

# Abschlussbericht HIVAM

Universitätsklinikum Hamburg-Eppendorf (UKE)

Förderkennzeichen: 16SV8880

Bewilligungszeitraum: 01.03.2022-31.08.2025

## Teil 1

### Aufgabenstellung

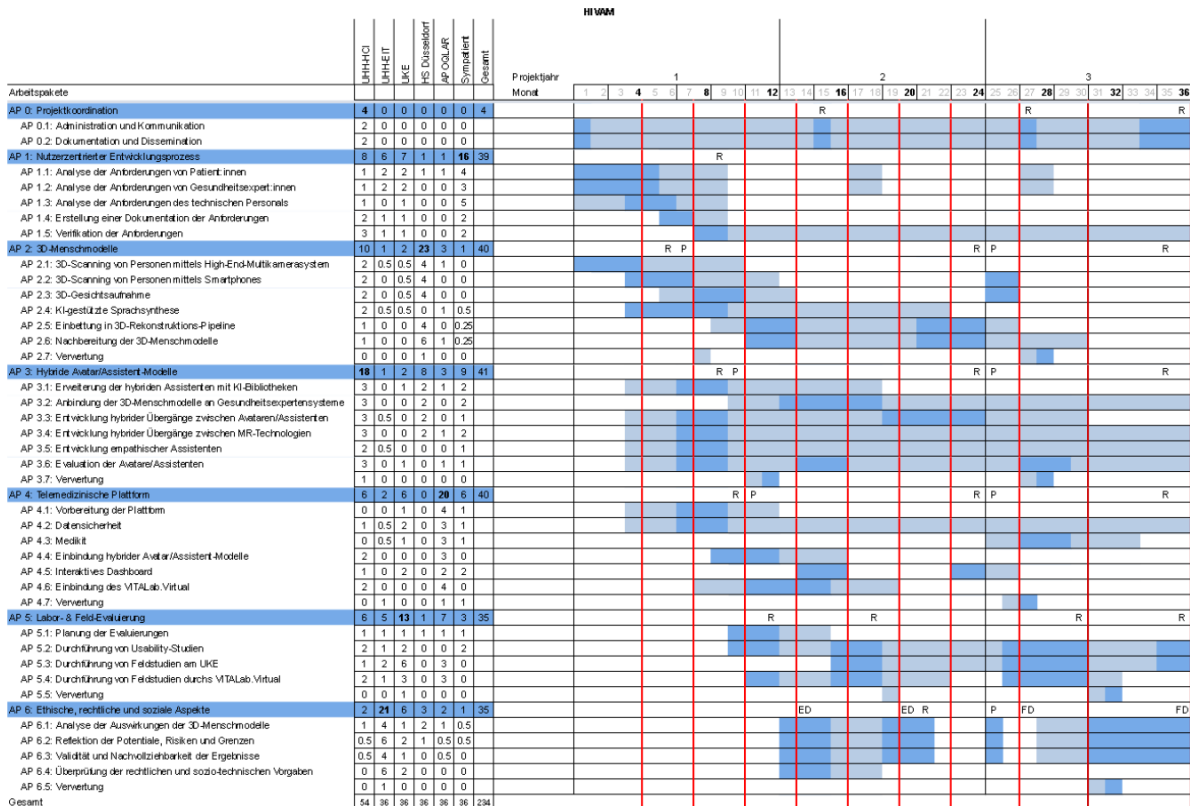
Das Verbundprojekt HIVAM entwickelt und erforscht eine immersive, MR-gestützte Telemedizinplattform, die die Expertise realer Gesundheitsexpert:innen mit hybriden Avatar- und Assistenzmodellen kombiniert. Patient:innen können so ohne physischen Kontakt sowohl mit den virtuellen Avataren realer Fachpersonen als auch mit intelligenten virtuellen Agenten interagieren – in einer Form, die räumliche, soziale Präsenz und Co-Präsenz erlebbar macht. Dies ermöglicht die kontaktlose Erfassung, Übertragung und Auswertung von Gesundheitsdaten sowie deren Nutzung für Diagnostik, Monitoring und Therapie. Neben der erweiterten Behandlung über die Telemedizinplattform entsteht dadurch ein KI-gestütztes Assistenzsystem zur Entscheidungsunterstützung, das die medizinische und pflegerische Versorgung optimiert, den Folgen sozialer Isolation und psychischer Belastung entgegenwirkt und Beratung, Visiten sowie Behandlungen vor Ort effizienter gestalten kann.

