpersonen wurde auf Basis der erarbeiteten Inhalte mithilfe von ChatGPT (OpenAI, 2025) formuliert und anschließend eigenständig angepasst.

Die visuelle Struktur der Präsentationen wurde für beide Fachdomänen identisch gehalten. Darüber hinaus folgten alle Lehrvideos demselben inhaltlichen Aufbau, bestehend aus einer Begrüßung, einer kurzen Einführung in die Inhalte, einem Hauptteil mit Wissensvermittlung sowie einem abschließenden Fazit. Die Lehrvideos waren hinsichtlich ihrer Länge vergleichbar und dauerten jeweils etwa fünf Minuten. Variiert wurde ausschließlich das Geschlecht der KI-generierten Lehrperson. Hierfür wurden zwei von HeyGen bereitgestellte Avatare verwendet. Die verwendeten Lehrvideos sind im Anhang unter Abschnitt 1 („HeyGen-Lehrvideos“) abzurufen.

### Abbildung 1

*Titelbilder der Lehrvideos in den vier experimentellen Bedingungen. Dargestellt sind die kongruenten und inkongruenten Kombinationen der Variablen Fachbereich (Informatik vs. Arbeitsgesundheit) und Geschlecht der Lehrperson (männlich vs. weiblich).*



Zur Erfassung des Lernerfolgs wurde ein eigens entwickelter Wissenstest eingesetzt, der sich inhaltlich auf die im jeweiligen Lehrvideo vermittelten Informationen

bezog. Die Testfragen wurden selbst erstellt und anschließend mithilfe von Chat-GPT (OpenAI, 2025) hinsichtlich ihrer sprachlichen Formulierung überprüft. Der Test bestand aus insgesamt 15 dichotomen Aussagen, die von den Teilnehmenden mit „Wahr“ oder „Falsch“ beantwortet werden konnten. Für jede korrekt beantwortete Aussage wurde ein Punkt vergeben, sodass maximal 15 Punkte erreicht werden konnten. Der Gesamtpunktwert diente als Maß für den Lernerfolg. Die Fragen des Wissenstest sind unter Abschnitt 2.1 („Items zur Messung des Lernerfolgs im Anhang“) nachzulesen.

Die wahrgenommene didaktische Kompetenz der KI-generierten Lehrperson wurde mithilfe einer fünfstufigen Likert-Skala erfasst, die von 1 („Ich stimme gar nicht zu“) bis 5 („Ich stimme voll zu“) reichte. Die Skala umfasste fünf Items, die inhaltlich an die Teacher Effectiveness Scale in Higher Education (TESHE; Calaguas, 2012) angelehnt waren. Die Anlehnung an die TESHE erschien sinnvoll, da die Lehrvideos im Kontext der Erwachsenenbildung eingesetzt wurden und die zugrunde liegenden Items validiert sind. Es wurden ausschließlich jene Items ausgewählt, die auf eine digitale Lehrperson sinnvoll übertragbar erschienen. Die interne Konsistenz der Skala ist in Tabelle 2 zu sehen und ist im Allgemeinen ausreichend bis gut zu bewerten (Danner, 2015). Die verwendeten Items sind im Anhang unter Punkt 2.2 („Items zur Wahrnehmung der didaktischen Kompetenz“) aufgeführt.

**Tabelle 2**

*Interne Konsistenz der Variable Wahrnehmung didaktischer Kompetenz nach experimenteller Gruppe*

Wahrgenommene Didaktische Kompetenz	Cronbach's $\alpha$
Männlich, Informatik (kongruent)	0.84
Weiblich, Informatik (inkongruent)	0.69
Weiblich, Arbeitsgesundheit (kongruent)	0.82
Männlich, Arbeitsgesundheit (inkongruent)	0.75

Die Manipulationskontrolle wurde mithilfe von insgesamt acht selbstkonzipierten Items erhoben, die auf einer fünfstufigen Likert-Skala von 1 („Ich stimme gar nicht zu“) bis 5 („Ich stimme voll zu“) beantwortet wurden. Jeweils vier Items erfassten die wahrgenommene Geschlechterstereotypisierung der beiden Fachdomänen Informatik und Arbeitsgesundheit (z. B. „Informatik gilt allgemein als „typisch männlicher“ Bereich.“ oder „Menschen, die im Bereich Arbeitsgesundheit arbeiten, sind

typischerweise Frauen.“). Für die Wahrnehmung von Arbeitsgesundheit als weiblich konnotierte Fachdomäne ergab sich ein Cronbach's  $\alpha = 0.81$  und für die Wahrnehmung von Informatik als männliche Fachdomäne ein Cronbach's  $\alpha = 0.80$ . Eine vollständige Übersicht der Items findet sich im Anhang unter Punkt 2.3 („Items zur Messung der Manipulationskontrolle“).

Das Vorwissen der Teilnehmenden wurde mittels dichotomer Skala gemessen. Dabei gaben die Teilnehmenden an, ob sie in den Themenbereichen Cloud-Speicherung und Stressprävention am Arbeitsplatz einen Lernbedarf aufweisen oder nicht. Unabhängig von diesen Angaben erfolgte die zufällige Zuteilung zu den experimentellen Gruppen. Die verwendeten Items sind im Anhang unter Punkt 2.4 („Items zur Erhebung des Vorwissens aufgeführt“).

Explorativ wurde die wahrgenommene Realitätsnähe des KI-Avatars erhoben. Die Messung erfolgte mithilfe einer fünfstufigen Likert-Skala (1 = „Ich stimme gar nicht zu“ bis 5 = „Ich stimme voll zu“). Die Variable wurde anhand von fünf Items erfasst, die sich an den Arbeiten von MacDorman und Chattopadhyay (2016) anlehnen. Da die ursprünglichen Items als semantische Differentialskala konzipiert waren, wurden sie für den vorliegenden Fragebogen in Likert-Items umformuliert. Die Skala wies eine sehr gute interne Konsistenz auf (Cronbach's  $\alpha = 0.92$ ). Eine vollständige Übersicht der Items findet sich im Anhang unter Punkt 2.5 („Items zum wahrgenommenen Realismus“)

### **3.4 Versuchsablauf**

Die Studie wurde online im Rahmen der ethischen Richtlinien der APA (2017) durchgeführt. Die Teilnehmenden erhielten eine kurze Einführung über die Studie, in der ausschließlich berichtet wurde, dass sie ein Lehrvideo sehen werden und anschließend Fragen zu dem Video beantworten werden. Außerdem wurden sie über die freiwillige Teilnahme sowie ihr Recht auf einen jederzeitigen Abbruch informiert und gaben ihr Einverständnis zur Datenverarbeitung. Zu Beginn der Untersuchung wurden demografische Angaben, darunter Alter und Geschlecht, erhoben. Im Anschluss wurde das selbstberichtete Vorwissen in Bezug auf die jeweilige Fachdomäne als Kontrollvariable erfasst, um Teilnehmende mit hoher fachlicher Spezialisierung identifizieren zu können. Anschließend erfolgte eine randomisierte Zu-

weisung der Teilnehmenden zu einer der vier experimentellen Bedingungen. Daraufhin sahen die Teilnehmenden ein Lehrvideo und bearbeiteten direkt im Anschluss den Wissenstest zur Erfassung des Lernerfolgs. Anschließend wurde die wahrgenommene didaktische Kompetenz des KI-Avatars erfasst. Zur Validierung der experimentellen Stimuli wurde außerdem die subjektive Stereotypisierung der jeweiligen Fachdomäne erhoben. Im Anschluss an die Hauptmessungen wurde eine explorative Erhebung durchgeführt, in der die wahrgenommene Realitätsnähe und Authentizität des KI-Avatars erfasst wurden. Abschließend erhielten die Teilnehmenden ein Debriefing. Hierbei wurde erläutert, dass die Studie den Einfluss des Geschlechts der Lehrperson untersuchte und auf einem Design mit vier experimentellen Gruppen basierte. Der Zeitaufwand pro Person lag bei circa 15 Minuten.

