Edilife

Ethik, Recht und Sicherheit des digitalen Weiterlebens

Teilvorhaben: Datenschutz und Datensicherheit beim digitalen Weiterleben

Sachbericht zum Verwendungsnachweis

Teil II: Eingehende Darstellung

Fraunhofer-Institut für Sichere Informationstechnologie (SIT)

Zuwendungsempfänger:	Fraunhofer-Institut für Sichere Informations-technologie (SIT)
Förderkennzeichen:	16INS114B
Titel des Teilvorhabens:	Datenschutz und Datensicherheit beim digitalen Weiterleben
Projektleitung:	Ulrich Waldmann
Autoren	Ines Geissler, Thomas Kunz, Annika Selzer, Ulrich Waldmann
Laufzeit des Vorhabens:	01.07.2022 - 31.03.2024
Projektpartner:	Universität Tübingen: Internationales Zentrum für Ethik in den Wissenschaften (IZEW)

Die in diesem Bericht enthaltenen Arbeitsergebnisse sind sorgfältig und unter Zugrundelegung des bekannten Standes der Wissenschaft erstellt worden, stellen jedoch Forschungsansätze dar. Eine Haftung oder Garantie dafür, dass die Arbeitsergebnisse/Informationen die Vorgaben der aktuellen Rechtslage erfüllen, wird aus diesem Grund nicht übernommen. Gleiches gilt für die Brauchbarkeit, Vollständigkeit oder Fehlerfreiheit, so dass jede Haftung für Schäden ausgeschlossen wird, die aus der Benutzung dieser Arbeitsergebnisse/Informationen entstehen können. Diese Haftungsbeschränkung gilt nicht in Fällen von Vorsatz.

Inhalt

1 Zusammenfassung der Projektziele 4

2 Erzielte Ergebnisse 4

2.1 Theoretische Rahmung, Identifizierung relevanter Fragestellungen 5

2.2 Aktuelle und zukünftige Technikentwicklung 5

2.2.1 Umsetzungsmöglichkeiten für Anwendungen des digitalen Weiterlebens 5

2.2.2 Technische Herausforderungen 7

2.2.3 Zukünftige Technikentwicklung und ausgewählte Konzepte zum Schutz von Avataren 8

2.3 Datenschutz und Persönlichkeitsrecht 10

3 Wichtigste Positionen des zahlenmäßigen Nachweises 13

4 Notwendigkeit und Angemessenheit der geleisteten Arbeit 14

5 Darstellung des Verwertungsplans 14

6 Darstellung des Fortschritts auf dem Gebiet des Vorhabens 15

7 Darstellung der erfolgten oder geplanten Veröffentlichungen 15

8 Literatur 16

1 Zusammenfassung der Projektziele

Das Thema „digitales Weiterleben“ berührt existenzielle Fragen, die potenziell jeden und jede in der digitalen Gesellschaft betreffen. Schon heute umfasst der digitale Nachlass unterschiedliche interaktive Inhalte (z. B. Blogs, Avatare, Soziale Medien), die Verstorbene ihrer Nachwelt hinterlassen. Angesichts der zu erwartenden Technikentwicklung werden diese Veränderungen eine immer größere Bedeutung erhalten. Neue Technologien, welche eine sinnbildliche Unsterblichkeit anbieten, stoßen auf großes Interesse und betreffen Fragen, die bisher in Religion und Weltanschauung verankert waren.

Dank des technologischen Fortschritts insbesondere in der KI und einer zunehmenden Kommerzialisierung der DAI (Digital Afterlife Industry) werden Avatare auf der Basis von echtem, zu Lebzeiten aufgenommenem Material, immer authentischer (Ajder et al., 2019). Zugleich aber ist jede Erschaffung eines Avatars wie bei jedem Medium mit Fragen der Auswahl und Kombination nur bestimmter Aspekte bzw. der Inszenierung einer Identität verbunden. Die Anstrengung, ein „ewiges“ digitales Abbild von sich zu schaffen, kann mit den Herausforderungen verglichen werden, die auch für die Lebenden durch ihre Beteiligung an den Sozialen Medien gegeben ist (Schmidt & Taddicken, 2017). Dazu gehört, ein möglichst optimales Bild von sich und seinem sozialen Beziehungsnetzwerk zu zeichnen. Diese Formen der Selbstinszenierung sind Gegenstand zahlreicher kritischer Betrachtungen, die sich vor allem mit Fragen des Konformitätsdrucks und der Erfüllung von Erwartungen an Attraktivität und Erfolg auseinandersetzen (Weissman, 2021; Heesen, 2022).

Wenn durch KI bald ein digitaler Zwilling erzeugt werden kann, der visuell, charakterlich, emotional und intellektuell einer/einem Verstorbenen gleicht, entstehen insbesondere neue datenschutzrechtliche und (sicherheits-)technische Fragestellungen, die durch das Teilvorhaben „Datenschutz und Datensicherheit beim digitalen Weiterleben“ im Rahmen des Projekts Edilife untersucht wurden, u. a.: Welche datenschutzrechtlichen und sicherheitstechnischen Implikationen hat der Wunsch nach dem virtuellen aktiven Weiterleben, z. B. in Form eines digitalen Avatars? Inwiefern besteht die Notwendigkeit, das postmortale Datenschutzrecht für zukünftige Formen des digitalen Weiterlebens in Form eines Avatars zu regeln bzw. zu konkretisieren. Wie können die Datenschutzrechte der mit dem Avatar interagierenden Personen gewährleistet werden? Wie kann eine gezielte Manipulation von KI-Eingaben und -Ausgaben, die einen erheblichen Einfluss auf die elektronische Repräsentation von natürlichen Personen haben kann und zu einer verfälschten Darstellung der repräsentierten und identifizierbaren Personen führen kann, verhindert werden?

Das Teilvorhaben hatte das Ziel, Lösungsvorschläge für diese und weitere Fragen zu erarbeiten und zukünftige Entwicklungen zu antizipieren. Damit sollten Fehlentwicklungen entgegengesteuert und sollten Handlungsbedarfe in Politik und Gesellschaft gekennzeichnet werden. Zukünftige Lösungsoptionen wurden skizziert und auf die Anforderungen der IT-Sicherheit und des Datenschutzes hin überprüft.

2 Erzielte Ergebnisse

Das folgende Kapitel stellt die von Fraunhofer SIT erzielten Projektergebnisse vor.