### Voraussetzungen

Zu Beginn der Projektlaufzeit war die Corona-Pandemie noch nicht vollständig überwunden. Dies führte zwar zu gewissen Einschränkungen bei persönlichen Gesprächen mit Patient:innen und Gesundheitspersonal, dennoch konnte die Projektdurchführung dank geeigneter Schutz- und Hygienekonzepte erfolgreich umgesetzt werden. Ferner gestaltete sich die Planung und Durchführung von Studien zum Einsatz generativer KI bei psychiatrischen Patient:innen aufgrund der Neuartigkeit des Forschungsfeldes sowie der noch nicht abschließend geklärten ethischen und datenschutzrechtlichen Rahmenbedingungen als herausfordernd.

### Planung und Ablauf

Das Projekt wurde vor Projektbeginn in Arbeitspakete strukturiert. Das Haupt-Arbeitspaket von UKE-Seite war AP5 „Labor- und Feldevaluationen“. Daneben waren wir an den Arbeitspaketen AP1 „Nutzerzentrierter Entwicklungsprozess“, AP2 „3D-Menschmodelle“, AP4 „Telemedizinische Plattform“ und AP6 „Ethische, rechtliche und soziale Aspekte“ eingebunden. Im Gantt-Chart unten ist die Struktur aufgezeigt.



## Wissenschaftlicher Stand

Die zunehmende Digitalisierung des Gesundheitswesens hat in den letzten Jahren auch die psychotherapeutische Versorgung nachhaltig verändert (1). Vor dem Hintergrund einer weiterhin hohen Prävalenz psychischer Erkrankungen bei gleichzeitig begrenzten Behandlungskapazitäten rücken digitale Unterstützungsangebote zunehmend in den Fokus der Forschung. Klassische Psychotherapie ist für viele Betroffene nur eingeschränkt zugänglich, etwa aufgrund langer Wartezeiten, hoher Kosten, regionaler Versorgungsunterschiede oder individueller Barrieren wie Scham oder Selbststigmatisierung (2,3). Vor diesem Hintergrund werden KI-basierte Anwendungen als potenzielles Mittel diskutiert, um Versorgungslücken zu überbrücken oder therapeutische Prozesse zu ergänzen.

Ein besonders dynamisches Forschungsfeld stellt der Einsatz KI-gestützter dialogischer Systeme dar. Solche Systeme können unterschiedliche Funktionen im psychotherapeutischen Kontext übernehmen: Sie werden als eigenständige niedrigschwellige Unterstützungsangebote, als Ergänzung zu laufenden Therapien (z. B. zur Übungsbegleitung, Motivation oder Psychoedukation) oder als Trainingsinstrumente für Therapeut:innen untersucht (4). Während frühe Ansätze computerbasierter Therapie bereits in den 1960er-Jahren entstanden (5), haben insbesondere Fortschritte in der Entwicklung großer Sprachmodelle (Large Language Models, LLMs) zu einer neuen Qualität dialogischer Systeme geführt. Diese ermöglichen flexibel generierte, kontextbezogene und zunehmend als empathisch wahrgenommene Interaktionen (6,7).

Intelligente virtuelle Agenten (IVAs) erweitern diese Systeme um eine visuelle und häufig verkörperte Darstellung, etwa in Form von Avataren. Aus psychotherapeutischer Perspektive ist dies insbesondere relevant, da nonverbale Kommunikationsaspekte, soziale Präsenz und die Entwicklung einer therapeutischen Allianz zentrale Wirkfaktoren psychotherapeutischer Behandlung darstellen. Erste Studien deuten darauf hin, dass IVAs durch ihre Gestaltung Zugänglichkeit fördern, Hemmschwellen senken und Selbstoffenbarung erleichtern können, insbesondere in sensiblen Themenbereichen (8). Gleichzeitig zeigen randomisiert kontrollierte Studien, dass chatbot-basierte Interventionen depressive und ängstliche Symptome reduzieren können, teilweise mit positiven Effekten auf wahrgenommene therapeutische Allianz (9,10,11). Allerdings bleibt bislang unklar, welche spezifischen Wirkmechanismen für diese Effekte verantwortlich sind und in welchem Ausmaß menschliche Interaktion ersetzbar oder lediglich ergänzbar ist.

