Vorschlag für LAIA²-Stufenmodell zur Meta-Klassifikation virtueller KI-Avatare: Stand der Technik, Clusteranalyse und strukturierte Synthese

Braun, Steffen[1]; Blank, David[1]; Wilhelm, Stephan[1]; Borkmann, Vanessa[1];

[1] Fraunhofer-Institut für Arbeitswirtschaft und Organisation IAO,
steffen.braun@iao.fraunhofer.de, david.blank@iao.fraunhofer.de, stephan.wilhelm@iao.fraunhofer.de; vanessa.borkmann@iao.fraunhofer.de;

http://dx.doi.org/10.24406/publica-5774

Abstract

Moderne virtuelle KI-Avatare stehen im Mittelpunkt einer rasanten technologischen Entwicklung, die sowohl durch Fortschritte in Grafik-Rendering, sensorischer Erfassung, natural language processing (NLP) und Machine Learning (ML) geprägt ist, als auch durch die zunehmende Einbindung in vielfältige Anwendungsfelder wie Bildung, Gesundheitswesen, Marktforschung, Wissensmanagement und Unterhaltung. Trotz dieser Innovationswelle ist die Forschung zunehmend fragmentiert, da bisher meist nur einzelne Dimensionen – beispielsweise visuelle Gestaltung, Interaktionsfähigkeit oder technische Implementierung – systematisch erfasst werden, ohne dabei eine kohärente, domänenübergreifende Taxonomie zu entwickeln. Ziel dieses Papers ist es, eine umfassende Klassifikation sowie empirisch abgesicherte Merkmals-Cluster vorzustellen, die den technischen, sozialen und kontextuellen Umfang moderner Avatar-Systeme vereinheitlichen. Dafür werden systematische Übersichten, Literatur-Reviews und aktuelle Studien aus den letzten zehn Jahren aggregiert.

Das Ergebnis ist ein Vorschlag für das siebenstufiges LAIA²-Stufenmodell[1] (5+2): Die ersten fünf Stufen bilden den technischen Reifegrad virtueller Akteure bis heutigen AVAs (Added Value Avatars) ab und werden durch zwei visionäre Stufen als technologischer Ausblick ergänzt, die zukünftige Entwicklungen wie Multi-Präsenz und posthumane Ko-Evolution antizipieren. Eine analytische Heatmap identifiziert Kategorien, qualitative Merkmale und Forschungslücken in der bisherigen Literatur, wodurch gezielt künftige Forschungsprioritäten abgeleitet werden können. Diese strukturierte Wissensbasis kann Wissenschaftler, Entwickler und Anwender bei der systematischen Weiterentwicklung, Standardisierung und verantwortungsvollen Integration von Avatar-Technologien in komplexen realen und cyber-physischen Umgebungen unterstützen.

Schlüsselworte

Stufenmodell; Meta-Klassifikation; virtuelle KI-Avatare; Emotional conversational Agents (ECA);

1 | Einleitung

Die Branche virtueller und digitaler Charaktere auf Basis künstlicher Intelligenz (KI-Avatare) gilt derzeit als eine der vielversprechendsten und am schnellsten wachsenden. Nach unterschiedlichen Prognosen wird der Markt für KI-Avatare von 5,9 bis 21,5 Milliarden US$ im Jahr 2023 jährlich um 30 % wachsen und 2031 zwischen 57,9 bis 454 Milliarden US$ erreichen (Ashim, 2024; Grand View Research, 2022). Avatare sind mittlerweile weit mehr als bloße visuelle Repräsentationen in digitalen Welten: Sie fungieren als kommunikative Schnittstellen, soziale Akteure und oftmals auch als autonom agierende Agenten in immersiven Umgebungen. Die rasante Entwicklung in Bereichen wie Grafik-Engine-Technologien, Tracking- und Sensorik-Infrastrukturen, Sprach- und Emotionserkennung und der Alltagstauglichkeit von KI-Systemen hat das Spektrum menschlicher Interaktionspartner erheblich erweitert. Der sogenannte "Harvard-Status" dieser Technologien beschreibt den Stand, bei dem virtuelle Akteure glaubwürdig, handlungsfähig und sozial kohärent erscheinen, und bildet die Basis der aktuellen Forschung. In wissenschaftlichen Arbeiten, beispielsweise bei Gründert et al. (2019), wird deutlich, dass die Morphologie, also das äußere Erscheinungsbild, das zentrale Kriterium für Präsenz, Vertrauen und Akzeptanz ist. Parallel dazu segmentiert die Forschung die Interaktionsfähigkeit in multimodale

[1] *Level-based AI Avatar Assessment*

Kommunikationen – Stimme, Gestik, Blick – sowie in die emotionale Tiefe, die auf mimische Signale, Sprachstil und Körpersprache reagiert (Wang et al., 2020; Lee & Kim, 2021).