2.1 Theoretische Rahmung, Identifizierung relevanter Fragestellungen

Fraunhofer SIT hat sicherheitstechnische Fragestellungen hinsichtlich der Nutzung personenbezogener KI-Lösungen und Avataren des digitalen Weiterlebens identifiziert und bzgl. ihrer Relevanz für Anwendungen des digitalen Weiterlebens analysiert. Hierzu zählen unter anderem die Sicherstellung der Integrität und Authentizität der für das Training des KI-Modells benötigten Datenquellen, sowie den Schutz der Integrität von Avataren für das digitale Weiterleben während der Nutzungsphase. Weitere Fragestellungen betreffen die Identifizierung und Authentifizierung von Avataren und den Nutzer:innen von Anwendungen des digitalen Weiterlebens, die Umsetzung vertrauenswürdiger Herkunftsnachweise von Avataren, die Umsetzung von Nutzungsrechten an Avataren, das Erkennen gefälschter oder manipulierter Avatare sowie die sichere Löschung von Avataren am Ende ihrer Nutzungsphase.

In rechtlicher Hinsicht hat Fraunhofer SIT die datenschutzrechtlichen Rollen der Akteure im digitalen Weiterleben untersucht. Dabei sind vier Rollen aus datenschutzrechtlicher Sicht besonders relevant: die der betroffenen noch lebenden Person, die Ausgangspunkt des weiterlebenden Avatars ist (1), die des weiterlebenden Avatars (2), die der mit den Avataren agierenden betroffenen Personen wie beispielsweise Verwandte (3) sowie mögliche datenschutzrechtliche Verantwortliche und Auftragsverarbeiter, die Dienstleistungen im Bereich des digitalen Lebens anbieten (4). Bei den lebenden Personen, die entweder Ausgangspunkt eines weiterlebenden Avatars sind oder mit den weiterlebenden Avataren agieren (1 und 3), handelt es sich um betroffene Personen im Sinne des Datenschutzrechts. Der weiterlebende Avatar (2) ist keine natürliche Person und somit keine betroffene Person. In dem Zusammenhang ergeben sich Fragestellungen, die in AP 4 berücksichtigt werden. Bei Organisationen, die Dienstleistungen im Bereich des digitalen Weiterlebens anbieten (4) kommt es darauf an, ob sie über den Zweck und die Mittel der personenbezogenen Datenverarbeitung entscheiden. Liegt diese Entscheidung bei der betreffenden Organisation, ist sie Verantwortlicher. Setzt sie nur einen Teilbereich der Verarbeitung personenbezogener Daten um und liegt die Entscheidung über den Zweck und die Mittel der Verarbeitung bei einer anderen den Teilbereich der Verarbeitung beauftragenden Organisation, ist sie Auftragsverarbeiter.

2.2 Aktuelle und zukünftige Technikentwicklung

2.2.1 Umsetzungsmöglichkeiten für Anwendungen des digitalen Weiterlebens

Fraunhofer SIT hat die Umsetzungsmöglichkeiten für Anwendungen des digitalen Weiterlebens aus technischer Sicht analysiert und hierbei besondere technische Herausforderungen identifiziert. Bei Anwendungen des digitalen Weiterlebens sollen Nutzer:innen mit digitalen Repräsentanten von verstorbenen Personen (den „Avataren“) kommunizieren können, also beispielsweise Informationen aus dem Leben der verstorbenen Person erfragen können oder komplexe Unterhaltungen mit dem Avatar führen können. Die Umsetzungsmöglichkeiten reichen von einfachen Frage-Antwort-Anwendungen, bei denen Nutzer:innen aus vorgegebenen Fragen auswählen können und vorgefertigte Antworten bekommen, über Chatbots, bei denen die Nutzer:innen entweder durch Eingabe von Freitext oder sprachbasiert Unterhaltungen mit der Computeranwendung führen können, bis hin zu komplexen Virtual Reality (VR)- oder Augmented Reality (AR)-Anwendungen. Während bei Chatbot-Anwendungen die visuelle Darstellung in der Regel entweder gar nicht erfolgt oder nur sehr vereinfacht ist (z. B. Anzeige eines Fotos der verstorbenen Person), wird in VR-Anwendungen

der Avatar in einer virtuellen dreidimensionalen Umgebung dargestellt, in der er sich bewegen und mit der er interagieren kann. Die Nutzer:innen selbst tauchen in diese virtuelle Welt ein, werden dort ebenfalls als Avatar repräsentiert und können über ihren Avatar mit dem Avatar der verstorbenen Person kommunizieren. Bei AR-Anwendungen verwenden die Nutzer:innen sogenannte Smartglasses, über die der Avatar der verstorbenen Person in die reale Umgebung der Nutzer:innen eingeblendet wird, so dass sie mit dem Avatar kommunizieren können. Zukünftig sind zudem als Weiterentwicklung von VR- und AR-Anwendungen auch Anwendungen des digitalen Weiterlebens in Metaversen denkbar (vgl. Abbildung 1).