Neben diesen Potenzialen werden in der klinisch-psychotherapeutischen Forschung zunehmend Risiken und Limitationen diskutiert. Dazu zählen Fragen der Akzeptanz und des Vertrauens in KI-basierte Systeme, insbesondere vor dem Hintergrund sensibler Gesundheitsdaten und bestehender Skepsis gegenüber automatisierten Entscheidungsprozessen (12,13). Zudem zeigen Studien, dass Nutzer:innen emotionale Bindungen zu IVAs entwickeln können, was Auswirkungen auf Abhängigkeit, Rollenerwartungen und das Verhältnis zu menschlichen Therapeut:innen haben könnte (14,15,16). Aus klinischer Sicht besonders relevant sind Risiken fehlerhafter oder unangemessener Reaktionen in kritischen Situationen, etwa bei Suizidalität oder Substanzmissbrauch. Systematische Reviews weisen darauf hin, dass generative KI-Modelle Defizite in der Risikoerkennung, Krisenintervention und Bereitstellung adäquater Hilfsangebote aufweisen können (17,18).

Zum Projektbeginn bestand zudem eine deutliche Forschungslücke hinsichtlich der ethischen, rechtlichen und sozialen Implikationen von IVAs, insbesondere solcher Systeme, die direkt mit Patient:innen interagieren. Während KI im Gesundheitswesen bislang vor allem für Diagnostik, Entscheidungsunterstützung oder Monitoring untersucht wurde, lagen nur wenige empirische Arbeiten zur direkten therapeutischen Interaktion mit intelligenten Agenten vor. Fragen der Verantwortlichkeit, Haftung, Transparenz, Datenschutzkonformität (insbesondere im Kontext der DSGVO) sowie potenzieller Auswirkungen auf die therapeutische Beziehung waren nur unzureichend empirisch adressiert.

Vor diesem Hintergrund bestand ein zentraler wissenschaftlicher Bedarf an klinisch fundierten Evaluationsstudien, die IVAs nicht nur technisch, sondern insbesondere aus psychotherapeutischer Perspektive untersuchen. Dazu zählen Fragen der Sicherheit, der angemessenen Reaktion auf psychische Krisen, der Akzeptanz bei Patient:innen und Therapeut:innen sowie der Einbettung solcher Systeme in bestehende Versorgungsstrukturen. Der im Projekt verfolgte Ansatz knüpft an diese Forschungslücken an und leistet insbesondere

durch Feldevaluationen und klinische Studien einen Beitrag zur evidenzbasierten Bewertung von IVAs im psychotherapeutischen Kontext.