## 4 Ergebnisse

Im Ergebnisteil werden zunächst die Ergebnisse der Manipulationskontrolle präsentiert (Abschnitt 4.1). In Abschnitt 4.2 wird die Auswertung der Hypothesen vorgenommen. In Abschnitt 4.3 werden relevante explorative Analysen zu den erhobenen Daten vorgestellt.

### 4.1 Manipulationskontrolle

Zur Überprüfung der Manipulation wurde erfasst, inwieweit die Teilnehmenden die Fachdomäne Informatik als männlich und die Fachdomäne Arbeitsgesundheit als weiblich wahrnahmen. Die Bewertung erfolgte auf einer fünfstufigen Likert-Skala mit einem Skalenmittelpunkt von 3. Die Informatik-Items wiesen einen Mittelwert von  $M = 3.97$  ( $SD = 0.73$ ) auf. Für die Items zur Arbeitsgesundheit ergab sich ein Mittelwert von  $M = 3.10$  ( $SD = 0.86$ ). Zur statistischen Überprüfung wurde jeweils ein t-Test gegen den Skalenmittelpunkt von 3 durchgeführt. Für die Informatik-Items zeigte sich ein signifikanter Unterschied zum Skalenmittelpunkt,  $t(128) = 15.20$ ,  $p < .001$ ,  $d = 1.34$ . Für die Items zur Arbeitsgesundheit ergab sich kein signifikanter Unterschied zum Skalenmittelpunkt,  $t(128) = 1.33$ ,  $p = .187$ ,  $d = 0.12$ . Zwischen dem Informatik-Items und dem Arbeitsgesundheit-Items zeigte sich zudem eine signifikant positive Korrelation,  $r(127) = .386$ ,  $p < .001$ .

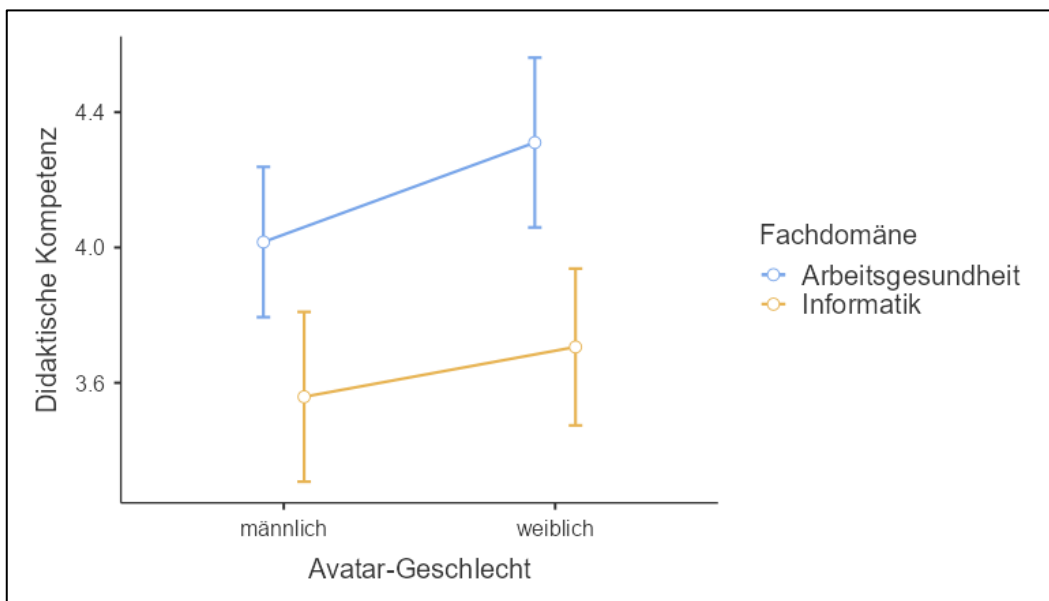
Zusätzlich wurde erhoben, ob die Teilnehmenden über Vorwissen zu den Lehrvideoinhalten verfügten. Insgesamt gaben 27 Teilnehmende an, keinen Lernbedarf in den Themenbereichen Cloud-Speicherung oder Stressprävention am Arbeitsplatz zu haben. Diese Teilnehmenden wurden dennoch zufällig derselben Versuchsbedingung zugewiesen. Der Ausschluss dieser 27 Datensätze führte zu keinen substantiellen Veränderungen der Ergebnisse der Hypothesentests. Entsprechend wurden die weiteren Analysen auf Basis des vollständigen Datensatzes ( $N = 129$ ) durchgeführt.

## 4.2 Befunde zu den Hypothesen

In Hypothese 1 wurde angenommen, dass rollenkongruente Bedingungen zwischen Geschlecht des Avatars und Fachdomäne zu einer höheren Bewertung der didaktischen Kompetenz führen als rolleninkongruente Bedingungen. Abbildung 2 zeigt die Verteilung der Mittelwerte der didaktischen Kompetenz für die einzelnen experimentellen Gruppen. Die höchsten Mittelwerte fanden sich in der Gruppe weibliche, Arbeitsgesundheit (kongruent) ( $M = 4.31$ ,  $SD = 0.63$ ), während die niedrigsten Werte in der Gruppe männlich, Informatik (kongruent) berichtet wurden ( $M = 3.56$ ,  $SD = 0.89$ ). Insgesamt fiel die Bewertung des weiblichen Avatars im Mittel etwas höher aus ( $M = 3.98$ ,  $SD = 0.69$ ) als die des männlichen Avatars ( $M = 3.89$ ,  $SD = 0.77$ ). Zudem wurde der Avatar im Kontext des Lehrvideos zur Arbeitsgesundheit im Durchschnitt positiver bewertet ( $M = 4.15$ ,  $SD = 0.62$ ) als im Informatik-Lehrvideo ( $M = 3.64$ ,  $SD = 0.75$ ).