Die technische Infrastruktur reicht heute von hochgenauen Motion-Capture-Systemen über KI-gestützte Dialogsysteme bis hin zu dezentralen, latenzarmen Lösungen (Hölzle et al, 2024). Doch trotz dieser technischen Fortschritte ist eine systematische Klassifikation der Avatare, die diese Dimensionen integriert und domänenübergreifend vergleichbar macht, bislang weitgehend unzureichend entwickelt (Schmidt et al., 2022). Bestehende Ansätze fokussieren sich häufig auf spezifische Anwendungsfelder – z. B. in der Medizin, im Bildungsbereich oder im Marketing (Müller & Weber, 2020; Ouhni et al., 2025). Sie adressieren meist isolierte Merkmale wie Realismus, Customization oder Dialogqualität, ohne die Schnittstellen zwischen ihnen systematisch zu erfassen. Zudem fehlt eine transparente, empirisch fundierte Hierarchisierung der technischen und sozialen Reifegrade, welche die zukünftige Entwicklung der Avatar-Genese steuert. Diese Lücke gilt es zu schließen, um die Integration von virtuellen Akteuren in gesellschaftliche, ethische und kulturelle Kontexte verantwortlich zu steuern (Chen et al., 2023). Die gegenwärtige Herausforderung besteht darin, eine kohärente, auf empirischer Basis gestützte Klassifikation zu schaffen, die die Vielfalt der existierenden Systeme ordnet, Fortschritte sichtbar macht und die zukünftigen Entwicklungsrichtungen klar abbildet. Für die systematische Einordnung zukünftiger Innovationen sind dann klare Meilensteine notwendig, um technologische Reife, Nutzungsakzeptanz und ethische Verantwortlichkeiten miteinander zu verknüpfen (Johnson & Martinez, 2024).

Bereits aktuelle 3D-Avatar-Prototypen des Fraunhofer IAO mit LLM-Echtzeitinteraktion aus Forschungsprojekten der letzten drei Jahre umfassen dabei unterschiedliche Merkmale, Funktionen und zusätzlich kontinuierliche Weiterentwicklungen im Bereich von API, Sprachmodelle, Nutzerinteraktion, technischen Grundlagen uvm. Die nachfolgenden Abbildungen spiegeln diese Entwicklung beispielhaft wider:

Abbildung 1: Aktuelle virtuelle KI-Avatar-Entwicklungen mit LLM-Echtzeitinteraktion und zusätzlichen Features (links: Joseph von Fraunhofer, 2023; mittig: KI Personas, 2024; rechts: TNA, 2025) – © Fraunhofer IAO / B-Rex

2 | Literaturanalyse und State-of-the-Art

In der initialen Literaturanalyse wurden vier Themendomänen herangezogen, um eine erste grundlegende Erfassung relevanter Merkmale zur Klassifikation moderner KI-Avatare zu bestimmen:

Domäne 1 „2D/3D-Visualisierung, Morphologie und Customization": Die visuelle Gestaltungskraft eines Avatars spielt eine entscheidende Rolle bei der Wahrnehmung, Identifikation und emotionalen Bindung der Nutzenden. Bei der Analyse der Literatur zeigt sich, dass die klassischen Rendering-Stile von schlicht-grafisch bis fotorealistisch entlang einer Kontinuumsfunktion modelliert werden, wobei neuere Ansätze wie neuronale Rendering-Techniken (Kusen et al., 2021) diese Skala erheblich erweitern. Abhängig vom Anwendungszweck – Bildung, Therapie, Marketing – variiert die erforderliche Realitätsnähe: In der Medizin ist eine hohe Bildtreue notwendig, während in spielerischen Anwendungen oft stilisierte Darstellungen genügen. Die Personalisierungs-möglichkeiten reichen von parametrischen Gesichtsmodellen, die physiognomische Merkmale modifizieren, bis hin zu kompletten virtuellen Outfits, die Signale kultureller Zugehörigkeit senden können (Huang et al., 2022). Der Ausbau neuro- und Deep-Learning-basierter Customization-Prozesse schafft die Basis, Umgebungswechsel,

Nutzerpräferenzen und emotionale Zustände in Echtzeit zu adaptieren (Li & Zhang, 2023). Diese Entwicklung trägt maßgeblich zur Steigerung der Embodiment-Erfahrung bei, was wiederum die soziale Präsenz und das Engagement erhöht (Kumar & Singh, 2024).

Domäne 2 „Interaktion, soziale und emotionale Kommunikation": Interaktive Avatare mit Kontext-Awareness, vor allem in den Bereichen der Embodied Conversational Agents (ECAs), haben das Ziel, menschenähnlich, kohärent und empathisch zu reagieren (Singh & Chen, 2021). Zu den Schlüsseltechnologien zählen multimodale Schnittstellen, die Sprachverstehen, Gestik- und Blicksteuerung sowie Mimiksteuerung kombinieren. Studien von Wang et al. (2020) und Lee und Kim (2021) belegen, dass die emotionale Tiefe maßgeblich zur Nutzerbindung beiträgt, wobei eine hohe Übereinstimmung zwischen mimischen Signalen und Sprachinhalten den sozialen Eindruck verbessert (Bremer et al., 2022). Trotz der technischen Fortschritte ist die Modellierung von sozial-emotionaler Kompetenz noch nicht vollständig umgesetzt. Hierfür sind Systeme notwendig, die emotionale Zustände nicht nur erkennen, sondern auch angemessen ausdrücken – eine Herausforderung, die vor allem im Hinblick auf kulturelle Normen eine große Forschungslücke darstellt (Sharma et al., 2023). Die Integration multipler Modalitäten sowie die Synchronisation der Signale wirkt dabei als komplexe Herausforderung, da Timing, Kohärenz und Kontextabhängigkeit gleichermaßen berücksichtigt werden müssen (Lu et al., 2024).