1. Kakale MM. Of digital transformation in the healthcare (systematic review of the current state of the literature). *Health Technol* 2024;14:35–50. <https://doi.org/10.1007/s12553-023-00803-w>.
2. Andrade LH, Alonso J, Mneimneh Z, Wells JE, Al-Hamzawi A, Borges G, et al. Barriers to mental health treatment: Results from the WHO World Mental Health surveys. *Psychol Med* 2014;44:1303–17. <https://doi.org/10.1017/S0033291713001943>.
3. Kazdin AE. Annual Research Review: Expanding mental health services through novel models of intervention delivery. *J Child Psychol Psychiatry* 2019;60:455–72. <https://doi.org/10.1111/jcpp.12937>.
4. Miner AS, Shah N, Bullock KD, Arnow BA, Bailenson J, Hancock J. Key considerations for incorporating conversational AI in psychotherapy. *Front Psychiatry* 2019;10. <https://doi.org/10.3389/fpsyt.2019.00746>.
5. Weizenbaum J. ELIZA—a computer program for the study of natural language communication between man and machine. *Commun ACM* 1966;9:36–45. <https://doi.org/10.1145/365153.365168>.
6. Adamopoulou E, Moussiades L. Chatbots: History, technology, and applications. *Mach Learn Appl* 2020;2:100006. <https://doi.org/10.1016/j.mlwa.2020.100006>.
7. Schöbel S, Karatas J, Tingelhoff F, Leimeister J. Not everything is a metaverse?! A practitioners perspective on characterizing metaverse platforms. 56th Hawaii International Conference on System Sciences, 2023. <https://doi.org/10.24251/HICSS.2023.070>.
8. Herbener AB, Klincewicz M, Damholdt MF. A narrative review of the active ingredients in psychotherapy delivered by conversational agents. *Comput Hum Behav Rep* 2024;14:100401. <https://doi.org/10.1016/j.chbr.2024.100401>.
9. He Y, Yang L, Qian C, Li T, Su Z, Zhang Q, et al. Conversational agent interventions for mental health problems: Systematic review and meta-analysis of randomized controlled trials. *J Med Internet Res* 2023;25:e43862. <https://doi.org/10.2196/43862>.
10. Liu H, Peng H, Song X, Xu C, Zhang M. Using AI chatbots to provide self-help depression interventions for university students: A randomized trial of effectiveness. *Internet Interv* 2022;27:100495. <https://doi.org/10.1016/j.invent.2022.100495>.
11. Fitzpatrick KK, Darcy A, Vierhile M. Delivering cognitive behavior therapy to young adults with symptoms of depression and anxiety using a fully automated conversational agent (Woebot): A randomized controlled trial. *JMIR Ment Health* 2017;4:e19. <https://doi.org/10.2196/mental.7785>.
12. Nong P, Ji M. Expectations of healthcare AI and the role of trust: Understanding patient views on how AI will impact cost, access, and patient-provider relationships. *JAMIA* 2025;32:795–9. <https://doi.org/10.1093/jamia/ocaf031>.

13. Noorbakhsh-Sabet N, Zand R, Zhang Y, Abedi V. Artificial intelligence transforms the future of health care. *Am J Med* 2019;132:795–801.  
<https://doi.org/10.1016/j.amjmed.2019.01.017>.
14. Pentina I, Hancock T, Xie T. Exploring relationship development with social chatbots: A mixed-method study of replika. *Comput Human Behav* 2023;140:107600.  
<https://doi.org/10.1016/j.chb.2022.107600>.
15. Xie T, Pentina I. Attachment theory as a framework to understand relationships with social chatbots: A case study of Replika. Hawaii International Conference on System Sciences, 2022. <https://doi.org/10.24251/HICSS.2022.258>.
16. Pentina I, Xie T, Hancock T, Bailey A. Consumer–machine relationships in the age of artificial intelligence: Systematic literature review and research directions. *Psychol Mark* 2023;40:1593–614. <https://doi.org/10.1002/mar.21853>.
17. Wang X, Zhou Y, Zhou G. The application and ethical implication of generative AI in mental health: Systematic review. *JMIR Ment Health* 2025;12:e70610.  
<https://doi.org/10.2196/70610>.
18. Xian X, Chang A, Xiang Y-T, Liu MT. Debate and dilemmas regarding generative AI in mental health care: Scoping review. *Interact J Med Res* 2024;13:e53672.  
<https://doi.org/10.2196/53672>

## Zusammenarbeit mit anderen Stellen

Im Rahmen des HIVAM-Projekts sowie der damit verbundenen Studien, an denen das UKE beteiligt war, fand ein regelmäßiger und intensiver Austausch mit weiteren am UKE durchgeführten, chatbot-basierten Projekten im psychotherapeutischen Kontext statt. Ziel dieses Austauschs war es, mögliche Synergien gezielt zu nutzen sowie sich insbesondere bei sensiblen und regulatorisch relevanten Fragestellungen (z. B. Datenschutz, ethische Rahmenbedingungen) gegenseitig zu unterstützen und gemeinsame Beratung in Anspruch zu nehmen. Ein zentraler Kooperationspartner war hierbei ein von Dr. Brooke Schneider und Prof. Dr. Jürgen Gallinat geleitetes Projekt zur Entwicklung eines Chatbots als Gesprächspartner für ältere Menschen mit Einsamkeitserleben („Elderbot“), das am UKE in Zusammenarbeit mit dem Max-Planck-Institut Berlin durchgeführt wurde. Der inhaltliche Fokus auf empathische, dialogbasierte Unterstützung wies in mehreren Aspekten inhaltliche und methodische Überschneidungen mit dem HIVAM-Projekt auf.