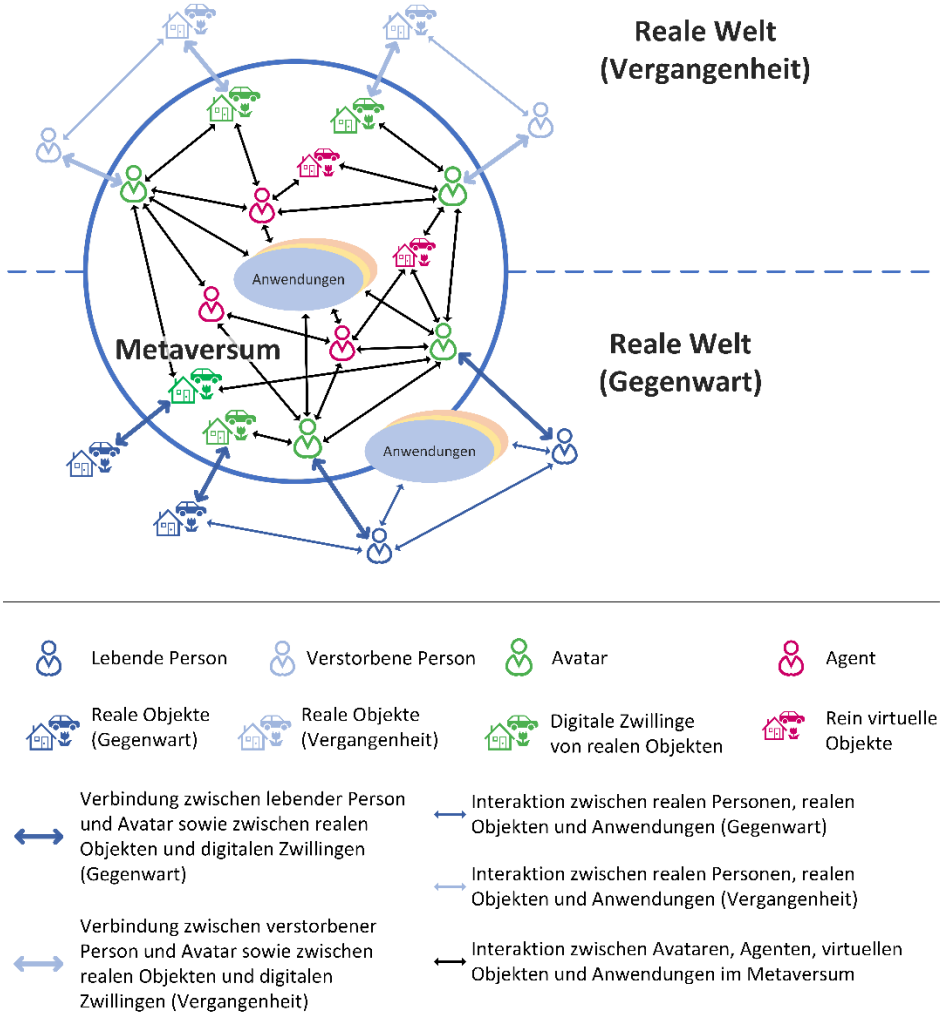

Abbildung 1: Anwendungen des digitalen Weiterlebens in Metaversen

Da die Gestaltung der Avatare in den jeweiligen Umsetzungsmöglichkeiten sehr stark variieren kann, hat Fraunhofer SIT verschiedene Ausprägungsformen von Avataren definiert. Hierbei wird zwischen der äußeren Gestaltung und der inhaltlichen Gestaltung unterschieden. Bei der äußeren Gestaltung geht es um die visuelle Darstellung des Avatars. Bei Anwendungen des digitalen Weiterlebens ist davon auszugehen, dass es das Ziel ist, eine verstorbene Person möglichst realistisch darzustellen, auch wenn es sich bei Avataren im Allgemeinen auch um reine Kunstfiguren handeln kann. Bei der inhaltlichen Gestaltung geht es um das Wissen des Avatars, die Imitation von Gedanken der verstorbenen Person, sich ähnlich ausdrücken zu können wie die verstorbene Person sowie die Fähigkeit, Emotionen ausdrücken zu können.

2.2.2 Technische Herausforderungen

Insbesondere die inhaltliche Gestaltung von Avataren stellt einige besondere technische Herausforderungen bereit. Aktuelle KI-gestützte Chat-Anwendungen wie ChatGPT zeigen, dass die Umsetzung solcher Anwendungen sehr aufwändig und auch sehr kostenintensiv ist. Soll der Avatar zudem möglichst realistisch hinsichtlich des Wissens und der Ausdrucksformen die verstorbene Person imitieren, besteht in vielen Fällen die Herausforderung darin, überhaupt ausreichend Trainingsdaten über die verstorbene Person zu sammeln. Eine grundsätzliche Herausforderung bei der Entwicklung KI-basierter Sprachmodelle besteht zudem darin, zu verhindern, dass die Anwendung Fakten „erfindet“ oder „toxische“ Antworten gibt (z. B. rassistische oder sexistische Äußerungen). Eine weitere Herausforderung besteht darin, KI-basierte Sprachmodelle zu entwickeln, die kontinuierlich während der Nutzungsphase weiter lernen, ohne hierbei durch gezielte Eingaben der Nutzer:innen in eine bestimmte Richtung hin manipuliert werden zu können, so dass der Avatar bzgl. der inhaltlichen Gestaltung irgendwann gar nicht mehr die verstorbene Person imitiert, sondern eher der Nutzer:in ähnelt.

Bei der äußeren Gestaltung entstehen besondere technische Herausforderungen insbesondere dann, wenn der Avatar in virtuellen Umgebungen dargestellt wird oder mittels Augmented Reality in die reale Umgebung eingeblendet werden soll. Heutige Anwendungen stoßen schnell an ihre Grenzen, wenn viele Avatare in Echtzeit gerendert oder soziale Interaktionen abgebildet werden sollen. So sind beispielsweise die Latenzzeiten für die Erwiderung des direkten Blicks eines Kommunikationspartners entscheidend für die soziale Bedeutung des Blickkontakts, ebenso die Latenzzeiten für die erneute Unterbrechung des Blickkontakts. Erwidert der Avatar den Blick der Nutzer:innen zu spät oder dauert der Blick zu lange, wird das von den Nutzer:innen als Desinteresse bzw. als unangemessenes Anstarren gewertet. Die Auswirkungen von Gesichts- oder Bewegungsmimik auf reale Personen sind ebenfalls zeitabhängig. So wird beispielsweise die Spiegelung der Körperhaltung nur dann als relevantes soziales Signal wahrgenommen, wenn eine gewisse Zeitverzögerung vorliegt.

Von den verschiedenen Umsetzungsmöglichkeiten für Anwendungen des digitalen Weiterlebens und den damit verbundenen technischen Herausforderungen hat Fraunhofer SIT im Anschluss typische Angriffe und Bedrohungen abgeleitet. Beispielsweise kann ein Angreifer versuchen, während der Entwicklungsphase des Avatars Zugriff auf personenbezogene Daten, die für das Training verwendet werden, zu erlangen, oder er kann versuchen, die Trainingsdaten oder das KI-Modell selbst zu manipulieren. Hierdurch kann beispielsweise das Sprachmodell des Avatars entweder unbrauchbar werden oder auf bestimmte Eingaben mit „toxischen“ Antworten reagieren.

Während der Nutzungsphase könnte ein Angreifer versuchen, die Antworten eines Avatars zu manipulieren, um auf diese Weise Falschinformationen zu verbreiten oder andere Personen zu diskreditieren. Ein Angreifer kann darüber hinaus versuchen, die äußere Gestaltung eines Avatars zu manipulieren (bspw. veränderte Stimme, unheimliches Erscheinungsbild, was bei den Nutzer:innen einen sogenannten „Uncanny Valley Effekt“ auslösen kann). Bei VR-Anwendungen besteht die Gefahr, dass Angreifer unbemerkt in die virtuelle Welt eindringen (beispielsweise mittels eines „unsichtbaren“ Avatars), um die Kommunikation zwischen Avataren und Nutzer:innen abzuhören oder die virtuelle Welt zu manipulieren.

Alternativ kann ein Angreifer auch versuchen, Zugriff auf das VR-Headset der Nutzer:innen zu erlangen, um auf diese Weise unbemerkt Daten zu extrahieren (z. B. die Videodaten der

Kamera). Im Falle von AR-Anwendungen kann ein Angreifer hierdurch auch Informationen über die reale Umgebung der Nutzer:innen extrahieren, die von den verschiedenen Sensoren erfasst werden (Kameras, Mikrofone, GPS usw.). Hierzu zählen auch personenbezogene Daten von unbeteiligten Personen, die sich in der Nähe der Nutzer:innen aufhalten.