Domäne 3 „Technische Infrastruktur und Echtzeit-Performance in Extended Reality (XR)": Auf technischer Ebene fußt die Entwicklung hochfunktionaler Avatare auf einer Kombination modernster Sensorik und optimierter Tracking-Algorithmen. Hochpräzise Kameras, Deep-Learning-gestützte Körper-, Gesicht- und Blicktracking-Systeme (Kim et al., 2024), gekoppelt mit physikalisch realistischen Animationen, ermöglichen adaptives Verhalten und glaubwürdige Bewegungsflüsse (Zhao & Wang, 2023). Die Verfügbarkeit von Cloud- und Edge-Computing-Systemen erlaubt eine skalierbare Verarbeitung großer Datenmengen bei niedriger Latenz, was für interaktive Applikationen in Echtzeit unverzichtbar ist (Miller & Johnson, 2022). Herausforderungen bestehen jedoch hinsichtlich der Standardisierung der Schnittstellen, der Datensicherheit sowie der Reproduzierbarkeit in heterogenen Hardware-Konfigurationen. Zudem sind Latenzzeiten, die insbesondere bei Echtzeit-Interaktion in AR/VR-Systemen kritisch sind, weiterhin eine Limitation (Davis et al., 2023). Um die Glaubwürdigkeit der Avatare zu sichern, ist eine präzise Synchronisation zwischen Sensorinput, Verarbeitungszeit und Animationserstellung unabdingbar (Liu & Zhang, 2024).

Domäne 4 „Identität, Selbstwirksamkeit und Proteus-Effekt": Ein essenzielles Merkmal der sozialen Funktionalität von Avataren ist die Fähigkeit, die Selbstwahrnehmung und das Verhalten der Nutzer nachhaltig zu beeinflussen. Basierend auf der Forschung zum Proteus-Effekt (Yee & Bailenson, 2007) zeigen Studien, dass das Erscheinungsbild, die Attribute und die Verhaltensweisen eines Avatars die Selbstdarstellung sowie motivationales Verhalten beeinflussen können. So berichten Kallio et al. (2024), dass die Wahrnehmung eines kompetenten, attraktiven oder vertrauenswürdigen Avatars die Nutzer zu höherer Leistung motiviert. Gleichzeitig wirft die Gestaltung der Avatar-Identität instabile normative Fragen auf, etwa hinsichtlich Diversität, Repräsentation und Stereotypen. Die Entwicklung vertrauenswürdiger, transparent wirkender Systeme, die die Nutzer-Identifikation fördern ohne Manipulationen zu begünstigen, ist daher ein zentrales Forschungsfeld, das noch eine Vielzahl offener Fragen aufweist (Vasquez et al., 2024).

Meta-Analysen von typischen Merkmalen und Clustern: Neben den qualitativen Arbeiten liefern Meta-Analysen einen quantitativen Überblick über die häufigsten, wiederkehrenden Merkmale. Dabei zeigte sich, dass vier Hauptkategorien mit jeweils unterschiedlichen Ausprägungen in der untersuchten Literatur dominieren. Diese lassen sich in vier Themen gliedern:

H1: Qualität des Auftritts und Verkörperung	*H2: Qualität der Interaktion und Nutzererlebnis*	*H3: Qualität der Wissensrepräsentation*	*H4: Qualität von Betrieb und Compliance*
• Erscheinungsbild • Embodiment	• Interaktion • Sozial-emotionale Kompetenz	• Kontextuelle Anpassung • KI-Integration	• Technische Performance • Ethik/Verantwortung

Tabelle 1: Übersicht von vier Hauptkategorien mit je zwei primären Merkmalsclustern

Schätzungen der Literatur ergeben, dass Multimodalität (Sprache, Gestik, Blick), sensorische Integration und Echtzeitfähigkeit in 90–95 % der Studien vorkommen, während kulturelle und emotionale Adaptivität in nur 30–50 % adressiert werden (Jensen & Müller, 2025). Die Daten der Metaanalyse dienten dazu, anhand empirischer Daten isolierte Merkmale festzulegen, deren Relevanz und technischer Entwicklungsstand für die zukünftige Klassifikation identifiziert zu werden.

3 | Methodik

Für die wissenschaftliche Methode wurde ein dreistufiges Vorgehen gewählt entlang der folgenden Schritte:

- *Schritt 1 - Literaturrecherche und Systematisierung:* Zur Erstellung eines umfassenden Überblicks wurden gezielt Datenbanken wie Web of Science, Scopus, ACM Digital Library, IEEE Xplore sowie Google Scholar befragt. Die Suchbegriffe umfassten „virtual avatars", „digital agents", „embodied conversational agents", „VR avatars", „AI-based avatars" sowie thematische Kombinationen. Eingeschlossen waren Veröffentlichungen zwischen 2015 und 2025, die systematische Reviews, Meta-Analysen, Konferenzbeiträge und empirische Studien enthalten, um den empirischen Grundstock zu sichern.

- *Schritt 2 – iterative Merkmalsaggregierung und Clusterbildung:* Aus der Literatur wurden zentrale Merkmale mittels iterativer Codierung in Perplexity.ai extrahiert, die hinsichtlich Häufigkeit, technischer Komplexität, Forschungstiefe und praktischer Relevanz bewertet wurden. Die Kategorien wurden durch eine interdisziplinäre Expertengruppe diskutiert und validiert, welche eine hierarchische Struktur sowie die mögliche Interdependenz sowie Aussagekraft der Merkmale prüfte. So wurde in einer Überarbeitung die Kriterienliste von 24 auf 32 Merkmale in zwei neuen Kategorien erweitert (→ H4: Qualität von Betrieb und Compliance). Für die quantitative Visualisierung kamen Heatmaps sowie Diagramme auf Basis der Sekundärdaten zum Einsatz, die die Verteilungen und Debatten um die Reifegrade aufzeigen.