### Abbildung 2

Mittelwerte der didaktischen Kompetenz nach Avatar-Geschlecht und Fachdomäne (Fehlerbalken = 95%-Konfidenzintervalle)



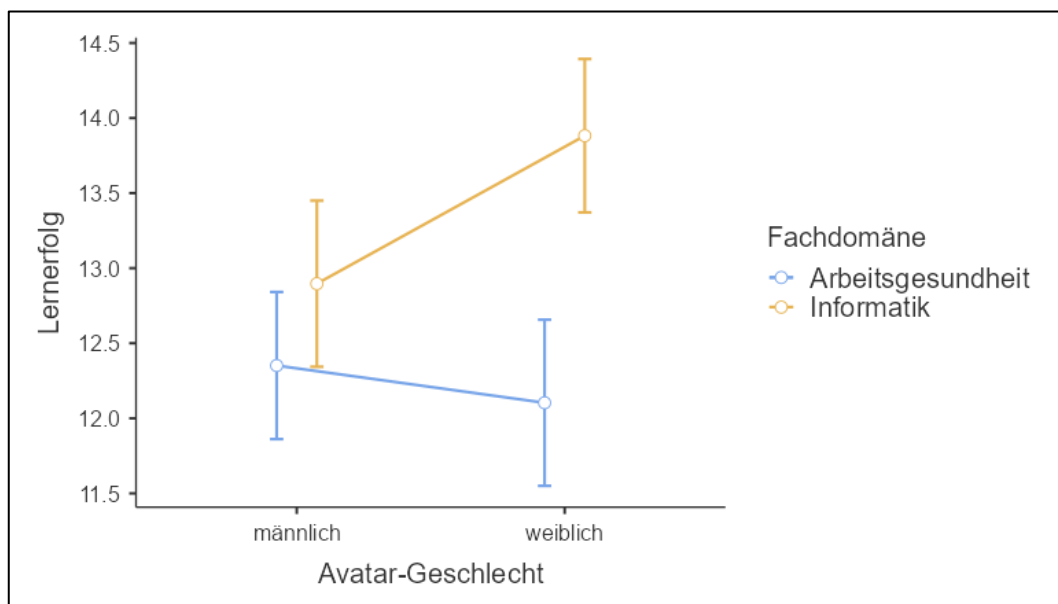
Die Varianzanalyse zeigte einen höchstsignifikanten Haupteffekt der Fachdomäne,  $F(1,125) = 19.29$ ,  $p < .001$ ,  $\eta^2_p = .134$ . Der Avatar im Arbeitsgesundheit-Lehrvideo wurde demnach im Mittel signifikant didaktisch kompetenter bewertet im Vergleich zum Informatik-Lehrvideo. Der Haupteffekt des Avatar-Geschlechts zeigte keine statistische Signifikanz,  $F(1,125) = 3.33$ ,  $p = .070$ ,  $\eta^2_p = .015$ . Die Interaktion zwischen Avatar-Geschlecht und Fachdomäne war ebenfalls nicht signifikant,

$F(1,125) = 0.37, p = .545, \eta^2_p = .003$ . Damit konnte Hypothese 1 nicht bestätigt werden. Es zeigte sich kein signifikanter Unterschied zwischen rollenkongruenten und rolleninkongruenten Bedingungen hinsichtlich der Bewertung der didaktischen Kompetenz des KI-Avatars.

Als zweite Hypothese wurde formuliert, dass die kongruente Bedingung einen positiven Effekt auf den Lernerfolg hat, im Vergleich zur inkongruenten Bedingung. Abbildung 3 veranschaulicht die Mittelwerte der im Wissenstest erzielten Punkte in den einzelnen experimentellen Gruppen. Die höchsten durchschnittlichen Punktzahlen erzielte die Gruppe weiblich, Informatik (inkongruent) ( $M = 13.9, SD = 1.20$ ). Die niedrigsten Werte zeigten sich in der Gruppe weiblich, Arbeitsgesundheit (kongruent) ( $M = 12.1, SD = 1.3$ ). Die Teilnehmenden erzielten bessere Ergebnisse im Informatik Lehrvideo ( $M = 13.4, SD = 1.7$ ) im Vergleich zum Arbeitsgesundheit Lehrvideo ( $M = 12.2; SD = 1.3$ ). Auch erreichten die Teilnehmenden bei einem weiblichen Avatar im Durchschnitt leicht höhere Punktzahlen ( $M = 13.1, SD = 1.5$ ) im Vergleich zum männlichen Avatar ( $M = 12.6, SD = 1.7$ ).

### Abbildung 3

Lernerfolg (Erreichte Punktzahlen im Wissenstest) nach Avatar-Geschlecht und Fachdomäne (Fehlerbalken = 95%-Konfidenzintervalle)



Es zeigte sich ebenfalls ein höchstsignifikanter Haupteffekt der Fachdomäne,  $F(1,125) = 19.00, p < .001, \eta^2_{partial} = .132$ . Beim Informatik Lehrvideo erzielten die Teilnehmenden im Mittel signifikant höhere Punktzahlen im Vergleich zum Ar-

beitsgesundheit Lehrvideo. Das Geschlecht des Avatars erzielte keinen signifikanten Haupteffekt  $F(1,125) = 1.92, p = .169, \eta^2_{\text{partial}} = .015$ . Die Interaktion von Avatar-Geschlecht und Fachdomäne war signifikant,  $F(1,125) = 5.35, p = .022, \eta^2_{\text{partial}} = .041$ . Die Hypothesen H2a und H2b konnten jedoch nicht bestätigt werden. Im Gegensatz zur Erwartung fielen die Mittelwerte in den inkongruenten Bedingungen höher aus als in den kongruenten Bedingungen.

### 4.3 Weiterführende Analysen

Aufgrund der hohen weiblichen Teilnehmeranzahl von 77,5% wurde in einer 2 x 2 x 2 ANOVA der Einfluss des Teilnehmergeschlechts auf die abhängigen Variablen der Hypothese 1 didaktische Kompetenz und Hypothese 2 Lernerfolg geprüft. Das Geschlecht hatte keinen Einfluss auf die Bewertung der didaktischen Kompetenz der KI-Avatare. Auch gab es keine signifikante Zweifach- oder Dreifachinteraktion mit Avatar-Geschlecht und Fachdomäne. Lediglich die Effektstärke des Haupteffekts der Fachdomäne veränderte sich ( $F(1, 121) = 9.29, p = .003, \eta^2_{\text{partial}} = .071$ ), wobei der Effekt weiterhin hochsignifikant blieb. Die vollständigen Ergebnisse der Varianzanalysen sind in Tabelle 3 und Tabelle 4 dargestellt.

**Tabelle 3**

*Dreifaktorielle Varianzanalyse für die wahrgenommene didaktische Kompetenz unter Einbezug des Teilnehmergeschlechts*

	Quadratsumme	df	Mittlere quad. Abw.	F	p	$\eta^2_p$
Geschlecht	0.1352	1	0.1352	0.286	.594	.002
Avatar-Geschlecht	0.8892	1	0.8892	1.884	.172	.015
Fachdomäne	4.3836	1	4.3836	9.287	.003	.071
Geschlecht × Avatar-Geschlecht	0.0513	1	0.0513	0.109	.742	.001
Geschlecht × Fachdomäne	0.5721	1	0.5721	1.212	.273	.010
Avatar-Geschlecht × Fachdomäne	0.3214	1	0.3214	0.681	.411	.006
Geschlecht × Avatar-Geschlecht × Fachdomäne	0.2678	1	0.2678	0.567	.453	.005
Residuen	57.1133	121	0.4720			

*Anmerkung.*  $\eta^2_p$  = partielles Eta-Quadrat. Signifikanzniveau:  $p < .05$

**Tabelle 4**

*Dreifaktorielle Varianzanalyse für den Lernerfolg unter Einbezug des Teilnehmergeschlechts*