2.2.3 Zukünftige Technikentwicklung und ausgewählte Konzepte zum Schutz von Avataren

Im Rahmen der aktuellen und zukünftigen Technikentwicklung beschäftigte sich Fraunhofer SIT schwerpunktmäßig mit folgenden Fragestellungen:

1. Welches Potential bieten KI-basierte Sprachmodelle für die Entwicklung von möglichst realistisch erscheinenden Avataren des digitalen Weiterlebens?
2. Inwieweit können in zukünftigen VR-basierten Anwendungen des digitalen Weiterlebens aktuelle Konzepte für dezentrale Sicherheitsmechanismen genutzt werden, um beispielsweise fälschungssichere Identitäten von Avataren und deren Beziehungen zu der repräsentierten Person abzubilden oder den Besitz und Rechte an Avataren nachzuweisen.

Zu den weiteren Themen, mit denen sich Fraunhofer SIT befasste, gehören unter anderem die Authentizität und Integrität der Trainingsdaten von KI-basierten Anwendungen, Herausforderungen beim Schutz der Privatheit bei Anwendungen des digitalen Weiterlebens, die sichere Löschung von Avataren sowie Herausforderungen bzgl. der realistischen Darstellung von Avataren (z. B. Erkennen und Imitieren von sozialen Interaktionen, Nutzung von Deepfake-Technologien).

KI-basierte Sprachmodelle

Die derzeitigen Angebote von Anwendungen des digitalen Weiterlebens nutzen für die inhaltliche Gestaltung in der Regel Wissensdatenbanken, die während der Entwicklung des Avatars manuell durch die Entwickler erstellt werden, indem sie zum Beispiel die zu repräsentierende Person interviewen. Diese Wissensdatenbanken enthalten vorgefertigte und unter Umständen bereits ausformulierte Antworten, die der Avatar während eines Gesprächs mit der anwendenden Person abrufen kann. Aktuell sind jedoch insbesondere auf dem Gebiet der generativen KI und bei der Entwicklung großer Sprachmodelle enorme Fortschritte zu beobachten. ML-basierte Sprachmodelle haben das Potential, die inhaltliche Gestaltung der Avatare enorm zu erweitern. Sie ermöglichen es, nahezu beliebige Gespräche mit einem Avatar zu führen. Die Texte, die ein solcher Avatar generiert, sind in der Regel kaum noch von durch Menschen verfasste Texte zu unterscheiden. Es ist zu erwarten, dass zukünftige Anwendungen des digitalen Weiterlebens sich diese Entwicklungen zunutze machen und in ihre Angebote integrieren. Aus diesem Grund bildeten KI-basierte Sprachmodelle einen Schwerpunkt der Untersuchungen des Fraunhofer SIT.

Die Untersuchungen haben gezeigt, dass es noch einige Herausforderungen zu bewältigen sind, wenn KI-basierte Sprachmodelle in Anwendungen des digitalen Weiterlebens genutzt werden sollen. Eine Hürde bilden derzeit die hohen Entwicklungskosten und die große Menge benötigter Trainingsdaten. Um die Entwicklungskosten zu senken, ist es unerlässlich, Sprachmodelle für möglichst viele Avatare weiterzuverwenden und nur gezielt an spezifische Eigenschaften eines speziellen Avatars anzupassen. Die Qualität von Sprachmodellen könnte

zukünftig durch sogenanntes bestärkendes Lernen (Reinforcement Learning from Human Feedback) verbessert werden. Hierbei wird durch die repräsentierte Person ein Belohnungsmodell trainiert, das dann wiederum als Basis für ein weitergehendes, selbstständiges Training des Avatars dient. Voraussetzung hierfür ist allerdings, dass die Entwicklung eines solchen Belohnungsmodells einfach ist, so dass sie auch von technischen Laien bewältigt werden kann. Darüber hinaus ist die Anpassung von Sprachmodellen an soziale Kontexte für eine natürliche und angemessene Kommunikation wichtig. Hierbei muss zwischen einer möglichst authentischen Darstellung der repräsentierten Person und einer hohen Dialogsicherheit abgewogen werden. Eine hohe Dialogsicherheit könnte bedeuten, dass sich die Avatare immer mehr ähneln und den anwendenden Personen zu sachlich, weich und politisch korrekt erscheinen - im Widerspruch zu den persönlichen Erinnerungen an die repräsentierte Person. Zudem ist es wichtig, Beziehungen zwischen verschiedenen anwendenden Personen und dem Avatar (bzw. der repräsentierten Person) sowie Änderungen bei diesen Beziehungen durch das Sprachmodell zu berücksichtigen. Eine weitere Schwierigkeit bei aktuellen Sprachmodellen besteht darin, ihnen eine Art „Zeitbewusstsein“ anzutrainieren, damit beispielsweise die Ereignisse im Leben der repräsentierten Person und aktuelle Ereignisse in der korrekten Reihenfolge berücksichtigt werden.

Interoperable, dezentrale Sicherheitsmechanismen

Mit der Integration in Internet-basierten VR- und AR-Umgebungen sind die Avatare auch von deren IT-Sicherheitsrisiken betroffen. Repräsentierte und anwendende Personen in virtuellen Welten können beispielsweise durch Identitätsdiebstahl, Manipulation von VR-Ausgaben und Angriffen auf Sensor- und Kommunikationsdaten bedroht sein. Der dynamische Wechsel eines bestimmten Avatars zwischen verschiedenen virtuellen Anwendungen setzt voraus, dass die beteiligten Sicherheitsmechanismen interoperabel sind und ohne nennenswerte Latenzzeiten synchronisiert werden können. Anwendungsübergreifende Synchronisation und Interoperabilität sind bisher in der Regel nicht gegeben. Die heutigen Avatare sind meist anwendungsspezifisch, d. h. können nicht exportiert und in andere Anwendungen integriert werden. Entsprechend sind auch die implementierten Sicherheitsmechanismen meist auf die jeweilige Plattform des jeweiligen Anbieters begrenzt.

