

KI im Einsatz: Interdisziplinäre Herausforderungen im Projekt UAV-Rescue

Martina Philippi*

Ruhr-Universität Bochum
Institut für Philosophie
Deutschland
martina.philippi@ruhr-uni-bochum.de

Meral Ari

Fraunhofer Institut für Kurzzzeitdynamik
Ernst-Mach-Institut
Deutschland
Meral.Ari@emi.fraunhofer.de

Victoria Heusinger-Heß

Fraunhofer Institut für Kurzzzeitdynamik
Ernst-Mach-Institut
Deutschland
Victoria.Heusinger@emi.fraunhofer.de

Emily Krüger

Bundesanstalt Technisches Hilfswerk
Deutschland
Emily.Krueger@thw.de

ZUSAMMENFASSUNG

Im Projekt „UAV-Rescue: UAV-getragene Sensorik zur KI-basierten Unterstützung von Rettungsmissionen“ wird an einem unbemannt fliegenden System geforscht, das in schwer zugänglichen Räumen bei der Rettungspriorisierung hilft, indem es den Einsatzkräften eine Visualisierung des Raumes mit darin lokalisierten Lebenszeichen zur Verfügung stellt. Diese neue Technologie wirft verschiedene ethische Fragestellungen auf, die unter anderem die Nutzung von Künstlicher Intelligenz zur Lebenszeichendetektion und die Einbettung der neuen Technologie in die Einsatzpraxis betreffen. Die ethische Begleitforschung des Projekts adressiert diese Fragestellungen in einem interdisziplinären und iterativen Diskurs mit Technikpartnern und Endanwendern. Einige dieser Fragestellungen werden hier vor dem Hintergrund der eingesetzten KI-Methoden und der bestehenden Einsatzpraxis beim THW vorgestellt.

1 EINLEITUNG

KI-gestützte Digitalisierung gibt es in allen Bereichen. Die Nutzung von Mustererkennung und die schnelle Generierung von Information auf der Basis umfassender Datenmengen birgt hohes Potenzial, erfordert aber auch einen verantwortungsvollen Einsatz. Auch die Einsatzpraxis im Kontext des *Heavy USAR (Urban Search and Rescue)* [1] nutzt diese Entwicklung in Form von KI-gestützter Technologie zur effizienteren Situationsbewertung. Der Kontext ist für die ethisch und praktisch motivierte Frage nach einem verantwortungsvollen Einsatz von Künstlicher Intelligenz hochinteressant: Zum einen geht es bei der Organisation und Priorisierung der Rettung um Entscheidungen auf Leben und Tod, eine KI-gestützte Informationsquelle als Entscheidungshilfe muss also hohen ethischen Ansprüchen genügen. Hier ist besonders nach dem verantwortungsvollen Umgang mit Restunsicherheit zu fragen, wie es sie auch bei bisher bewährten Suchmethoden gibt. Sensibilisierung der Nutzer

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s).

Veröffentlicht durch die Gesellschaft für Informatik e.V.

in K. Marky, U. Grünefeld & T. Kosch (Hrsg.):

Mensch und Computer 2022 – Workshopband, 04.-07. September 2022, Darmstadt

© 2022 Copyright held by the owner/author(s).

<https://doi.org/10.18420/muc2022-mci-ws10-332>

ist hier ebenso wichtig wie die Gestaltung des User Interfaces (UI). Zum anderen ist die Einsatzpraxis nach strikten Richtlinien und Routinen organisiert, so dass ein verantwortungsvoller Umgang mit dem neuen Werkzeug – über die Sensibilisierung der unmittelbaren Systemnutzer hinaus – über diese Wege in die Praxis eingebettet und stabilisiert werden kann. In unserem Beitrag stellen wir ein solches KI-gestütztes Ortungsinstrument vor, das im Rahmen des vom Bundesministerium für Bildung und Forschung (BMBF) geförderten, derzeit laufenden Verbundprojektes UAV-Rescue konzipiert wird, und fragen nach den ethischen Herausforderungen und möglichen Lösungen im soeben skizzierten Kontext.