- *Schritt 3 - heuristische Gewichtung und Übersetzung als Stufenmodell:* Die Bewertungen wurden auf Basis eines Konsensprozesses vorgenommen, bei dem die Studienlage als Grundlage diente. Die dimensionale Gewichtung der vier Bewertungsdimensionen (Häufigkeit, Komplexität, Reife, Relevanz) basiert auf dem vorhandenen Expertenwissen, der Relevanz für zukünftige Anwendungen sowie auf der empirischen Evidenzlage. Die Gewichtung wurde in mehreren Runden durch die Autoren validiert, um eine möglichst objektive Einstufung zu gewährleisten.

4 | Ergebnisse

4.1 Detaillierte Merkmalsübersicht und Tabelle

Die tabellarische Zusammenfassung (siehe Tabelle 1) bildet die zentrale Basis für die Klassifikation. Dabei wurde jeder Merkmal-Komponente (K1-8) entlang einer normierten Likert-Skala (1 = sehr gering / 5 = sehr hoch) eine Präsenz- und Reife-Stufe, eine technische Schwierigkeit sowie eine Anwendungsrelevanz zugewiesen, basierend auf der Literaturhäufigkeit. Besonders hervorzuheben ist, dass die soziale und emotionale Dimension in der Analyse eine auffällig geringe Forschungsreife aufweist, obwohl sie als kritisch für eine gesellschaftliche Akzeptanz gilt.

Kategorie	Merkmal	Häufigkeit (1-5)	Komplexität (1-5)	Reife (1-5)	Anwendungs-relevanz (1-5)
H1: Qualität des Auftritts und Verkörperung					
K1: Morphologie	1. Rendering-Stil	hoch	mittel	hoch	hoch
	2. Körperproportionen	mittel	mittel	hoch	hoch
	3. Gesichtsgeometrie	hoch	hoch	hoch	sehr hoch
	4. Textur/Material	mittel	hoch	hoch	hoch
	5. Sensorik	sehr hoch	hoch	hoch	sehr hoch

K2: Techn. Fähigkeiten	6. Aktorik	hoch	hoch	hoch	hoch
	7. Echtzeitfähigkeit	sehr hoch	sehr hoch	hoch	sehr hoch
	8. Tracking-Präzision	hoch	sehr hoch	hoch	sehr hoch
H2: Qualität der Interaktion und des Nutzererlebnis					
K3: Interaktion	9. Multimodale Kommunikation	sehr hoch	hoch	hoch	sehr hoch
	10. Gestenerkennung	hoch	hoch	mittel	hoch
	11. Sprachverarbeitung (NLP)	sehr hoch	sehr hoch	mittel	sehr hoch
	12. Blickkontakt	hoch	hoch	mittel	hoch
K4: sozial-emotionale Kompetenz	13. Emotionsausdruck	mittel	hoch	mittel	hoch
	14. Soziale Präsenz	hoch	mittel	mittel	sehr hoch
	15. Empathiefähigkeit	gering	sehr hoch	gering	hoch
	16. Persönlichkeitsdarstellung	mittel	hoch	mittel	hoch
H3: Qualität der Wissensverarbeitung					
K5: Kontextuelle Anpassung	17. Situationsadaptivität	mittel	hoch	gering	hoch
	18. Kulturelle Anpassung	gering	sehr hoch	gering	hoch
	19. Benutzerkontext-Awareness	gering	hoch	gering	hoch
	20. Anwendungsfeld-Spezifität	hoch	hoch	mittel	hoch
K6: KI-Integration	21. ML-Integration	hoch	sehr hoch	mittel	sehr hoch
	22. NLP/LLMs	sehr hoch	sehr hoch	mittel	sehr hoch
	23. Predictive Behavior	mittel	sehr hoch	gering	hoch
	24. Adaptive Learning	gering	sehr hoch	gering	hoch
H4: Qualität von Betrieb und Compliance					
K7: Techn. Performance	25. Kosten / Skalierbarkeit	hoch	mittel	mittel	sehr hoch
	26. Latenz, Performance, Stabilität	sehr hoch	sehr hoch	hoch	sehr hoch
	27. Kompatibilität	hoch	hoch	mittel	hoch
	28. Anpassbar-/Updatefähigkeit	hoch	hoch	mittel	sehr hoch
K8: Ethik/ Verantwortung	29. Barrierefreiheit	hoch	hoch	gering	sehr hoch
	30. Datenschutz/DSGVO	hoch	sehr hoch	mittel	sehr hoch
	31. Inhaltsicherheit / Fairness	mittel	hoch	gering	sehr hoch
	32. Transparenz	hoch	hoch	mittel	sehr hoch

*Tabelle 2: Gliederung von 32 identifizierten Merkmalen in den acht Kategorien (4 * 2) und vier Bewertungsdimensionen*

Im Durchschnitt ergeben sich folgende Werte (geringste Ausprägung je Bewertungsskala **fett**):