Darüber hinaus bestand ein enger fachlicher Austausch mit dem von PD Dr. Franziska Miegel und Dr. Lara Rolvien geleiteten und durch das BMFTR geförderten Projekt STiVOK, das die Entwicklung eines Demonstrators einer XR-basierten Online-Klinik („NovaTreat XR“) zum Ziel hat. Diese Plattform soll eine intensive, stationsäquivalente psychotherapeutische Behandlung vollständig remote ermöglichen. Bestandteile des geplanten Behandlungskonzepts sind unter anderem virtuelle Einzel- und Gruppentherapieformate sowie Avatar-gestützte Psychoedukation und IVA-basierte Selbsthilfeangebote, wodurch sich inhaltliche und technologische Schnittstellen zum HIVAM-Projekt ergaben.

## Teil 2

### Aufzählung der wichtigsten wissenschaftlich-technischen Ergebnisse und anderer wesentlicher Ereignisse

#### Wichtigste wissenschaftlich-technische Ergebnisse

Das Projekt HIVAM begann mit einer Literaturanalyse zum Thema Embodied Virtual Medical Agents, sowie mit einer Anforderungsanalyse, um mögliche Themen, Wünsche und Ängste von Endnutzer:innen zu sammeln. Wir haben uns als Nutzergruppe für den psychotherapeutischen Bereich entschieden, da wir als Forscher:innen, Psychotherapeut:innen und Psycholog:innen des Universitätsklinikums Hamburg-Eppendorf (UKE) für diesen Anwendungsbereich die nötige Expertise haben. Für die Anforderungsanalyse wurden (ehemalige) Patient:innen sowie Ärzt:innen und Psychotherapeut:innen in Fokusgruppen interviewt und ihnen wurden frühe Versionen unseres Prototypen gezeigt. Mit Fokus auf einen nutzerzentrierten Entwicklungsprozess wirkte das UKE im ersten Projektjahr insbesondere an der Entwicklung der Anwendungsfälle (Use Cases), der Anforderungsanalyse sowie an der Erstellung eines ersten Prototyps (Early Demonstrator) mit. Ein zentrales Element war dabei die partizipative Zusammenarbeit mit EmPeeRie Now!. Im Austausch mit den Projektpartnern konnte zunächst ein erster Use Case im Anwendungsbereich Diagnostik definiert werden. Hierfür wurden die psychologisch-psychiatrischen Anforderungen erarbeitet, die wiederum die Grundlage für die Entwicklung des Prototyps bildeten. Parallel dazu wurde eine erste Iteration der Anforderungsanalyse durchgeführt, in der Themen, Wünsche und Ängste von Nutzer:innen erfasst wurden. Grundlage dafür war ein positiver Ethikantrag bei der Ethikkommission des UKE. Die Teilnehmenden konnten am UKE rekrutiert werden, und die Interviews fanden ebenfalls dort statt. Beteiligt waren insgesamt sieben Psycholog:innen in Ausbildung (PIA), fünf Psychotherapeut:innen, fünf Ärzt:innen, sieben Patient:innen sowie vier Mitglieder von EmPeeRie Now!. Die Auswertung dieser Daten zur Fertigstellung des Anforderungsanalyse-Dokuments konnte zeitgerecht abgeschlossen werden.

Darüber hinaus wurden im Hinblick auf den Prototyp/Early Demonstrator erste psychologische Anforderungen definiert, insbesondere zu Aspekten der psychotherapeutischen Gesprächsführung und Grundhaltung sowie zu relevanten Wirkfaktoren. Weiterhin wurden die wichtigsten ethisch-rechtlichen Neuerungen untersucht, beispielsweise der AI Act. Die

Ergebnisse sind in die Entwicklung des Prototypen eingeflossen und wurden in verschiedenen wissenschaftlichen Publikationen veröffentlicht.