	Quadratsumme	df	Mittlere quad. Abw.	F	p	$\eta^2_p$
Geschlecht	3.41	1	3.41	1.523	.220	.012

	<i>Quadratsumme</i>	<i>df</i>	<i>Mittlere quad. Abw.</i>	<i>F</i>	<i>p</i>	$\eta^2p$
Avatar-Geschlecht	7.22	1	7.22	3.225	.075	.026
Fachdomäne	33.07	1	33.07	14.773	< .001	.109
Geschlecht $\times$ Avatar-Geschlecht	1.27	1	1.27	0.566	.453	.005
Geschlecht $\times$ Fachdomäne	1.27	1	1.27	0.568	.452	.005
Avatar-Geschlecht $\times$ Fachdomäne	3.22	1	3.22	1.437	.233	.012
Geschlecht $\times$ Avatar-Geschlecht $\times$ Fachdomäne	5.70	1	5.70	2.546	.113	.021
Residuen	270.85	121	2.24			

*Anmerkung.*  $\eta^2p$  = partielles Eta-Quadrat. Signifikanzniveau:  $p < .05$

Auch wurde der Einfluss des Alters der Teilnehmenden mithilfe einer ANCOVA auf die Wahrnehmung der didaktischen Kompetenz (Analog zu H1) und des Lernerfolgs (Analog zu H2) geprüft. Das Alter wurde als Kovariate in beide Modelle aufgenommen. Die Kovariate hatte keinen signifikanten Einfluss auf die Bewertung der didaktischen Kompetenz,  $F(1, 124) = 0.53, p = .469, \eta^2_{\text{partial}} = .004$  und auch nicht auf den Lernerfolg,  $F(1, 124) = 0.72, p = .398, \eta^2_{\text{partial}} = .006$ . Die Haupteffekte und Interaktion der unabhängigen Variablen Fachdomäne und Avatar-Geschlecht blieben bei beiden Analysen unverändert. Alle Ergebnisse sind in Tabelle 5 und Tabelle 6 dargestellt.

### **Tabelle 5**

*Kovarianzanalyse (ANCOVA) zum Einfluss von Fachdomäne und Avatar-Geschlecht auf die wahrgenommene didaktische Kompetenz unter Kontrolle des Alters*

	<i>Quadratsumme</i>	<i>df</i>	<i>Mittlere quad. Abw.</i>	<i>F</i>	<i>p</i>	$\eta^2p$
Alter	0.247	1	0.247	0.529	.469	.004
Avatar-Geschlecht	1.520	1	1.520	3.248	.074	.026
Fachdomäne	9.039	1	9.039	19.313	< .001	.135
Avatar-Geschlecht $\times$ Fachdomäne	0.163	1	0.163	0.349	.556	.003
Residuen	58.039	124	0.468			

*Anmerkung.*  $\eta^2p$  = partielles Eta-Quadrat. Signifikanzniveau:  $p < .05$

### **Tabelle 6**

*Kovarianzanalyse (ANCOVA) zum Einfluss von Fachdomäne und Avatar-Geschlecht auf den Lernerfolg unter Kontrolle des Alters*

	<i>Quadratsumme</i>	<i>df</i>	<i>Mittlere quad. Abw.</i>	<i>F</i>	<i>p</i>	$\eta^2p$
Alter	1.63	1	1.63	0.718	0.398	0.006
Avatar-Geschlecht	4.48	1	4.48	1.972	0.163	0.016
Fachdomäne	43.33	1	43.33	19.071	< .001	0.133

	<i>Quadratsumme</i>	<i>df</i>	<i>Mittlere quad. Abw.</i>	<i>F</i>	<i>p</i>	$\eta^2p$
Avatar-Geschlecht $\times$ Fachdomäne	11.94	1	11.94	5.256	0.024	0.041
Residuen	281.71	124	2.27			

*Anmerkung.*  $\eta^2p$  = partielles Eta-Quadrat. Signifikanzniveau:  $p < .05$

Explorativ wurde im Rahmen der Untersuchung auch der wahrgenommene Realismus des Avatars erfasst. Tabelle 7 zeigt die Ergebnisse der Varianzanalyse mit den experimentellen Gruppen und dem wahrgenommenen Realismus als abhängige Variable. Es ergaben sich weder für die Haupteffekte noch für die Interaktion signifikante Unterschiede. Für das Avatar-Geschlecht zeigte sich ein knapp nicht signifikanter Effekt ( $F(1, 125) = 3.15, p = .070, \eta^2_{\text{partial}} = .025$ ), wobei der männliche Avatar tendenziell als realistischer wahrgenommen wurde ( $M = 2.84, SD = 1.07$ ) als der weibliche Avatar ( $M = 2.52, SD = 0.91$ ). Insgesamt fiel der wahrgenommene Realismus der Avatare eher niedrig aus ( $M = 2.68, SD = 1.00$ ).

### **Tabelle 7**

*Varianzanalyse für den wahrgenommenen Realismus nach experimenteller Gruppe*

	<i>Quadratsumme</i>	<i>df</i>	<i>Mittlere quad. Abw.</i>	<i>F</i>	<i>p</i>	$\eta^2p$
Avatar-Geschlecht	3.16	1	3.16	3.15	.079	.025
Fachdomäne	0.14	1	0.14	0.14	.709	.001
Avatar-Geschlecht $\times$ Fachdomäne	0.03	1	0.03	0.03	.873	.000
Residuen	225.50	125	1.00			

*Anmerkung.*  $\eta^2p$  = partielles Eta-Quadrat. Signifikanzniveau:  $p < .05$

Es konnte ein signifikant positiver Zusammenhang zwischen Realismus und didaktischer Kompetenz festgestellt werden  $\beta = .28, t(127) = 4.77, p < .001$ . Dies war jedoch überraschend, da die vorherige Analyse zeigte, dass der männliche Avatar deskriptiv realistischer wahrgenommen wurde, aber die wahrgenommene didaktische Kompetenz für den weiblichen Avatar deskriptiv höher ausfiel, siehe Abschnitt 4.2 („Befunde zu den Hypothesen“). In Tabelle 8 und 9 wurde nun in zwei Regressionsmodellen geprüft wie sich die Variablen Avatar-Geschlecht und wahrgenommener Realismus des Avatars auf die didaktische Kompetenz auswirkten. Der Vergleich macht sichtbar, dass erst nach Aufnahme des Realismus im Regressionsmodell das Avatar Geschlecht einen signifikanten Einfluss auf die didaktische Kompetenz hatte ( $\beta = .27, t(126) = 2.25, p = .026$ ), wobei der weibliche Avatar im

Mittel als didaktisch kompetenter eingeschätzt wurde als der männliche Avatar. Zudem erhöhte sich die erklärte Varianz des Modells von  $R^2 = .014$  auf  $R^2 = .185$  ( $\Delta R^2 = 0.171$ ). Gleichzeitig nahm die Stärke des Zusammenhangs zwischen Realismus und didaktischer Kompetenz im Vergleich zum einfachen Regressionsmodell leicht zu. Diese Befunde deuten auf das Vorliegen eines Suppressionseffekts hin. Die inhaltliche Einordnung erfolgt im Diskussionsteil.