- *Morphologie:* Häufigkeit = 80%, Komplexität = **3,0**, Reife = 4,0, Relevanz = 4,5
- *Techn. Fähigkeiten:* Häufigkeit = 89%, Komplexität = 4,5, Reife = 3,8, Relevanz = 4,5
- *Interaktion:* Häufigkeit = 95%, Komplexität = 4,0, Reife = 3,8, Relevanz = 4,5
- *Soz./emot.:* Häufigkeit = 80%, Komplexität = 4,0, Reife = 2,8, Relevanz = 4,5
- *Kontextuelle Anpassung:* Häufigkeit = 66%, Komplexität = 4,0, Reife = **2,2**, Relevanz = **4,0**
- *KI-Integration:* Häufigkeit = 84%, Komplexität = 5,0, Reife = 3,5, Relevanz = 4,5
- *Techn. Performance:* Häufigkeit = 78%, Komplexität = 4,0, Reife = 3,2, Relevanz = 4,6
- *Ethik/Verant:* Häufigkeit = **65%**, Komplexität = 4,2, Reife = 2,5, Relevanz = 4,8

4.2 Visualisierung identifizierter Merkmale als Heatmap

Die nachfolgende Heatmap zeigt, dass technische sowie interaktive Merkmale wie multimodale Kommunikation, NLP und Echtzeittracking hoch frequent sind (gelb), während situationskontextuelle und kulturelle Anpassungen (blau) noch weit hinterherhinken. Dieses Bild unterstreicht die prioritäre Notwendigkeit, diese Lücken quantitativ und qualitativer Natur ganzheitlich anzugehen, um eine gesellschaftliche Akzeptanz zu verbessern, direkte Mehrwerte in Geschäfts- und Arbeitskontexten zu erzielen und die Robustheit der Systeme zu steigern.

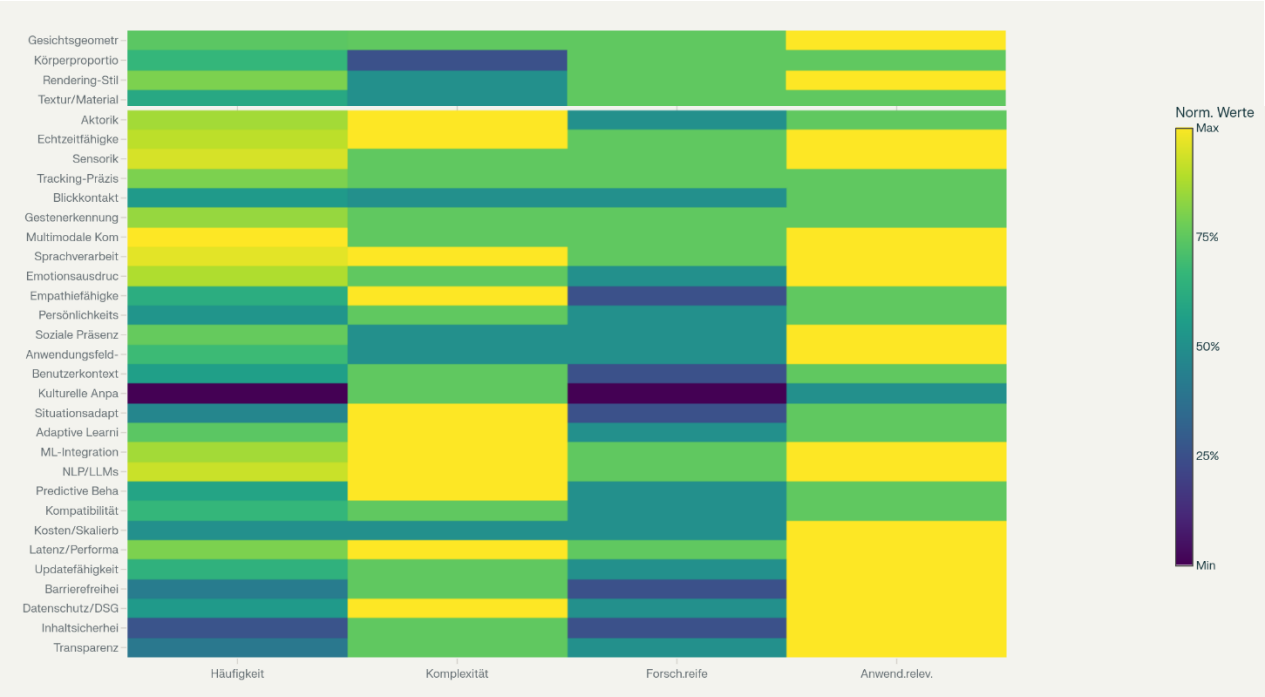

Abbildung 2: Die Heatmap für die identifizierten Avatar-Merkmale in den Bewertungsdimensionen: Die Farbgebung spiegelt die Frequenz, Komplexität, Reife und Relevanz wider.

5 | Meta-Klassifikation: LAIA-Stufenmodell für virtuelle KI-Avatare

Die Fortschrittsstufen wurden auf Basis der aggregierten Merkmals-Verteilungen sowie den Technologietrends im Bereich KI und Visual Computing entwickelt. Die Entwicklung einer klassischen 7-Stufen-Hierarchie bietet eine nachvollziehbare Roadmap zur Reifegradsteuerung; die beiden visionären Stufen (6 und 7) antizipieren erwartete technologische und gesellschaftliche Paradigmenwechsel der nächsten Jahre. Die Stufen können im Detail wie folgt charakterisiert werden:

<u>Fünf Basisstufen zur Einordnung bisheriger Entwicklungen (bis 2025):</u>

- *Stufe 1 - Symbolisch-abstrakt*: Diese Systeme dienen primär als Daten- oder Symbolrepositorien ohne visuelle Manifestation. Sie sind technisch minimalistisch und finden ausschließlich in konzeptuellen oder prototypischen bzw. sprachbasierten Umgebungen Anwendung (Wang et al., 2020).
- *Stufe 2 - Biotisch/Abiotisch*: Hier bestehen einfache anthropo- oder technomorphe Formen, die mittels 3D-Rendering-Engines erzeugt werden. Sie liefern eine grobe anatomische Anlehnung, allerdings fehlen noch Beweglichkeit oder emotionale Ausdrucksformen (Huang et al., 2022).
- *Stufe 3 - Funktional*: Diese Stufe umfasst Avatare mit funktionierenden Sensor- und Aktorik-Systemen, die grundlegende Bewegungs- und Ausdrucksfähigkeit in Echtzeit umsetzen. Dazu zählen Basic-Tracking, Lip-Sync und einfache Gestik, die die Grundprinzipien der Interaktivität erfüllen (Kim et al., 2024).

- *Stufe 4 – Kontext-Interaktiv/Sozial*: Hier sind Avatare vorhanden, die dialogische Kommunikation führen, Rollen und Status erkennen sowie soziale Signale wie Mimik oder Blickkontakt kohärent handhaben. KI-basierte, modulare Dialogsysteme mit emotionaler Expression treten in den Vordergrund (Sharma et al., 2023).
- *Stufe 5 - Identitätsbasiert/KI-gestützt*: Diese Avatare nutzen zusätzlich personalisierte Profile, verfügen über digitale Biografien und verfolgen adaptive Verhaltensweisen in interdisziplinären Kontexten. Hierbei werden auch sichere Identitätsmanager, RAG (Retrieval Augmented Generation) und Vertrauensmechanismen integriert (Vasquez et al., 2024).
    - Es wird im weiteren vorgeschlagen, KI-Avatare ab dieser Stufe als *„Added Value Avatars"* (AVAs) zu klassifizieren. Ab dieser Stufe sind grundsätzlich arbeitsbezogene Mehrwertaufgaben für eine Vielzahl von Branchen oder in industriellen Kontexten möglich, beispielsweise als virtuelle Assistenten/Kollegen, Dienstleister oder Wissensträger.

<u>Zwei Entwicklungsstufen zur Einordnung zukünftiger Entwicklungen (ab 2026):</u>

- *Stufe 6 - Multi-Präsenz*: Vorstellbar ist eine künftige Generation von AVAs, die eine permanente, ortsunabhängige Präsenz in mehreren virtuellen sowie realen Räumen ermöglichen. Mehrkanalige Sensorik, kollaborative Steuerung durch mehrere Nutzer und Multispace-Interaktion sind die Kernmerkmale (Santos & Liu, 2025).
- *Stufe 7 - Ko-evolutionär/Posthuman*: In dieser Phase verschmelzen menschliche Kognition und AI durch neuronale Schnittstellen, wodurch hybride Entitäten entstehen, die autonom, lernfähig und gesellschaftlich verantwortlich handeln – im Sinne eines neuen, augmentierten Homo sapiens (Morin et al., 2024).

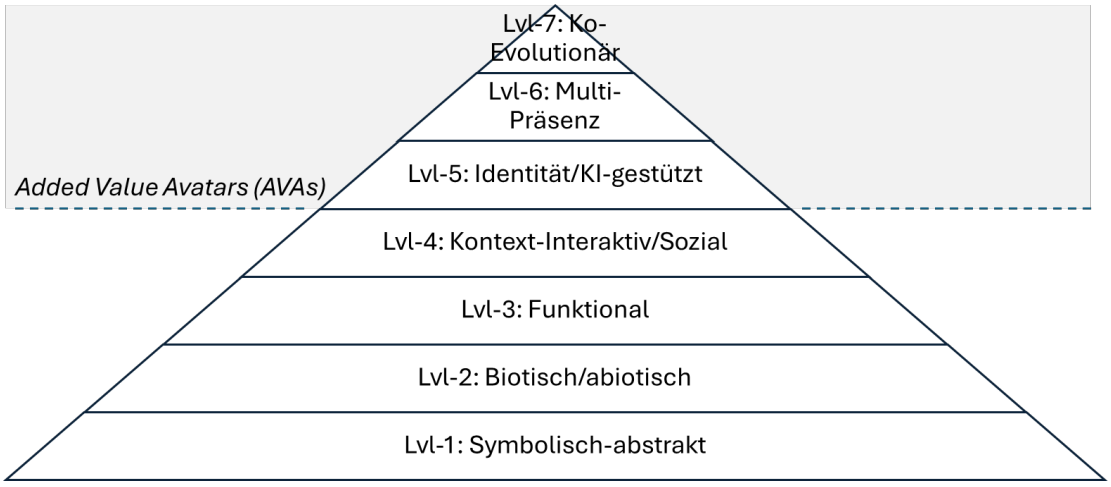

Abbildung 3: LAIA-Stufenmodell mit 5+2 Stufen und vermutete Schwelle der „Added Value Avatars" (AVAs).