Wenn in Zukunft viele virtuelle Welten miteinander vernetzt sind und über Metaversen ein offenes Gesamtsystem bilden, das auch mit der physischen Welt verbunden ist, dann erfordert dies interoperable digitale Identitäten sowohl für reale Objekte und Personen als auch für virtuelle Objekte und Avatare, um über die Grenzen der einzelnen virtuellen Welten hinweg nutzbar zu sein. Avatare sollten überprüfbar authentisch und integer sein, insbesondere wenn befürchtet werden muss, dass es zu einer repräsentierten Person auch nicht-autorisierte Avatare oder Manipulationsversuche geben wird. Fraunhofer SIT hat daher verschiedene Blockchain-basierte Technologien auf ihre Eignung in Anwendungen des digitalen Weiterlebens hin untersucht. Solche Blockchain-Technologien bieten Möglichkeiten, fälschungssichere Informationen (beispielsweise über die Herkunft der Avatare) öffentlich zur Verfügung zu stellen. Zum Schutz des geistigen Eigentums an Avataren können digitale Wasserzeichen eingebettet werden, mittels derer anwendende Personen überprüfen, ob ein Avatar von einem vertrauenswürdigen Anbieter stammt. Vielversprechend erscheinen aus Sicht von Fraunhofer SIT die Konzepte zu Self-Sovereign Identity (SSI) und Non-Fungible Tokens (NFTs). SSI-Konzepte können dazu dienen, repräsentierte Personen mit ihren rechtmäßigen Avataren zu verknüpfen und deren Identitäten gegenüber den anwendenden Personen nachzuweisen.

Ergänzend könnte der Nachweis von Besitz und Rechten an Avataren mittels NFTs erbracht werden. Die genannten Konzepte und Technologien befinden sich allerdings derzeit noch in Entwicklung und sind gerade in Bezug auf ML-basierte Anwendungen und deren Nutzung in virtuellen Welten noch weitgehend unausgereift und unerprobt, so dass sie noch nicht einheitlich einsetzbar sind. Herausforderungen sind auch darin begründet, dass es derzeit nur wenige interoperable VR-Plattformen gibt. Die dezentralen, weitgehend Blockchain-basierten Ansätze von SSI und NFT können aber als prinzipiell geeignet angesehen werden.

Abschließend wurden Handlungsempfehlungen aus technischer Sicht formuliert. Beispielsweise empfiehlt Fraunhofer SIT zur Implementierung eines Avatars des digitalen Weiterlebens die Nutzung eines generativen Sprachmodells. Dieses sollte gut dokumentiert sein und Mechanismen gegen schädliche Ausgaben unterstützen. Vor der Erstellung sollten klare Vereinbarungen mit den Auftraggebern bezüglich der inhaltlichen Gestaltung, des Anwendungskontextes und der Nutzung treffen. Dabei sollte geregelt sein, wie autonom der Avatar sein darf und in welchen Kommunikationsräumen er genutzt wird. Zudem ist eine klare Kennzeichnung der Avatare als KI-basierte Lösungen erforderlich, inklusive Nennung des verwendeten Sprachmodells und der personenspezifischen Anpassungsverfahren. Transparenzpflichten umfassen die Bereitstellung detaillierter Informationen über urheberrechtlich geschützte Inhalte, personenbezogene Trainingsdaten, und Sicherheitsmaßnahmen. Anwendende Personen sollten vor der Nutzung über die Gestaltung und den Kontext des Avatars informiert werden, besonders wenn kommerzielle Ziele oder Echtzeitanalysen verfolgt werden. Avatare sollten in ihrer Autonomie begrenzt werden, um Verwechselungen mit lebenden Personen zu vermeiden und die Integrität der menschlichen Kommunikation zu wahren. Anbieter sollten sich verpflichten, auf Emotionsanalysen und virtuelle Produktplatzierungen zu verzichten und sicherstellen, dass Avatare nicht autonom handeln oder durchsetzbare Entscheidungen treffen können, die reale Auswirkungen haben.

2.3 Datenschutz und Persönlichkeitsrecht

Im Rahmen der Analyse der rechtlichen Herausforderungen hat sich das Fraunhofer SIT mit Fragestellungen zum postmortalen Persönlichkeitsschutz sowie zum Datenschutz, dabei insbesondere zur transparenten Datenverarbeitung und zur Umsetzung von Informationspflichten, beschäftigt.

Untersuchung des postmortalen Persönlichkeitsschutzes

Der postmortale sieht sich einer Reihe von Herausforderungen gegenüber, die das digitale Weiterleben einer repräsentierten Person betreffen. Traditionelle Social-Media-Dienste ermöglichen bereits die Erinnerung und das Gedenken an Verstorbene, bleiben jedoch statisch. Der Einsatz von Diensten des digitalen Weiterlebens geht einen Schritt weiter, da dieser Menschen samt ihrer Persönlichkeit „weiterleben“ lassen und eine aktive Kommunikation und Interaktion ermöglichen kann. Die Tatsache, dass Avatare selbstbestimmt handeln können, führt dazu, dass eine neue Dimension an Risiken in Bezug auf den postmortalen Persönlichkeitsschutz entsteht. Dazu gehören das ungewollte Erscheinen als Avatar, die unbeabsichtigte Offenlegung vertraulicher Informationen, die ungewollte Interaktion des Avatars mit Hinterbliebenen sowie mögliche Veränderungen der Persönlichkeit des Avatars. Auch das ungewollte Löschen des Avatars durch Hinterbliebene oder den Dienstanbieter ist eine potenzielle Gefahr. Somit stellen sich bei der Nutzung eines weiterlebenden Avatars immer die

Fragen: Stehen die Handlungen und Entscheidungen des Avatars im Einklang mit den Werten und Überzeugungen der repräsentierten Person? Wie wird die Autonomie und Entscheidungsfreiheit des Verstorbenen gewahrt? Um diese Herausforderungen zu adressieren, können Nutzende durch digitale Vorsorgemaßnahmen und letztwillige Verfügungen ihren digitalen Nachlass regeln und damit Einfluss auf die Verwendung ihrer Daten nach ihrem Tod nehmen. Die Regelung des digitalen Nachlasses ermöglicht es Personen, bereits zu Lebzeiten Vorkehrungen zu treffen, um sicherzustellen, dass ihr Recht auf informationelle Selbstbestimmung auch nach ihrem Tod gewahrt bleibt. Durch vorsorgliche Regelungen, die in einer letztwilligen Verfügung oder einem Testament festgehalten werden können, können Personen steuern, wie ihre digitalen Daten nach ihrem Tod behandelt werden sollen. Dies kann beispielsweise die Anweisung beinhalten, bestimmte Daten nicht zu kommerziellen Zwecken weiterzugeben. Die Ernennung eines Testamentsvollstreckers kann dabei sicherstellen, dass der letzte Wille bezüglich des digitalen Nachlasses umgesetzt wird. Eine weitere Möglichkeit der Adressierung ist die Verpflichtung des Dienstanbieters, bestimmte Parameter für die Nutzung des digitalen Nachlasses festzulegen, um die postmortalen Persönlichkeitsrechte zu schützen. Dies könnte beinhalten, wie lange ein Avatar nach dem Tod aktiv sein soll oder mit wem er kommunizieren darf. Solche Vorgaben ermöglichen es den Verstorbenen, ihre Autonomie über ihren digitalen Nachlass zu wahren, und bieten klare Richtlinien für Hinterbliebene und Dienstanbieter, um potenzielle Missbräuche zu vermeiden.