**Tabelle 8**

*Lineare Regressionsanalyse zur Vorhersage der wahrgenommenen didaktischen Kompetenz durch das Avatar-Geschlecht*

Prädiktor	<i>B</i>	<i>SE</i>	<i>t</i>	<i>p</i>
Konstante	3.815	0.090	42.54	< .001
Avatar-Geschlecht (weiblich)	0.169	0.128	1.32	.190
$R^2$	.014			

*Anmerkung.* Signifikanzniveau:  $p < .05$

**Tabelle 9**

*Multiple Regressionsanalyse zur Vorhersage der didaktischen Kompetenz durch das Avatar-Geschlecht unter Kontrolle des wahrgenommenen Realismus*

Prädiktor	<i>B</i>	<i>SE</i>	<i>t</i>	<i>p</i>
Konstante	2.948	0.1872	15.75	< .001
Avatar-Geschlecht (weiblich)	0.267	0.1187	2.25	.026
Realismus	0.305	0.0593	5.15	< .001
$R^2$	.185			

*Anmerkung.* Signifikanzniveau:  $p < .05$

## 5 Diskussion

Ziel der vorliegenden Studie war es, zu untersuchen, ob das Geschlecht einer KI-generierten Lehrperson die wahrgenommene didaktische Kompetenz des Avatars sowie den Lernerfolg der Teilnehmenden beeinflusst. Auf Grundlage der Role Congruity Theory wurde angenommen, dass Avatare in rollenkongruenten Bedingungen positiver bewertet werden und zu höheren Lernleistungen führen als in rolleninkongruenten Bedingungen. Diese Annahmen konnten jedoch nicht bestätigt werden. Für die Kompetenzbewertung zeigte sich kein signifikanter Interaktionseffekt zwischen Avatar-Geschlecht und Fachdomäne. Auch der Haupteffekt des Geschlechts war in der Varianzanalyse zunächst nicht signifikant. Erst nach Kontrolle des wahrgenommenen Realismus als Kovariate zeigte sich, dass der weibliche Avatar im Mittel als signifikant kompetenter wahrgenommen wurde als der männliche. Der wahrgenommene Realismus könnte in diesem Zusammenhang als Suppressionsvariable wirken, indem er irrelevante Varianz in der Prädiktorvariable Avatar-Geschlecht kontrolliert und dadurch den Zusammenhang zwischen Avatar-Geschlecht und wahrgenommener didaktischer Kompetenz deutlicher hervortreten lässt. Dafür spricht, dass sich der Regressionskoeffizient für das Avatar-Geschlecht unter Einbezug des Realismus erhöhte und zugleich das Bestimmtheitsmaß ( $R^2$ ) stark anstieg.

Für den Lernerfolg zeigte sich ebenfalls kein hypothesenkonformes Muster. Entgegen den Erwartungen erzielten die Teilnehmenden insgesamt höhere Lernleistungen in rolleninkongruenten Bedingungen im Vergleich zu rollenkongruenten Bedingungen. Deskriptiv zeigte sich zudem ein leichter Vorteil des weiblichen Avatars im Lernerfolg, wenngleich dieser Unterschied statistisch nicht signifikant war. Insgesamt widersprechen die Befunde somit den Annahmen der Role Congruity Theory im vorliegenden Kontext. Die Fachdomäne hatte in Kombination mit dem Geschlecht des Avatars weder einen Einfluss auf die Kompetenzbewertung noch führte sie in rollenkongruenten Bedingungen zu einem höheren Lernerfolg.

Ein zentraler Erklärungsansatz ergibt sich aus der Manipulationskontrolle. Die Fachdomäne Arbeitsgesundheit wurde von den Teilnehmenden nicht eindeutig als weiblich konnotiert wahrgenommen, im Vergleich zur Fachdomäne Informatik.

Damit ist fraglich, ob die intendierte Rollenkongruenz tatsächlich aktiviert wurde. Während frühere Forschung zeigt, dass die Geschlechterverteilung in Berufen stereotype Zuschreibungen beeinflusst, deuten Befunde aus dem Lehrkontext darauf hin, dass insbesondere die interne Geschlechterverteilung innerhalb einer Institution oder Domäne entscheidend sein kann (Aragón et al., 2023). Ein isoliertes Lehrvideo ohne expliziten Hinweis auf die Geschlechterverteilung scheint möglicherweise nicht ausreichend, um eine klare geschlechtliche Konnotation zu aktivieren. Darüber hinaus ist zu berücksichtigen, dass der Lehrberuf insbesondere im Bereich der Primar- und Sekundarbildung häufig als eher weiblich konnotiert wahrgenommen wird (Jugovic et al., 2022; Kammermeier et al., 2025). Diese übergeordnete Rollenzuschreibung könnte potenzielle Inkongruenzeffekte abgeschwächt oder überlagert haben, selbst wenn sich das Lehrvideo an Erwachsene richtete.

Ein weiterer Erklärungsansatz betrifft die Avatare selbst. Der wahrgenommene Realismus wurde insgesamt eher niedrig eingeschätzt. Erste Befunde legen nahe, dass hochrealistische Avatare leichter verarbeitet werden (de Borst & de Gelder, 2015) und eher als glaubwürdige soziale Akteure wahrgenommen werden (Tao et al., 2025; Baake et al., 2025). Insbesondere in professionellen Kontexten scheint der visuelle Realismus maßgeblich zur Zuschreibung von Kompetenz und Vertrauenswürdigkeit beizutragen (Tao et al., 2025). Wenn Avatare jedoch als wenig realistisch wahrgenommen werden, werden sie möglicherweise nicht ausreichend als „echte“ soziale Rolleninhaber kategorisiert. Geschlechtsbezogene Rollenerwartungen könnten dadurch weniger stark aktiviert werden, sodass Effekte im Sinne der Role Congruity Theory abgeschwächt oder verhindert werden. Gleichzeitig zeigen aktuelle Studien, dass weiblich verkörperte Avatare in digitalen Kontexten teilweise präferiert werden, da sie als wärmer und menschlicher wahrgenommen werden (Craiut & Iancu, 2022). Diese Präferenz könnte sowohl die positivere Kompetenzbewertung des weiblichen Avatars als auch die deskriptiv höheren Lernleistungen erklären.

Die Ergebnisse sind zudem vor dem Hintergrund methodischer Limitationen zu interpretieren. Es wurde kein Pretest zur Überprüfung der wahrgenommenen Schwierigkeit der Lehrvideos durchgeführt. Unterschiede in der Fachdomäne könnten daher teilweise auf Unterschiede in Verständlichkeit oder Komplexität der Inhalte zu-

rückzuführen sein. Zudem war die Stichprobe überwiegend weiblich und vergleichsweise jung. Explorative Analysen zeigten zwar keinen signifikanten Einfluss von Alter oder Geschlecht der Teilnehmenden auf die abhängigen Variablen, dennoch könnte eine diversere Stichprobe die Generalisierbarkeit zukünftiger Befunde erhöhen.