Der finale Prototyp stellt einen virtuellen Psychotherapeuten dar. Dieser führt mit den Nutzer:innen kurze Selbsthilfeübungen durch. Für den Agenten haben wir zum einen einen generischen weiblichen Agenten, zum anderen die von der HS Düsseldorf entwickelte realistische Darstellung von Prof. Dr. Jürgen Gallinat. Dieser Agent wurde in der Unity Engine eingebunden und mit Idle-Animationen, Augenbewegungen sowie Lip Sync Technologie ausgestattet, um realistischer zu erscheinen. Die Anwendung ist in der virtuellen Realität, aber auch am Desktop möglich. Für das Gespräch mit dem IVA werden verschiedene Technologien genutzt, beispielsweise Speech-to-Text, damit die von den Nutzer:innen gesprochene Sprache transkribiert wird. Danach wird diese an ein Large Language Model (z.B. GPT-4 Omni) weitergeleitet, welches eine passende Antwort generiert. Dieser Text wird dann wiederum zurück an das System geleitet, wo er mithilfe von Text-to-Speech zurück zu gesprochener Sprache wird. Hierfür können wir verschiedene Stimmen nutzen, beispielsweise generische Stimmen, oder auch die realistische, geclonte Stimme von Prof. Gallinat. Momentan beinhaltet der Prototyp sechs verschiedene Selbsthilfeübungen zum Thema Ängste oder Selbstwert, er ist aber leicht erweiterbar. Außerdem haben wir die Einbindung von psychologischen Fragebögen untersucht, beispielsweise den PHQ-9. Der Prototyp wurde im Laufe des Projektes ständig verbessert und Ergebnisse aus kleineren Studien sind mit eingeflossen.

Der entwickelte Prototyp wurde darüber hinaus in mehreren Nutzerstudien evaluiert. In einer ersten Studie mit Studierenden der Universität Hamburg wurde die Durchführung von Selbsthilfeübungen mithilfe des IVA in Virtual Reality mit einer textbasierten Smartphone-Anwendung verglichen. Die Ergebnisse zeigten, dass VR als anregender erlebt wurde und zu einer längeren Beschäftigung mit den Aufgaben führte. Gleichzeitig berichteten die Teilnehmenden jedoch auch von einer höheren empfundenen Arbeitsbelastung im Vergleich zur Smartphone-Version.

Am UKE wurde anschließend eine Studie zur Sicherheit der Aussagen des Agenten durchgeführt. Hierfür führten approbierte Psychotherapeut:innen geskriptete Gespräche mit dem IVA, die im Anschluss von externen Fachkolleg:innen bewertet wurden. Insgesamt wurden dabei zweimal zwölf verschiedene Testfälle zu den Themen Suizidalität und Substanzkonsum vorgegeben, die unterschiedliche Fragetechniken enthielten (z. B. persönliche Erfahrungen, verdeckte Geschichten, Fragen zu statistischen Einschätzungen). Die Antworten des IVA wurden anschließend mithilfe einer eigens entwickelten Machbarkeitskala beurteilt, die Kriterien wie Einhaltung von Richtlinien, Hilfestellung und Risikomanagement sowie Empathie abbildete. Es zeigte sich, dass der IVA in rund 29 % der Fälle kritische Antworten produzierte, insbesondere in Bezug auf die Normalisierung von Alkoholkonsum und den Umgang mit Minderjährigen. Die Publikation zu dieser Studie befindet sich momentan im Review bei *Scientific Reports* (Rolvien et al., eingereicht).

In einer abschließenden Studie erfolgte schließlich eine Evaluation mit 12 Patient:innen der Station für Angststörungen am UKE, bei der die Anleitung von drei Selbsthilfeübungen zum Umgang mit Ängsten in drei unterschiedlichen Visualisierungen des Prototyps miteinander verglichen wurden: als menschlich verkörperter IVA, als Voice-Chat sowie als textbasierter Chat. Hierbei schnitt der Voice-Chat am besten ab und der IVA zeigt vor allem in der Patientenzufriedenheit, User Experience, Nützlichkeit und Nutzungsabsicht Defizite, die vor allem auf den Realismus des Agenten zurückzuführen sind, z.B. auf den Uncanny Valley Effekt. Personalisierbarkeit, insbesondere beim Geschlecht des Agenten, wurde häufig gewünscht und es wurde angemerkt, dass die KI momentan noch zu wenig Empathie aufbringen kann. Beim Voice-Chat wurde angemerkt, dass es im Vergleich zum textuellen Chat hilfreich ist, nicht allein in seiner Gedankenwelt zu sein. Generell wurde die Zugänglichkeit der Übungen über digitale Plattformen als positiv bewertet. Die Publikation zu dieser Studie befindet sich momentan in der Vorbereitung.