1.1 Das Projekt

Im Projekt UAV-Rescue [2] wird ein sensortragendes *Unmanned Aerial Vehicle* (UAV, auch „Drohne“) für den humanitären Einsatz entwickelt. Das System erkundet bei Bergungsaktionen schwer bis durch den Menschen zunächst gar nicht zugängliche Räume und liefert die Daten an eine Bodenstation, die dann ein Lagebild als Entscheidungshilfe für die Einsatzkräfte bei der Lokalisierung von Hilfebedürftigen und der Priorisierung der Rettung zur Verfügung stellt. Zielvorgabe ist die Unterstützung bei der Erkundung der Schadenslage, verbunden mit einer Effizienzsteigerung der Suchaktion in frühen Phasen der Rettung, innerhalb deren die Chancen, menschliches Leben zu retten, am höchsten sind. UAV-Rescue ist ein bilaterales Projekt mit einem österreichischen Partnerkonsortium, das auf die Erkundung im Außenbereich spezialisiert ist. In diesem Paper konzentrieren wir uns auf die deutsche Projektseite, in der es um die Erkundung von Innenräumen geht.

Im angestrebten *Unmanned Aerial System* (UAS), das aus dem UAV sowie einer Bodenstation und einem Controller besteht, kommen ein Radar- und ein LiDAR-Sensor zum Einsatz. LiDAR (light detection and ranging) wird auch im Automobilbau verwendet. Mithilfe von (unsichtbaren und unschädlichen) Laserstrahlen werden Abstände vermessen. Durch eine große Menge ermittelter Messpunkte kann ein Raum als Punktwolke in 3D dargestellt werden. Auf diese Weise wird eine Innenraumkarte erstellt, die den Einsatzkräften als Lagebild übermittelt wird. Das Radar dient zunächst der geplanten teilautonomen Navigation, in der durch den Einsatz einer KI auch autonome Manöver möglich sein sollen, zur Kollisionsvermeidung. Darüber hinaus ermittelt es Bewegungen im Raum, die

mittels Eigenbewegungskompensation von der eigenen Bewegung im Raum unterschieden werden können. Eine zweite KI ermittelt aus diesen Bewegungen jene Muster, die den feinen Bewegungen von Atem und Puls entsprechen, und detektiert so Lebenszeichen. Die detektierten Lebenszeichen werden in der 3D-Karte angezeigt, so dass überlebende Personen lokalisiert werden können.

UAV-Rescue schließt in Details an bereits durchgeführte Projekte an (z. B. radarbasierte Lebenszeichendetektion in LUPE+ [3], Entscheidungsunterstützungssystem in I-LOV [4], Mensch-Maschine-Interaktion in TRADR [5, 6], UAV-Einsatz in Airshield [7]). Der technisch neuartige Entwicklungsschritt ist die Fusion der unterschiedlichen Sensordaten auf einer fliegenden Plattform, die bei ihren Messungen ihre Eigenbewegung berücksichtigen und kompensieren muss. Der Fokus von UAV-Rescue – Innenraumbefliegung mit einem UAV zur Kartierung und Lokalisierung von Lebenszeichen auch bei schlechten Sichtverhältnissen – grenzt das Projekt beispielsweise vom ebenfalls gegenwärtig laufenden Projekt FALKE [8] ab.

1.2 Ethische Begleitforschung in UAV-Rescue: Ziele, Methoden und interdisziplinäre Zusammenarbeit

Das beschriebene System wirft in seiner Gestaltung und in seiner Nutzung verschiedene ethisch relevante Fragestellungen auf, die im Projekt im Rahmen einer ethischen Begleitforschung durchgeführt werden. Aufgabe der ethischen Begleitforschung im Projekt ist eine Identifizierung und Minimierung möglicher Risiken. Neben den forschungsethischen Fragen, die hinsichtlich der Durchführung von Messkampagnen und Systemtests mit Personen, Datenschutz und einer Dual-Use-Bewertung adressiert werden, geht es in der anwendungsethischen Perspektive um folgende zentrale Aspekte des Systems:

- Das System ist hinsichtlich Personal, Transport und Energieversorgung ein ressourcenintensives Device. Dies ist insbesondere bei einer Kosten-Nutzen-Abwägung zu berücksichtigen und in die Gestaltung miteinzubeziehen.
- Das System erhebt Daten und speichert sie voraussichtlich auch, so dass Datenschutzfragen einbezogen werden müssen.
- UAV fliegt zwischen Personen, darf also selbst möglichst keine Quelle von Gefährdung und Verängstigung darstellen.
- Das System verfügt über einen teilautonomen Modus, was hinsichtlich Zuverlässigkeit der pfadplanenden und kollisionsvermeidenden KI sowie der Regelung von Verantwortlichkeit ähnliche Überlegungen auf den Plan ruft wie das automatisierte Fahren [9].
- Das System liefert mithilfe von Sensoren und KI ein Lagebild und beeinflusst daher die Entscheidung bei der Priorisierung der Suche und der Rettung.

Ressourcen-, Datenschutz- und Sicherheitsfragen (Punkt 1-3) waren im Austausch zwischen Endanwendern und Technikpartnern von Anfang an sehr präsent und wurden in der Technikgestaltung direkt adressiert (angestrebte Gewichtseinsparung, datenschutzkonformer Umgang mit Daten, Schutzvorrichtungen am UAV wie

Rotorkäfig). Die Punkte 4-5 sind dezidiert Probleme der Mensch-Maschine-Interaktion und der Künstlichen Intelligenz und als solche anschlussfähig an Diskurse aus Technikethik, Technikphilosophie und Medientheorie. In diesem Paper konzentrieren wir uns auf Punkt 5.

Für alle Risiken gilt, dass sie bei einer Kosten-Nutzen-Abwägung berücksichtigt werden und Maßnahmen zu ihrer Minimierung erarbeitet werden müssen, auch wenn diese Maßnahmen nicht innerhalb des Projektverlaufs realisierbar sind, denn das Projekt ist auf zwei Jahre angelegt und das angestrebte System hat noch nicht Marktreife. Da diese sehr frühe Entwicklungsphase jedoch maßgeblich ist für die Einschätzung der ersten antizipierbaren Folgerisiken und deren Bewertung durch Technikpartner und Endanwender, ist hier ein systematisches Vorgehen gefragt, um alle Parteien zu konkreten Themenkomplexen zu Wort kommen zu lassen und die Ergebnisse für weitere, technikethisch fundierte Diskussionen im Projektverlauf aufzubereiten.

Das hierzu in UAV-Rescue gewählte Vorgehen entspricht Ansätzen zur begleitenden Technikgestaltung. Hierbei sind die konstruktive Technikfolgenabschätzung und, näher am Gestaltungsprozess, das Value-sensitive Design zu nennen. Technikfolgenabschätzung [10, 11] sieht eine zentrale gesellschaftliche und wissenschaftliche Verantwortung im Antizipieren möglicher Technikfolgen bei der Entwicklung einer neuartigen Technologie. Sie stellt sich mit diesem Vorhaben der Herausforderung des sogenannten Collingridge-Dilemmas, demzufolge Technikfolgen erst dann sichtbar würden, wenn die Technik bereits irreversibel in die Praxis eingegangen ist [12]. Im Gegensatz dazu geht sie von der grundsätzlichen sozialen und politischen Gestaltbarkeit von Technik aus und macht es sich zur Aufgabe, einen partizipativen und interdisziplinären Diskurs zu etablieren. Als konstruktive Technikfolgenabschätzung (KTA) begleitet sie gestaltend bereits frühe Entwicklungsstadien.

Technikfolgenabschätzung arbeitet weitgehend interdisziplinär und partizipativ, lässt sich jedoch auch im Bereich der Geistes- und Sozialwissenschaften ansiedeln und als Methode der Angewandten Ethik auch in der Philosophie verorten, die sich bekanntlich durch konzeptionelle Arbeit und argumentativem, erst sekundär empirischem Vorgehen auszeichnet. Für eine ethische Begleitforschung in diesem Sinne bedeutet dies: Auch wenn Erfahrungen aus verwandten Bereichen mit einbezogen werden können und sollten, sind nicht zwingend empirische Untersuchungen nötig, um Bedenken an einer neuartigen Technologie und ihrem Einsatz zu formulieren; die Legitimation von Problemen als behandlungswürdig erfolgt auf argumentativer Basis und im Sinne einer Identifizierung möglicher Risiken, Wertekonflikte und Problemszenarien, gefolgt von der Konzeption möglicher Strategien zur Minimierung oder Vermeidung dieser Risiken.