Datenschutzrechtliche Untersuchungen

-Datenschutzrechtliche Rollen-

Grundvoraussetzung für ein rechtskonformes Handeln im Zusammenhang mit dem virtuellen Weiterleben ist, dass allen Beteiligten klar sein muss, welche datenschutzrechtlichen Rollen sie innehaben und welche Rechte und Pflichten sie entsprechend ihrer datenschutzrechtlichen Rollen erfüllen müssen. Folgende Rollen können den Akteuren im virtuellen Weiterleben zugeordnet werden (Geissler, 2023):

Dienstanbieter	Verantwortliche
Repräsentierte Personen	Lebende Personen: betroffene Personen
	Verstorbene Personen: keine datenschutzrechtliche Rolle
Weitere in das KI-Training involvierte Personen	Betroffene Personen
Kommunikationspartner	Betroffene Personen
Avatar	Keine Datenschutzrechtliche Rolle

Abbildung 2: Anwendungen des digitalen Weiterlebens in Metaversen

Die Rolle der repräsentierten Person im virtuellen Weiterleben ist aus datenschutzrechtlicher Sicht von besonderer Bedeutung, da von ihr i.d.R. zu Lebzeiten zum Anlernen des weiterlebenden Avatars eine Vielzahl personenbezogener Daten verarbeitet werden, sie nach ihrem Ableben jedoch keinen direkten Einfluss mehr auf den Avatar und die damit verbundenen Verarbeitungsprozesse nehmen kann. Vor diesem Hintergrund soll die Rechtmäßigkeit der Verarbeitung personenbezogener Daten der repräsentierten Person nachfolgend – als eines der im Projektverlauf betrachteten Rollen – skizziert werden.

-Rechtmäßigkeit der Verarbeitung personenbezogener Daten-

Die Rechtmäßigkeit der Verarbeitung personenbezogener (u.a. Namen, Informationen zu Hobbys und zum Bildungsweg) im Rahmen des virtuellen Weiterlebens richtet sich primär nach der Regelung der Art. 6 Abs. 1 DSGVO. Eine Verarbeitung ist danach rechtmäßig, wenn mindestens einer der Tatbestände aus lit. a bis f erfüllt ist (Leupold & Wiebe & Glossner, 2021). Die Rechtmäßigkeit der Verarbeitung besonderer Kategorien personenbezogener Daten im virtuellen Weiterleben (u.a. Sprachaufnahmen des Avatar-Inspirators – damit der Avatar die gleiche Stimme wie der Avatar-Inspirator hat –, Informationen zur Gesundheit und ethnischen Herkunft) richtet sich hingegen primär nach Art. 9 DSGVO, wonach für derartige Daten grundsätzlich ein Verarbeitungsverbot besteht (Abs. 1), sofern keine Ausnahme vorliegt (Abs. 2).

Die personenbezogenen Daten des Avatar-Inspirators werden zu seinen Lebzeiten vor allem für das Training der KI verwendet. Es wird hier davon ausgegangen, dass der Avatar-Inspirator selbst tätig wird und seine Daten somit willentlich verarbeitet werden. Die Verarbeitung der personenbezogenen Daten könnte hier zur *Erfüllung eines Vertrags* gem. Art. 6 Abs. 1 lit. b DSGVO erfolgen. Dafür müsste der Avatar-Inspirator einen Nutzervertrag mit dem Dienstanbieter des virtuellen Weiterlebens abschließen. Darüber hinaus muss das Training des Avatars selbst Vertragsgegenstand sein (Rostalski, 2022) und der Nutzungsvertrag alle materiellrechtlichen Voraussetzungen des allgemeinen Vertragsrechts erfüllen (Leupold & Wiebe & Glossner, 2021).

Zu berücksichtigen ist jedoch, dass eine Verarbeitung besonderer Kategorien personenbezogener Daten gem. Art. 9 DSGVO nicht auf dieser Rechtsgrundlage vorgenommen werden kann. Für Verarbeitungen, die nicht unter Art. 6 Abs. 1 lit. b DSGVO subsumiert werden können, kommt vorwiegend die Einwilligung als Rechtsgrundlage in Betracht. Diese könnte bspw. in Verbindung mit Art. 9 Abs. 1 lit. a DSGVO die Verarbeitung personenbezogener Daten besonderer Kategorien legitimieren. Eine Einwilligung nach Art. 6 Abs. 1 lit. a DSGVO setzt voraus, dass die betroffene Person vor ihrer Erteilung in nachvollziehbarer Weise über Art und Umfang der Datenverarbeitung sowie über deren Zweck informiert wird. Des Weiteren muss sie durch eine eindeutige bestätigende Handlung erfolgen, mit der freiwillig, für den konkreten Fall, in informierter Weise und unmissverständlich zum Ausdruck gebracht wird, dass die betroffene Person der Verarbeitung der sie betreffenden personenbezogenen Daten einwilligt. Ebenfalls muss über das Widerrufsrecht informiert werden. Eine Einwilligung nach Art. 9 Abs. 1 lit. a DSGVO setzt darüber hinaus voraus, dass sich die Einwilligung explizit auf die besonderen Kategorien personenbezogener Daten bezieht. Der Avatar-Inspirator wird diese Einwilligung i. d. R. elektronisch erteilen, so dass eine aktive Handlung etwa in Form von Anklicken einer/mehrerer Checkbox(en) erfolgen kann (Enzmann & Selzer & Spychalski, 2019).

Die personenbezogenen Daten des Avatar-Inspirators werden nach seinem Tod vor allem für die Bereitstellung seines weiterlebenden Avatars verwendet. Die DSGVO gilt gem. Art. 4 Nr. 1 DSGVO nur für natürliche Personen. Daten verstorbener Personen sind vom Anwendungsbereich ausgeschlossen (s. Erwägungsgrund 27). Die Mitgliedstaaten können zwar durch die Öffnungsklausel Vorschriften für die Verarbeitung personenbezogener Daten Verstorbener vorsehen, von dieser wurde aber in Deutschland keinen Gebrauch gemacht. Bei der Verarbeitung von Daten von Verstorbenen sind jedoch jene Gesetze zu berücksichtigen, die das Andenken Verstorbener bzw. das postmortale Persönlichkeitsrecht schützen. Sofern Angehörige des verstorbenen Avatar-Inspirators nach dessen Tod Initiatoren des Verarbeitungsprozesses durch den Dienst des virtuellen Weiterlebens sind, ist vor diesem

Hintergrund darauf zu achten, dass das postmortale Persönlichkeitsrecht gewahrt wird (Gola & Heckmann, 2022).