Für zukünftige Studien erscheint es zentral, die Manipulation der Fachdomäne stärker abzusichern. Neben der Erhebung der wahrgenommenen Konnotation könnten Teilnehmende explizit zu stereotypen Zuschreibungen, etwa kommunalen und agentischen Eigenschaften, befragt werden. Ebenso könnte experimentell variiert werden, ob Informationen zur Geschlechterverteilung einer Domäne bereitgestellt werden, um Rollenkongruenz gezielter zu aktivieren. Darüber hinaus bleibt grundsätzlich offen, in welchem Ausmaß Theorien wie die Role Congruity Theory auf KI-generierte Lehrpersonen übertragbar sind. Zwar zeigen Studien, dass soziale Kategorisierungsprozesse auch bei technologischen Entitäten auftreten können, die empirische Befundlage im Kontext moderner KI-Avatare ist jedoch bislang begrenzt.

Aus praktischer Perspektive verdeutlichen die Befunde, dass das Geschlecht KI-generierter Lehrpersonen nicht als neutraler Designparameter betrachtet werden sollte. Vielmehr zeigen sich Effekte auf die Kompetenzzuschreibung sowie auf den Lernerfolg, die jedoch nicht im Sinne klassischer Annahmen der Role Congruity Theory ausfallen. Die insgesamt positivere Bewertung des weiblichen Avatars weist darauf hin, dass in digitalen Lehrkontexten möglicherweise andere soziale Wahrnehmungsprozesse wirksam sind als bei realen Lehrpersonen. Für die Konzeption KI-gestützter Lehrformate implizieren die Ergebnisse, dass sowohl das Avatar-Geschlecht als auch dessen visuelle Ausgestaltung als relevante Einflussfaktoren systematisch berücksichtigt werden sollten. Angesichts der dynamischen technologischen Entwicklungen im Bereich KI-gestützter Lehrformate besteht erheblicher Forschungsbedarf, um soziale Wahrnehmungsprozesse im Umgang mit digitalen Lehrpersonen systematisch zu untersuchen.

## Literatur

- 14 Best Practices zur Sicherung Ihrer Cloud-Umgebung.* (o. J.). ManageEngine Log360. Abgerufen 8. März 2026, von <https://www.manageengine.com/de/log-management/cyber-security/secure-cloud-14-best-practices.html>
- 14 Best Practices zur Sicherung Ihrer Cloud-Umgebung.* (o. J.). ManageEngine Log360. Abgerufen 8. März 2026, von <https://www.manageengine.com/de/log-management/cyber-security/secure-cloud-14-best-practices.html>
- Aragón, O. R., Pietri, E. S., & Powell, B. A. (2023). Gender bias in teaching evaluations: The causal role of department gender composition. *Proceedings of the National Academy of Sciences*, *120*(4), e2118466120.  
<https://doi.org/10.1073/pnas.2118466120>
- Armando, M., Ochs, M., & Régner, I. (2022). The Impact of Pedagogical Agents' Gender on Academic Learning: A Systematic Review. *Frontiers in Artificial Intelligence*, *5*. <https://doi.org/10.3389/frai.2022.862997>
- Baake, J., Schmitt, J., & Metag, J. (2025). Balancing realism and trust: AI avatars in science communication. *Journal of Science Communication*, *24*(2), A03.
- Baylor, A. L., & Kim, Y. (2004). Pedagogical Agent Design: The Impact of Agent Realism, Gender, Ethnicity, and Instructional Role. In J. C. Lester, R. M. Vicari, & F. Paraguaçu (Hrsg.), *Intelligent Tutoring Systems* (S. 592–603). Springer.  
[https://doi.org/10.1007/978-3-540-30139-4\\_56](https://doi.org/10.1007/978-3-540-30139-4_56)
- Boring, A. (2017). Gender biases in student evaluations of teaching. *Journal of public economics*, *145*, 27–41.

- Breucker, G., Sochert, R. & Mißler, M. (2015, März). *Kein Stress mit dem Stress – Eine Handlungshilfe für Beschäftigte*. Initiative Neue Qualität der Arbeit. Bundesministerium für Arbeit und Soziales.  
<https://www.inqa.de/DE/mediathek/publikationen/kein-stress-mit-dem-stress-eine-handlungshilfe-fuer-beschaefigte.html>
- Calaguas, G. M. (2012). Teacher effectiveness scale in higher education: Development and psychometric properties. *International Journal of Research Studies in Education*, 2(1). <https://doi.org/10.5861/ijrse.2012.108>
- Cheryan, S., Ziegler, S. A., Montoya, A. K., & Jiang, L. (2017). Why are some STEM fields more gender balanced than others? *Psychological Bulletin*, 143(1), 1–35.  
<https://doi.org/10.1037/bul0000052>
- Cohen, J. (2013). *Statistical Power Analysis for the Behavioral Sciences* (2. Aufl.). Routledge. <https://doi.org/10.4324/9780203771587>
- Craiu, M.-V., & Iancu, I. R. (2022). Is technology gender neutral? A systematic literature review on gender stereotypes attached to artificial intelligence. *Human Technology*, 18(3), 297–315.
- Croft, A., Schmader, T., & Block, K. (2015). An underexamined inequality: Cultural and psychological barriers to men’s engagement with communal roles. *Personality and Social Psychology Review*, 19(4), 343–370.
- Danner, D. (2015). *Reliabilität—Die Genauigkeit einer Messung (GESIS Survey Guidelines)* (Version 1.1). GESIS - Leibniz Institute for the Social Sciences.  
[https://doi.org/10.15465/GESIS-SG\\_011](https://doi.org/10.15465/GESIS-SG_011)

- de Borst, A. W., & de Gelder, B. (2015). Is it the real deal? Perception of virtual characters versus humans: An affective cognitive neuroscience perspective. *Frontiers in psychology*, 6, 576.
- Eagly, A. H., & Karau, S. J. (2002). Role congruity theory of prejudice toward female leaders. *Psychological Review*, 109(3), 573–598. <https://doi.org/10.1037/0033-295X.109.3.573>
- Eagly, A. H., Nater, C., Miller, D. I., Kaufmann, M., & Sczesny, S. (2020). Gender stereotypes have changed: A cross-temporal meta-analysis of U.S. public opinion polls from 1946 to 2018. *American Psychologist*, 75(3), 301–315. <https://doi.org/10.1037/amp0000494>
- E-Learning-Studie 2025: Die Ergebnisse.* (2025, Juni 25). WTT CampusONE. [https://wtt-campusone.com/ueber-uns/news/e-learning-studie-2025-wie-digitale-weiterbildung-in-deutschland-wirklich-funktioniert?utm\\_source=Mafo&utm\\_medium=eLearningjournal&utm\\_campaign=Pressemitteilung](https://wtt-campusone.com/ueber-uns/news/e-learning-studie-2025-wie-digitale-weiterbildung-in-deutschland-wirklich-funktioniert?utm_source=Mafo&utm_medium=eLearningjournal&utm_campaign=Pressemitteilung)
- Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G\*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, 39(2), 175–191. <https://doi.org/10.3758/BF03193146>
- Fink, M. C., Robinson, S. A., & Ertl, B. (2024). AI-based avatars are changing the way we learn and teach: Benefits and challenges. *Frontiers in Education*, 9, 1416307.
- Frauenanteil in der technischen Forschung und Entwicklung binnen zehn Jahren von 11 % auf 18 % gestiegen.* (2024, April 23). DESTATIS Statistisches Bundesamt. [https://www.destatis.de/DE/Presse/Pressemitteilungen/Zahl-der-Woche/2024/PD24\\_17\\_p002.html](https://www.destatis.de/DE/Presse/Pressemitteilungen/Zahl-der-Woche/2024/PD24_17_p002.html)