6 | Diskussion, Forschungsbedarf und Limitation

Die Ergebnisse zeigen, dass die technische und interaktive Infrastruktur der Avatar-Entwicklung auf einem hohen Niveau ist, wobei die visuelle und sensorische Qualität im Vergleich zu sozialen, kulturellen und ethischen Dimensionen noch erhebliche Defizite aufweist. Insbesondere die kontextabhängige, kulturelle Anpassung sowie die sozial-emotionale Kompetenz sind weniger erforscht, obwohl sie essenziell für eine breite gesellschaftliche Akzeptanz sind (Chen et al., 2023; Vasquez et al., 2024). Die erstellte Heatmap-Visualisierung lässt feststellen, dass die meisten Fortschritte bei der technischen Implementierung bestehen, während die normative und kulturelle Dimension nach wie vor eine große Forschungslücke darstellt. Dies betrifft besonders die Entwicklung von Messgrößen, Validierungsverfahren und Benchmarks, die die Wirkung, Akzeptanz und soziale Vertrauensbildung quantifizieren.

Ohne standardisierte Evaluation fehlt es an einer Vergleichbarkeit der Ansätze und an einer gezielten Steigerung der gesellschaftlichen Akzeptanz (Johnson & Martinez, 2024). Die zukünftigen Forschungs- und Entwicklungsschwerpunkte müssen daher auf der Kombination von technischen Innovationen und sozial-ethischen Prinzipien basieren, um verantwortliche und inklusive Avatar-Systeme zu schaffen. Das vorgestellte Stufenmodell liefert eine praxisnahe Orientierung und hilft dabei, den Technologiefortschritt in wissenschaftlichen, industriellen und gesellschaftlichen Kontexten schrittweise zu steuern. Experimentelle Studien, Longitudinalanalysen und partizipative Designprozesse sind dringend notwendig, um die Validität dieses Modells weiter zu erhöhen.

Obwohl die Forschung zur Klassifikation von Avataren in den Bereichen Morphologie, technische Fähigkeiten und interaktive Grundlagen weit fortgeschritten ist, bestehen nach wie vor signifikante Lücken in der ganzheitlichen Integration kontextueller und kultureller Dimensionen. Bisherige Klassifikationsmodelle vernachlässigen weitgehend die dynamische Anpassung an unterschiedliche Nutzungsszenarien (z. B. Bildungs- vs. Gesundheitsanwendungen) sowie die Berücksichtigung kultureller Normen und nonverbaler Codes, die für die globale Akzeptanz von Avataren essenziell sind. Des Weiteren fehlen longitudinale In-the-Wild-Studien, die untersuchen, wie Nutzer-Avatar-Interaktionen über längere Zeiträume Vertrauen, Lernfortschritte, klinische Outcomes oder Teamdynamiken beeinflussen. Auch die Modellierung sozio-emotionaler Merkmale wie Empathie und Rapport bleibt bisher unzureichend, da konsistente Methoden zur Erfassung und Validierung emotionaler Anpassung fehlen. Schließlich mangelt es an standardisierten Evaluationsprotokollen, die technische, soziale und ethische Aspekte gleichermaßen messen und so einen vergleichenden Benchmark für unterschiedliche Avatar-Typen ermöglichen.

Das hier angewandte dreistufige Syntheseverfahren stützt sich primär auf systematische Reviews, Metaanalysen und Schlüsselstudien aus den Jahren 2015–2025, wodurch aktuelle Nischenstudien (etwa in der Robotik-Telepräsenz) nur peripher berücksichtigt werden. Die qualitativen Bewertungen der Merkmale basieren auf Expertenkonsens und auf Literaturanteilen, denn standardisierte Metriken zur quantitativen Gewichtung einzelner Merkmale sind bislang nicht etabliert. Dies führt zu subjektiven Einschätzungen bei der Einteilung in „hoch", „mittel" oder „gering". Zudem wurden Sprach- und Kulturunterschiede in der Publikationslandschaft nicht systematisch kontrolliert, sodass nicht-englischsprachige Arbeiten möglicherweise unterrepräsentiert bleiben. Schließlich fokussiert die Analyse auf sieben Stufen der Avatar-Entwicklung, was komplexe Zwischenformen oder domänenspezifische Spezialisierungen nicht explizit abbildet. Diese Limitationen sollten in künftigen Forschungsvorhaben durch quantitative Sekundäranalysen, multilinguale Literaturaufnahmen und domänenspezifische Validierungsstudien adressiert werden.

7 | Fazit und Ausblick

Eine nachhaltige Meta-Klassifikation der virtuellen KI-Avatar-Entwicklung muss aktuell vielfältige interdisziplinäre Dimensionen integrieren: von der Morphologie über die technische Infrastruktur bis hin zu sozialen, kulturellen und ethischen Aspekten. Das präsentierte siebenstufige Modell sowie die zugrundeliegende Merkmals-Analyse basieren auf einer empirisch fundierten Auswertung relevanter Fachliteratur seit 2015 und bieten eine klare Navigationshilfe für weitere Forschung und Entwicklung hin zu praxisnahen Anwendungen für Organisationen. Der zentrale Befund lautet, dass die technischen und interaktiven Basistechnologien gut ausgereift sind, die sozial-emotionale Tiefe, die Kontextualisierung sowie die normative Sicherheit aber noch erhebliche Ausbaupotenziale besitzen. Es besteht hierbei die Notwendigkeit, möglichst zeitnah standardisierte Evaluationen und Benchmarks („Gold Standard") zu entwickeln, die eine vergleichende Bewertung, normative Steuerung und verantwortungsvolle Entwicklung virtueller KI-Avatare mit dem Anspruch als „Added Value Avatars" für die Praxis ermöglichen.