-Informationspflichten-

Im Kontext des digitalen Weiterlebens, bei dem KI-basierte Verarbeitungsprozesse eine große Rolle spielen, ist die Erfüllung der Informationspflichten gemäß der DSGVO von entscheidender Bedeutung. Diese Pflichten dienen dazu, betroffene Personen über die Verarbeitung ihrer personenbezogenen Daten zu informieren und ermöglichen es diesen, ihre weiteren Betroffenenrechte wie z.B. das Recht auf Auskunft, Berichtigung und Löschung wahrzunehmen. Dies ist besonders wichtig, da komplexe KI-Modelle und Algorithmen die Transparenz der Verarbeitungsprozesse erschweren können. Diese Verarbeitungsprozesse müssen jedoch für betroffene Personen erklärbar gemacht werden, damit der Transparenzgrundsatz der DSGVO gewahrt werden kann. Neben den grundsätzlich zu erfüllenden Anforderungen an die Erklärbarkeit sollte auch die Möglichkeit berücksichtigt werden, zielgruppenorientierte Informationen zu implementieren. Dienste des digitalen Weiterlebens können von verschiedenen Altersgruppen genutzt werden, von Kindern und Jugendlichen bis zu Senioren, die in Kontakt mit ihren verstorbenen Angehörigen bleiben wollen. Die Unterscheidung zwischen diesen Zielgruppen ist wichtig, da diese häufig unterschiedliche Fähigkeiten, Präferenzen und Bedürfnisse haben, um die KI und die Funktionsweise des Avatars zu verstehen. So sind jüngere Menschen als *Digital Natives* häufig versierter im Umgang mit digitalen Medien als ältere Menschen. Außerdem können Altersgruppen verschiedene Präferenzen für die Kommunikationsmittel und die Informationspräsentation haben (z.B. Bildsymbole oder kindgerechte Sprache).

Auch die Informationserteilung selbst gestaltet sich im digitalen Weiterleben komplex, da die Datenerhebung häufig in Form von Dritterhebung erfolgt und die Kontaktmöglichkeiten zu den betroffenen Personen begrenzt sein können. In diesen Fällen hat der Verantwortliche jedoch geeignete Maßnahmen zum Schutz der Rechte und Freiheiten sowie der berechtigten Interessen der betroffenen Person zu ergreifen, zu denen auch die Bereitstellung der Informationen über die Verarbeitung zählt. Die Umsetzung dessen stellt eine Herausforderung dar, insbesondere, wenn Personen nicht wissen, dass es diese Dienste gibt und ihre Daten dort verarbeitet werden können. Grundlegendste Voraussetzung sollte hierbei die Aufnahme der relevanten Informationen auf der Webseite des Dienstanbieters sein. Außerdem könnten die anlernenden Personen darum gebeten werden, Datenschutzhinweise an die betroffenen Personen, bspw. in Verbindung mit Info-Broschüren, die weiterführende Erklärungen zum angebotenen Dienst enthalten, weiterzuleiten. Zusätzlich könnte der Verantwortliche immer dann proaktiv auf seine Datenschutzhinweise verweisen, wenn er seinen Dienst bewirbt (Geissler, 2024).

3 Wichtigste Positionen des zahlenmäßigen Nachweises

Im Rahmen der Durchführung des Teilprojektes sind Kosten in Form von Personalkosten und Reisekosten (insbesondere zu Konferenzen und Workshops) entstanden über die im Detail in den Verwendungsnachweisen berichtet wurde.

4 Notwendigkeit und Angemessenheit der geleisteten Arbeit

Der Verlauf der Arbeit im Projekt folgte der im Projektantrag formulierten Planung. Alle im Arbeitsplan formulierten Aufgaben wurden erfolgreich bearbeitet, es waren keine zusätzlichen Ressourcen für das Projekt nötig.

Beide Teilprojekte wurden um einen Monat kostenneutral verlängert, da die Tagung „Unsterblich als Avatar? Ethik, Recht und Sicherheit des digitalen Weiterlebens“, die für November 2023 geplant war, erst am 19. Januar 2024 stattfinden wird. Dies hat eine leichte Verzögerung der Zusammentragung der finalen Inhalte der Studie zur Folge, wodurch auch die von Fraunhofer SIT übernommene Endredaktion der Studie einschließlich des Layouts und Designs sowie der Veröffentlichung und Verteilung etwas später als geplant erfolgen werden.

Insgesamt erarbeitete das Projekt Edilife eine Studie, in der die Ergebnisse der Teilvorhaben der beiden Forschungspartner umfassend dargestellt wurden. Darüber hinaus erreichte das Teilvorhaben des Fraunhofer SIT die in Abschnitt 7 dargestellten Publikationen und Vorträge in einschlägigen Fachzeitschriften und auf Fachkonferenzen. Dies alles, sowie die bereits erfolgte Verwertung von Projektergebnissen, die in dem Dokument „Kurzfassung zum Verwertungsplan“ näher dargestellt wird, bestätigen sowohl die wissenschaftliche als auch wirtschaftliche Relevanz der Ergebnisse des Fraunhofer-SIT-Teilprojekts von Edilife.

5 Darstellung des Verwertungsplans

Fraunhofer SIT plant, die im Rahmen von Edilife entwickelten Ergebnisse in die folgenden Projekte und Forschungsaktivitäten einfließen zu lassen und auf diese Weise zu verbreiten:

Industrie- und Forschungsprojekte. Durch das Projekt Edilife erweiterte Fraunhofer SIT seine Expertise im Bereich von Datenschutz und Datensicherheit. Dies ermöglicht es, im Rahmen von Industrie- und Forschungsprojekten auf diese Expertise zurückzugreifen. Insbesondere werden die Edilife-Ergebnisse zur Sicherheit von Avataren in VR- und AR- und Metaverse-Umgebungen sowie zu den KI-basierten Sprachmodellen der Avatare in das am 1. März 2024 gestartete, BMBF-geförderte Projekt „Ethik, Recht und IT-Sicherheit in Metaversen“ (PriMeta)[1] einfließen. PriMeta widmet sich den drängenden Privatheitsfragen bei der Nutzung virtueller Welten und wird dazu Themenfelder des Datenschutzrechts, der Ethik und der IT-Sicherheit bearbeiten. Es werden Risiken in Bezug auf die Privatheit in Metaversen aus ethischer, rechtlicher und technischer Sicht analysiert sowie technische und nichttechnische Gegenmaßnahmen entwickelt. Die Arbeit in PriMeta kann auf den Ergebnissen der Edilife-Studie aufbauen.