- Guzman, A. L., & Lewis, S. C. (2020). Artificial intelligence and communication: A human–machine communication research agenda. *New media & society*, 22(1), 70–86.
- Heilman, M. E., Caleo, S., & Manzi, F. (2024). Women at Work: Pathways from Gender Stereotypes to Gender Bias and Discrimination. *Annual Review of Organizational Psychology and Organizational Behavior*, 11(Volume 11, 2024), 165–192. <https://doi.org/10.1146/annurev-orgpsych-110721-034105>
- Jugovic, I. P., Maskalan, A., & Pavin Ivanec, T. (2022). Gender differences and motivation for the teaching profession: Why do men choose (not) to teach? *Australian Journal of Teacher Education (Online)*, 47(9), 1–18.
- Kammermeier, M., Muckenthaler, M., Weiß, S., & Kiel, E. (2025). Feminization of teaching: Gender and motivational factors of choosing teaching as a career. *Frontiers in Education*, 10, 1471015.
- Kim, Y., Baylor, A. I., & Shen, E. (2007). Pedagogical agents as learning companions: The impact of agent emotion and gender. *Journal of Computer Assisted Learning*, 23(3), 220–234. <https://doi.org/10.1111/j.1365-2729.2006.00210.x>
- Koenig, A. M., & Eagly, A. H. (2014). Evidence for the social role theory of stereotype content: Observations of groups' roles shape stereotypes. *Journal of Personality and Social Psychology*, 107(3), 371–392. <https://doi.org/10.1037/a0037215>
- Lange, P. A. M. van, Kruglanski, A. W., & Higgins, E. T. (2012). *Handbook of theories of social psychology*. Sage.
- Löhr B. (2022, März 23). *Stressbewältigung: Tipps für akuten und chronischen Stress*. AOK Gesundheitsmagazin. <https://www.aok.de/pk/magazin/wohlbefinden/stress/stressbewaeltigung-tipps-fuer-akuten-und-chronischen-stress/>

- McDonnell, M., & Baxter, D. (2019). Chatbots and gender stereotyping. *Interacting with Computers*, 31(2), 116–121.
- McPherson, M. A., Jewell, R. T., & Kim, M. (2009). What Determines Student Evaluation Scores? A Random Effects Analysis of Undergraduate Economics Classes. *Eastern Economic Journal*, 35(1), 37–51. <https://doi.org/10.1057/palgrave.eej.9050042>
- Moreno, K. N., Person, N. K., Adcock, A. B., Eck, R., Jackson, G. T., & Marineau, J. C. (2002). Etiquette and efficacy in animated pedagogical agents: The role of stereotypes. *AAAI symposium on personalized agents*, Cape Cod, MA.
- OpenAI. (2025). *ChatGPT* (Version GPT-5.3) [KI-Sprachmodell]. <https://chat.openai.com/>
- Rizvi, M. N. (2024). Unleashing the power of AI avatars: Revolutionizing education for the digital generation. *Journal of Emerging Technologies and Innovative Research*, 11(6), 36–44.
- Schmidt, J. (2024, März 11). *Reframing lernen: Wie kognitive Umstrukturierung Ihren Stress in Deutschland lindert*. Stress Burnout Management. <https://www.stress-burnout-management.de/wie-sie-durch-perspektivenwechsel-stressoren-um-60-weniger-bedrohlich-erleben-die-reframing-techniken/>
- Shiban, Y., Schelhorn, I., Jobst, V., Hörnlein, A., Puppe, F., Pauli, P., & Mühlberger, A. (2015). The appearance effect: Influences of virtual agent features on performance and motivation. *Computers in Human Behavior*, 49, 5–11. <https://doi.org/10.1016/j.chb.2015.01.077>
- So erkennen Sie Stresssymptome*. (2022, November 15). AOK Gesundheitsmagazin. <https://www.aok.de/pk/magazin/wohlbefinden/stress/so-erkennen-sie-stresssymptome/>

- Suhr, F. (2023, April 27). *Infografik: Frauenberufe, Männerberufe?* Statista Daily Data. <https://de.statista.com/infografik/29862/berufe-in-deutschland-nach-geschlechteranteil>
- Tao, Z., Liu, Y., Qiu, J., & Li, S. (2025). Impact of virtual avatar appearance realism on perceptual interaction experience: A network meta-analysis. *Frontiers in Psychology, 16*, 1624975.
- Urmersbach B. (2025, Dezember 17). *Themenseite: Frauen auf dem Arbeitsmarkt*. Statista. <https://de.statista.com/themen/12735/frauen-auf-dem-arbeitsmarkt/>
- Wadhvani, P. & Ambekar, A. (2024, Juni). *Marktgröße und Marktanteil von KI-Avataren | Statistikbericht 2024-2032*. Global Market Insights Inc. <https://www.gminsights.com/de/industry-analysis/ai-avatars-market>
- Wie Cloud Storage Ihre Infrastruktur voranbringt*. (2025, April 1). [PROACT]. <https://www.proact.de/blog/wie-cloud-storage-ihre-infrastruktur-voranbringt/>

## Übersicht verwendeter Hilfsmittel

Im Rahmen der Erstellung dieser Arbeit wurden KI-gestützte Tools unterstützend verwendet. Dabei kam insbesondere ChatGPT zur sprachlichen Überarbeitung eigener Textentwürfe zum Einsatz. Dies umfasste die Ausformulierung stichpunktartiger Vorüberlegungen, die stilistische Optimierung bereits verfasster Abschnitte sowie die Anpassung an einen wissenschaftlichen Sprachstil. Auch bei der konzeptionellen Reflexion einzelner Arbeitsschritte diente das Tool als unterstützendes Diskussionsinstrument. Zudem wurde ChatGPT und HeyGen bei der Erstellung der Lehrvideos, wie in Abschnitt 3.3 Versuchsmaterialien beschrieben, unterstützend genutzt. Zur Unterstützung der Literaturrecherche wurde zudem das Tool Consensus verwendet, um relevante wissenschaftliche Publikationen effizienter zu identifizieren und einen Überblick über bestehende Forschungsergebnisse zu erhalten. Zusätzlich wurde die Arbeit mit der Software Turnitin auf mögliche Übereinstimmungen mit anderen Quellen überprüft.

Die inhaltliche Konzeption, die Entwicklung der Forschungsfragen und Hypothesen, die methodische Umsetzung, die Datenauswertung sowie sämtliche inhaltlichen Argumentationen wurden eigenständig erbracht. Es wurden keine Textabschnitte ohne eigene inhaltliche Vorleistung generiert oder unverändert übernommen.