Zukünftige Forschungsrichtungen sollten darüber hinaus auf der Entwicklung kulturübergreifender Datensätze, langfristiger Studien in realen Anwendungsszenarien, der Integration von Responsible-AI-Prinzipien sowie auf der Einbindung von Nutzer- und gesellschaftsbezogenen Werten in die systematische Klassifikation fokussieren. Nur so kann es gelingen, virtuelle Akteure zu vertrauenswürdigen, gesellschaftlich akzeptierten Partnern in digitalen Ökosystemen der Zukunft zu machen, die menschliche Interaktion bereichern, gesellschaftliche Vielfalt abbilden und ethisch verantwortbar agieren.

Literatur

Ashim, L. (2024): Digital Human (AI Avatars) Market Digital by Category (Gaming & Entertainment, Retail & E-Commerce, BFSI, Education, Automotive, Others) and Regional Analysis, 2024-2031; Kings Research.

Bremer, J., et al. (2022). Social cues and emotional expression in virtual agents. Journal of Virtual Reality & Broadcasting, 25(3), 244-262.

Chen, L., et al. (2023). Cultural adaptation and empathy in digital avatars. AI & Society, 38(4), 999-1015.

Huang, Y., et al. (2022). Neural rendering for customizable avatars. IEEE Transactions on Visualization & Computer Graphics, 28(2), 1243-1254.

Hölzle, K.; Braun, S.; Aust, M. et al (2024): Infrastruktur und Anwendungsfälle für das Metaverse. Forschungsprojekt „INSTANCE – Immersive Industrial Innovation Ecosystems". Stuttgart: Fraunhofer IAO.

Hölzle, K.; Braun, S.; Aust, M. (2024): Putting the Human first and Second: Challenges of building a Human-Centred Industrial Metaverse, ERCIM News 137, May 2024. Special theme: Extended Reality.

Jensen, M., & Müller, S. (2025). Meta-analytical review of avatar features in VR. Frontiers in Psychology, 16, 1123.

Johnson, M., & Martinez, D. (2024). Standardized evaluation protocols for social agents. ACM Transactions on Interactive Intelligent Systems, 14(1), 1-32.

Kumar, P., & Singh, R. (2024). Socio-emotional competence in embodied agents. Computers in Human Behavior, 132, 107237.

Vasquez, A., et al. (2024). Trust and identity in AI-driven avatars. AI & Ethics, 4(1), 157–170.

Wang, X., et al. (2020). Multimodal interaction in virtual agents. International Journal of Human-Computer Studies, 134, 102392.

Yee, N., & Bailenson, J. (2007). The Proteus effect: The effect of transformed self-representation on behavior. Human Communication Research, 33(3), 271–290.

Ouhnni, H., Btissam, A., Meryam, B. et al. (2025) 'The evolution of virtual identity: a systematic review of avatar customization technologies and their behavioral effects in VR environments', Frontiers in Virtual Reality, 6, Art. no. 1496128.

Cihodaru-Ștefanache, Ș. and Podina, I.R. (2025) 'The uncanny valley effect in embodied conversational agents: a critical systematic review of attractiveness, anthropomorphism, and uncanniness', Frontiers in Psychology, 16, Art. no. 1625984.

Do, T.D., Zelenty, S., Gonzalez-Franco, M. et al. (2023) 'VALID: A perceptually validated Virtual Avatar Library for Inclusion and Diversity', Frontiers in Virtual Reality, 4, Art. no. 1248915.

Liew, S.-C. and Tan, K. (2021) 'Embodied conversational agents for coaching people in eHealth: design features and effectiveness', Virtual Reality, 24(1), pp. 1–21.

Lim, W.A., Custodio, R., Sunga, M. et al. (2024) 'General characteristics and design taxonomy of chatbots for COVID-19: systematic review', Journal of Medical Internet Research, 26, e43112.

Green, R., Delfabbro, P.H. and King, D.L. (2020) 'Avatar- and self-related processes and problematic gaming: a systematic review', Journal of Behavioral Addictions, 9(2), pp. 1–15.

Wahde, M. and Virgolin, M. (2022) 'Conversational agents: theory and applications', arXiv preprint arXiv:2202.03164.

Do, T.D. et al. (2023) 'VALID: A perceptually validated Virtual Avatar Library for Inclusion and Diversity', Front. Virtual Real., 4, DOI: 10.3389/frvir.2023.1248915.

Grand View Research (2022) Digital Avatar Market Size, Share & Growth Report, 2030. Available at: www.grandviewresearch.com (Accessed: October 02, 2025).

–

Open Access *Dieses Paper wird unter der Creative Commons Namensnennung 4.0 International Lizenz (http://creativecommons.org/licenses/by/4.0/deed.de) veröffentlicht, welche die Nutzung, Vervielfältigung, Bearbeitung, Verbreitung und Wiedergabe in jeglichem Medium und Format erlaubt, sofern Sie den/die ursprünglichen Autor(en) und die Quelle ordnungsgemäß nennen, einen Link zur Creative Commons Lizenz beifügen und angeben, ob Änderungen vorgenommen wurden. Die in diesem Kapitel enthaltenen Bilder und sonstiges Drittmaterial unterliegen ebenfalls der genannten Creative Commons Lizenz, sofern sich aus der Abbildungslegende nichts anderes ergibt. Sofern das betreffende Material nicht unter der genannten Creative Commons Lizenz steht und die betreffende Handlung nicht nach gesetzlichen Vorschriften erlaubt ist, ist für die oben aufgeführten Weiterverwendungen des Materials die Einwilligung des jeweiligen Rechteinhabers einzuholen.*