Know-How- und Wissenstransfer. Mitarbeiter des Fraunhofer SIT hielten wissenschaftliche Vorträge und veröffentlichten Fachbeiträge in Fachzeitschriften und Tagungsbänden. Die zentralen Ergebnisse der Arbeitspakete wurden in Form der Edilife-Studie für einen breiten Adressatenkreis aufbereitet. Darüber hinaus enthält die Studie auch Handlungsempfehlungen für zukünftige Anbieter von Avataren des digitalen Weiterlebens.

[1] https://www.forschung-it-sicherheit-kommunikationssysteme.de/projekte/primeta

Die Edilife-Studie wird unter den folgenden Adressen veröffentlicht: https://www.sit.fraunhofer.de/de/reports/ und https://www.sit.fraunhofer.de/de/edilife/ und https://publica.fraunhofer.de/home.

6 Darstellung des Fortschritts auf dem Gebiet des Vorhabens

Aktuell sind insbesondere auf dem Gebiet der generativen KI und bei der Entwicklung großer Sprachmodelle enorme Fortschritte zu beobachten. Entsprechende Anwendungen wie ChatGPT sind sehr populär. ML-basierte Sprachmodelle haben ein großes Potenzial, die inhaltliche Gestaltung der Avatare des digitalen Weiterlebens zu erweitern, da sie ein überzeugendes, nicht-deterministisch erscheinendes Verhalten der Avatare produzieren können. Sie ermöglichen es, nahezu beliebige Gespräche mit einem Avatar zu führen. Die Texte, die ein solcher Avatar generiert, sind in der Regel kaum noch von durch Menschen verfasste Texte zu unterscheiden. Daher ist zu erwarten, dass zukünftige Anwendungen des digitalen Weiterlebens sich diese Entwicklungen zunutze machen und in ihre Angebote integrieren. Aus diesen Gründen hat Fraunhofer SIT während der Projektlaufzeit das Thema generative KI aufgegriffen und als weiteren Schwerpunkt behandelt. Als Ergebnis empfiehlt Fraunhofer SIT eine Nutzung von generativen Sprachmodellen als Grundlage für die inhaltliche Gestaltung von Avataren des digitalen Weiterlebens, da auf diese Weise mit vertretbarem Aufwand kommunikationsfähige Avatare erstellt werden können.

Im Rahmen des Projektes war – als Grundvoraussetzung der daran anschließenden datenschutzrechtlichen Bearbeitung – zunächst zu klären, ob KI-basierten Avataren eine datenschutzrechtliche (Mit-)Verantwortlichkeit zukommen könnte. Im Rahmen dieser Betrachtungen war die zeitlich parallel geführte Abstimmung einer Verordnung zur Regulierung Künstlicher Intelligenz für die Europäische Union von hoher Relevanz. Aufgrund dieser hohen Relevanz wurde die am 1. August 2024 in Kraft getretene Verordnung als Entwurfsversion in dieser Betrachtung mitberücksichtigt und inhaltliche Diskussionen, Änderungen und der Zeitplan bis zum In-Kraft-Treten weiterhin verfolgt.

7 Darstellung der erfolgten oder geplanten Veröffentlichungen

Im Folgenden werden erfolgte Veröffentlichungen und Vorträge genannt, die der Verbreitung der Projektergebnisse dienten.

Vorträge:

- Ines Geissler: Datenschutz in Metaversen und im virtuellen Weiterleben, 4. Privacy & Security at Large Workshop, 27.9.2023, Berlin, GI Informatik 2023.

Veröffentlichungen:

- Ines Geissler: Datenschutzrechtliche Rollen in Metaversen und im virtuellen Weiterleben – Von betroffenen Personen über Auftragsverarbeiter zu Verantwortlichen, GI Informatik 2023, S. 497-510.
- Ines Geissler: Leben in Metaversen und im Virtual Afterlife – Chancen und Risiken zukünftiger Formen des virtuellen (Weiter-)Lebens, GI Informatik 2023, S. 511-521.

- Ines Geissler: Virtuelles Weiterleben – Rechtmäßigkeit der Datenverarbeitung und Umsetzung von Informationspflichten, DuD 2024, S. 164–168.
- Ines Geissler: Das postmortale Persönlichkeitsrecht beim virtuellen Weiterleben – Weiterleben durch einen KI-gestützten Avatar (eingereicht).

Sonstiges:

- Martin Hennig / Ulrich Waldmann: Projektvorstellung auf dem INSIGHT-Forum 2022 in Berlin - Interdisziplinäre Perspektiven des gesellschaftlichen und technologischen Wandels.

Studie:

- Ines Geissler / Jessica Heesen / Martin Hennig / Thomas Kunz / Matthias Meitzler / Ulrich Waldmann: Ethik, Recht und Sicherheit des digitalen Weiterlebens, Fraunhofer SIT und Universität Tübingen 2024.

8 Literatur

Ajder, H., Patrini, G., Cavalli, F., Cullen, L. (2019). *The State of Deepfakes: Landscape, Threats, and Impact*. Deeptrace.

Enzmann, M., Selzer, A., Spychalski, D. (2019). *Data Erasure under the GDPR – Steps towards Compliance*. EDPL, S. 416-420.

Geissler, I. (2023). *Datenschutzrechtliche Rollen in Metaversen und im virtuellen Weiterleben – Von betroffenen Personen über Auftragsverarbeiter zu Verantwortlichen.* GI Informatik, S. 497-510.

Geissler, I. (2024). *Virtuelles Weiterleben – Rechtmäßigkeit der Datenverarbeitung und Umsetzung von Informationspflichten*. DuD, S. 164–168.

Gola, P., Heckmann, D. (2022). *Datenschutz-Grundverordnung und Bundesdatenschutzgesetz (Kommentar)*. Beck-Verlag, 3. Auflage.

Heesen, J. (2022). *Verstorbene als Medienprodukt. Die Programmierung von Unendlichkeit als ethische Herausforderung*. In W. George & K. Weber (Hg.), Die Grenzen des Wachstums: Eigene Endlichkeit. Gießen 2022.

Leupold, A., Wiebe, A., Glossner, S. (2021). *IT-Recht.* Beck-Verlag, 4. Auflage.

Rostalski, F. (2022). *Künstliche Intelligenz. Wie gelingt eine vertrauenswürdige Verwendung in Deutschland und Europa?* Verlag: Mohr Siebeck, 1. Auflage.

Schmidt, J.-H., Taddicken, M. (2017). *Soziale Medien: Funktionen, Praktiken, Formationen*. In J.-H. Schmidt, M. Taddicken (Hrsg.), Handbuch Soziale Medien (S. 23-37). Wiesbaden: Springer.

Weissman, J. (2021). *The Crowdsourced Panopticon. Conformity and Control on Social Media*. London: Rowman & Littlefield.