## Anhang

### 1. HeyGen-Lehrvideos

Arbeitsgesundheit Lehrvideo männlicher Avatar: <https://syncandshare.lrz.de/get-link/fiEK1odymsu7U5wbsxX44q/Arbeitsgesundheit%20MA.mp4>

Arbeitsgesundheit Lehrvideo weiblicher Avatar: <https://syncandshare.lrz.de/get-link/fiGqvamvoew2eBNJurCfac/Arbeitsgesundheit%20WA.mp4>

Informatik Lehrvideo männlicher Avatar: <https://syncandshare.lrz.de/get-link/fiNp1Hke3zdi46taHh9Nkj/Cloud%20Speicherung%20MA.mp4>

Informatik Lehrvideo männlicher Avatar: <https://syncandshare.lrz.de/get-link/fiFC9wxmkNwht165kYgAyG/Cloud%20Speicherung%20WA.mp4>

### 2. Verwendete Items im Fragebogen

#### 2.1. Items zur Wahrnehmung des Lernerfolgs

Wissensfragen zum Lehrvideo Arbeitsgesundheit		
Stress ist eine biologische Aktivierungsreaktion und nicht lediglich ein subjektives Gefühl.	Richtig	
Kortisol wird ausschließlich dann ausgeschüttet, wenn die Arbeitsmenge dauerhaft zu hoch ist.		Falsch
Kurzfristiger Stress kann leistungsfördernd wirken, solange ausreichende Erholungsphasen vorhanden sind.	Richtig	
Stress entsteht nur dann, wenn eine Situation objektiv zu schwierig ist.		Falsch
Das autonome Nervensystem spielt eine zentrale Rolle bei der körperlichen Stressaktivierung.	Richtig	
Frühwarnsignale von Stress umfassen körperliche, emotionale und soziale Symptome umfassen.		Falsch
Reizbarkeit und Fehleranfälligkeit sind emotionale Symptome, die bei Stress auftreten können.		Falsch
Dauerunterbrechungen und unklare Erwartungen zählen zu den typischen Belastungsfaktoren.	Richtig	
Ignorierte Frühwarnsignale verschwinden meist von selbst, wenn man lange genug weitermacht.		Falsch
Stress entsteht häufig durch fehlende Kontrolle oder Unklarheit – nicht nur durch Arbeitsmenge.	Richtig	
Kurze Regenerationsphasen von 1–2 Minuten können die Stressreaktion bereits messbar reduzieren.	Richtig	
Kognitives Reframing bedeutet, belastende Gedanken vollständig zu stoppen.		Falsch
Grenzmanagement reduziert Stress, weil es Unsicherheiten und unklare Erwartungen strukturiert.	Richtig	
Die 4-6-Atmung wirkt, weil sie die körperliche Aktivierung verlangsamt und die Regulation unterstützt.	Richtig	

Soforttechniken machen langfristige Präventionsmaßnahmen überflüssig, wenn sie regelmäßig angewendet werden.		Falsch
--	--	--------

Wissensfragen zum Lehrvideo Informatik		
Cloud-Speicherung basiert auf verteilten Systemen, in denen Daten auf mehreren Servern parallel liegen können.	Richtig	
Eine Datei wird in der Cloud grundsätzlich nur an einem physischen Ort gespeichert, um Ladezeiten zu optimieren.		Falsch
Beim Speichern können Dateien in kleinere Fragmente zerlegt werden, die auf verschiedene Speicherkomponenten verteilt werden.	Richtig	
Für den Zugriff auf Cloud-Speicher ist üblicherweise eine aktive Internetverbindung erforderlich.	Richtig	
Cloud-Systeme skalieren automatisch, sodass Speicherressourcen auch ohne manuelle Eingriffe erweitert werden können.	Richtig	
Cloud-Anbieter setzen Verschlüsselung ein, sowohl während der Übertragung als auch bei gespeicherten Daten.	Richtig	
Die Cloud ist automatisch sicher, unabhängig davon, wie Nutzer*innen damit arbeiten.		Falsch
Zu weit gefasste Freigaben können dazu führen, dass Unbefugte Zugriff auf Daten erhalten.	Richtig	
Private Accounts bieten denselben Schutz wie Unternehmensaccounts, solange ein Passwort verwendet wird.		Falsch
Lokale Offline-Kopien können ein Sicherheitsrisiko darstellen, wenn ein Gerät verloren geht.	Richtig	
Die Cloud ermöglicht es, dass mehrere Personen gleichzeitig an derselben Datei arbeiten, ohne dass separate Versionen entstehen.	Richtig	
Automatische Synchronisation reduziert die Wahrscheinlichkeit von Versionskonflikten.	Richtig	
Dateien sollten bevorzugt lokal gespeichert werden, um sie unabhängig von Cloud-Systemen kontrollieren zu können.		Falsch
Freigaben sollten nur an Personen erfolgen, die sie tatsächlich benötigen, um Datenlecks vorzubeugen.	Richtig	
Die Nutzung des Unternehmensaccounts ist entscheidend, um Daten im Organisationskontext kontrollierbar und geschützt zu halten.	Richtig	

## 2.2. Items zur Wahrnehmung der didaktischen Kompetenz

Nr.	Item	Skala
1	Die Lehrperson zeigt eine hohe Beherrschung der vermittelten Inhalte.	1–5
2	Die Lehrperson verfügt über umfassendes Wissen zu den behandelten Themen.	1–5
3	Die Lehrperson teilt viele interessante und relevante Gedanken zu den Lerninhalten.	1–5
4	Die Lehrperson vermittelt Ihre Ideen klar und verständlich.	1–5
5	Die Lehrperson formuliert realistische Ziele, die Lernende erreichen können.	1–5

## 2.3. Items zur Messung der Manipulationskontrolle

Nr.	Item	Skala
1	Informatik wird in der Gesellschaft eher mit Männern als mit Frauen verbunden.	1–5
2	Ich persönlich nehme Informatik als eher männlich konnotiert wahr.	1–5
3	Menschen, die in der Informatik arbeiten, sind typischerweise Männer.	1–5
4	Informatik gilt allgemein als „typisch männlicher“ Bereich.	1–5
5	Arbeitsgesundheit wird in der Gesellschaft eher mit Frauen als mit Männern verbunden.	1–5
6	Ich persönlich nehme Arbeitsgesundheit als eher weiblich konnotiert wahr.	1–5

7	Menschen, die im Bereich Arbeitsgesundheit arbeiten, sind typischerweise Frauen.	1–5
8	Arbeitsgesundheit gilt allgemein als „typisch weiblicher“ Bereich.	1–5

#### 2.4. Items zur Erhebung des Vorwissens

Nr.	Item	Skala
1	Ich kenne mich mit dem Thema Cloud-Speicherung von Dateien sehr gut aus und habe keinen Lernbedarf.	„Ich stimme zu“ oder „Ich stimme nicht zu“
2	Ich kenne mich mit dem Thema Mentale Arbeitsgesundheit und Stressprävention sehr gut aus und habe keinen Lernbedarf.	„Ich stimme zu“ oder „Ich stimme nicht zu“

#### 2.5. Items zum wahrgenommenen Realismus

Nr.	Item	Skala
1	Die Lehrperson wirkte realistisch dargestellt.	1–5
2	Die Lehrperson erschien mir lebensecht und natürlich.	1–5
3	Ich hatte den Eindruck, dass die Lehrperson authentisch wirkte.	1–5
4	Die Lehrperson fühlte sich eher wie eine echte Person als wie eine computergenerierte Figur an.	1–5
5	Das Erscheinungsbild und die Bewegungen der Lehrperson wirkten natürlich und glaubwürdig.	1–5