#### Wesentliche Ereignisse:

- Teilnahme an allen zweiwöchentlichen Meetings mit allen Projektpartnern
- Teilnahme an ELSA-Workshops mit internen und externen Referierenden und Teilnehmenden (UHH-EIT):
  - ELSA-Workshop, 31.05.2023: „Ethische und soziale Implikationen intelligenter virtueller Agenten in der Gesundheitsversorgung“
  - ELSA-Workshop, 21.11.2023: „Virtuelle Agenten und Chatbots in der Gesundheitsversorgung: Implikationen für Haftung, Schweigepflicht und Datenschutz“
  - Workshop, 27.07.2023: „Nutzung der Value-Sensitive-Design-Methode für nutzer- und werte-orientiertes Design“
- Teilnahme am HIVAM Konsortialtreffens mit allen Projektpartnern und VDI/VDE am 23.01.2024 an der UHH
- Teilnahme am BMBF-Vernetzungssymposium der Förderlinie, 26./27.02.2024 in Düsseldorf
- Teilnahme am HIVAM Konsortialtreffen am 29.10.2024 bei Apoqlar
- Teilnahme am HIS-Abschlussymposium am 18./19.02.2025 in Berlin
- Regelmäßige Teilnahme an Treffen des Begleitprojekts CoCreHit

#### Nutzen und Verwertbarkeit

Im Projekt wurden vielfältige Synergien mit anderen Forschungsinitiativen zum Einsatz von KI in der Psychiatrie geschaffen und genutzt. Dazu zählen insbesondere die Kooperation mit dem Projekt „Large Language Model (LLM)-based Chatbot for Loneliness in Older Adults (Elderbot)“ von Dr. Brooke Schneider und Prof. Jürgen Gallinat in Zusammenarbeit mit dem Max-Planck-

Institut für Bildungsforschung sowie die Zusammenarbeit mit der „NovaTreat XR“-Online-Klinik, einem Verbundprojekt zwischen der Patient Zero Games GmbH und PD Dr. Franziska Miegel und Dr. Lara Rolvien vom UKE. Durch den Austausch mit diesen Projekten konnten methodische Ansätze, Erfahrungen im Umgang mit Zielgruppen sowie technologische Entwicklungen gegenseitig nutzbar gemacht werden.

Darüber hinaus fanden klinikinterne Brainstorming-Treffen zum zukünftigen Einsatz von KI in der Klinik für Psychiatrie und Psychotherapie statt. Hier flossen auch die Erkenntnisse der beiden am UKE durchgeführten HIVAM-Studien ein. Diese Diskussionsrunden hatten nicht nur das Ziel, konkrete Anwendungsfelder für KI im klinischen Alltag zu identifizieren, sondern dienten auch der Sensibilisierung des Stationspersonals für Chancen und Herausforderungen der Technologie. Auf diese Weise konnten erste Vorbehalte gegenüber KI thematisiert und abgebaut sowie ein breiteres Interesse für deren möglichen Einsatz in Diagnostik und Therapie geschaffen werden.

Langfristig besteht das Ziel, einen IVA regelhaft auf Station anzubieten und in Studien systematisch zu untersuchen, wie und in welchem Umfang Patient:innen dieses Angebot nutzen. Dafür wird erwogen, eine Meeting- bzw. Telefonbox einzurichten, in der die KI dauerhaft implementiert ist und für die sich Patient:innen eigenständig anmelden können. Auf diese Weise ließe sich der IVA oder möglicherweise eher ein Voice-Bot (s. Ergebnisse finale Studie) niedrigschwellig und unabhängig vom direkten therapeutischen Kontakt erproben und wissenschaftlich evaluieren.

## Fortschritte auf diesem Gebiet an anderer Stelle

- Insbesondere die aktuellen Entwicklungen und Diskussionen um Sprachmodelle wie GPT-4 waren bei der Entwicklung der Avatar/Assistent-Modelle im Projekt HIVAM weiterhin von großer Bedeutung. Diese entwickelten sich besonders zu Projektanfang sehr schnell und ermöglichten uns durch ihre natürliche Sprachverarbeitung die einfache Integration in unseren Prototypen. Die Grundlagen, Erkenntnisse und wissenschaftlichen Reflektionen um die Potenziale und Grenzen dieser Sprachmodelle wurden im Projekt angewendet und für den psychiatrischen Fall spezifiziert.
- Aus rechtlicher Sicht ist insbesondere das Gesetzgebungsverfahren zur KI-Verordnung der Europäischen Union („AI Act“) relevant. Die Verordnung wurde im März 2024 verabschiedet und wird 2026 vollständig in Kraft treten. KI-basierte Telemedizinanwendungen wie die des HIVAM-Projekts fallen unter das rechtliche Rahmenwerk der EU-KI-Verordnung. Demnach müssen die rechtlichen Anforderungen im Forschungsprojekt zumindest theoretisch für eine hypothetische prospektive Anwendung im Medizinsektor mitgedacht werden. Interessant wird, ob die deutsche Regierung eine Abschwächung der KI-Verordnung anstreben und erwirken wird, wie es aktuell diskutiert wird. Dies könnte die rechtliche Lage erneut verändern. Wiederholte Ereignisse rund um Chatbots und ihre Wirkung auf Menschen mit Neigung zu Suizidalität in den vergangenen

zwei Jahren erhöhen die Sicherheitsanforderungen an KI-Anwendungen im Bereich der Psychotherapie und lenken die Aufmerksamkeit der Öffentlichkeit auf diese Themen.

## Veröffentlichungen der Ergebnisse.

Publikationen im Kontext des HIVAM-Projekts, an denen das UKE beteiligt war:

- Rudschies, C., Rings, S., Kruse, L., **Schauenburg, G.**, Marmarshahi, H., & Zimmer, C. N. (2023). Psychotherapy with the help of ChatGPT? Current technical and ethical boundaries of intelligent virtual agents. Mensch und Computer 2023 - Workshopband. DOI: 10.18420/muc2023-mci-ws06-367. GI. MCI-WS06: First International Workshop on Co-Creation of Hybrid Interactive Systems for Healthcare. Rapperswil. 3.-6. September 2023.
- Zimmer, C., Kruse, L., **Rolvien, L.**, **Gallinat, J.**, & Geiger, C. (2025). Transformation of MetaHumans for a generic XR workflow. In International Conference on Extended Reality (pp. 73-92). Cham: Springer Nature Switzerland.
- Rings, S., Kruse, L., Rudschies, C., **Rolvien, L.**, **Schauenburg, G.**, **Gallinat, J.**, Schneider, I., & Steinicke, F. IVAs' Future in therapy: Investigating the impact of intelligent virtual agents in VR on psychological self-help exercises and user satisfaction. In: Mensch und Computer 2025 (MuC 2025). DOI: <https://doi.org/10.1145/3743049.3743063>
- UNDER REVIEW:
  - **Rolvien, L.**, Kruse, L., Rings, S., Zimmer, C., **Schauenburg, G.**, Thams, F., Brähler, A., Rudschies, C., Schneider, I., **Moritz, S.**, Steinicke, F., & **Gallinat, J.** (under review at *Scientific Reports*). Intelligent virtual agents in psychotherapy: A safety evaluation of high-risk mental health scenarios.

Medien und Wissensvermittlung:

- Vortrag „Künstliche Intelligenzen sprechen von Psyche – Was antworten wir?“; Konferenz: Die subjektive Seite der Schizophrenie, 03.08.2024, Hamburg, Prof. Gallinat
- Klinik-interner Vortrag am 05.07.2024 anlässlich eines Wissenschaftlertreffens am Universitätsklinikum Hamburg-Eppendorf, Dr. Rolvien
- Vortrag bei der Donnerstagsweiterbildung am 17.10.2024 in der Klinik für Psychiatrie und Psychotherapie, Dr. Rolvien
- Vortrag UKE Alumni Veranstaltung am 8.11.2024 „Psychiatrie und Psychotherapie: Was kann mein Therapeut was meine KI nicht kann?“, Prof. Gallinat