Value-sensitive Design (VSD) geht ebenfalls davon aus, dass ethische Fragen bereits im Designprozess neuartiger Technologien berücksichtigt werden sollten und zu diesem Zweck bereits in einer frühen Phase der Produktentwicklung ein moderierter Austausch zwischen Entwicklern und Endanwendern einzuplanen und durchzuführen ist [13, 14]. Während in der KTA vor allem Fragen nach langfristigen gesellschaftlichen Folgen behandelt werden, stellt der VSD-Ansatz, neben Werkzeugen zur Erfahrbarmachung neuartiger Technologien, konkrete und auch im Kontext des Rettungswesens bewährte Methoden zur Stakeholder- und Werte-Ermittlung bereit.

Der Schwerpunkt liegt hier mehr auf der ethischen Gestaltung einer neuartigen Technologie als an der Frage nach den langfristigen gesellschaftlichen Folgen ihres Einsatzes, wenngleich dies natürlich in die Betrachtung mit einbezogen werden sollte.

Beiden Ansätzen ist gemeinsam, dass sie zur verantwortungsvollen Technikgestaltung nach dem „*ethical by design*“-Grundsatz [15] beitragen wollen. Nach diesem Grundsatz sind Werte wie Sicherheit, Schutz personenbezogener Daten oder Autonomieerhalt der Nutzenden bereits im Design soweit berücksichtigt, dass dieses einer verantwortungsvollen Nutzung nicht zuwiderläuft und sie im Idealfall sogar unterstützt. Das bedeutet jedoch nicht zwingend, dass die verantwortungsvolle Nutzung dann „narrensicher“ gegeben ist, also ohne zusätzliche Reflexionsleistung durch die Nutzenden; in vielen Fällen muss dieses „*ethical by design*“ ergänzt werden durch ein „*ethical by practice*“. Auf diesen Aspekt wird im vierten Abschnitt noch einmal näher eingegangen.

1.3 Interdisziplinäre und bilaterale Zusammenarbeit in UAV-Rescue

In einer frühen Projektphase wurde in UAV-Rescue ein eintägiger Ethikworkshop veranstaltet, in dem Technikpartner und Endanwender zur Bearbeitung ethischer Fragestellungen zusammengebracht wurden. Mithilfe einer im Search-and-Rescue- und VSD-Kontext bereits bewährten Matrix [16] wurden in Bezug auf die oben genannten fünf Themenkomplexe die Stakeholder, ethisch relevante Werte wie Sicherheit, Verantwortung, Datenschutz und Autonomie und ihre Relevanz für die einzelnen Stakeholder ermittelt. Diesem Treffen gingen zwei Endanwenderworkshops mit der Bundesanstalt Technisches Hilfswerk (THW) zur Bedarfsermittlung und Szenariendefinition voraus. Die Einsatzszenarien des Systems wurden hier auf teileingestürzte Gebäude spezifiziert. Ziel des Verfahrens im Ethikworkshop war zum einen, die zuvor in zwei skizzierten ethisch relevanten Fragestellungen zu verifizieren, und zum anderen, eine erste interdisziplinäre Diskussion zu diesen Fragestellungen anzustoßen, an die im weiteren Projektverlauf bei unterschiedlichen Treffen wieder angeknüpft wird. Anlass zur Weiterdiskussion bieten die Verbundtreffen, die vom THW eingebrachten Präsentationen etwa zu Ortungsroutinen, Suchmethoden und Einsatz von UAVs sowie Ethik-Follow-Ups zu konkreten Fragestellungen.

Dies entspricht der VSD, in deren Kern die Iteration von Fragestellungen steht. Das bedeutet, dass die ethisch relevanten Fragen und Aspekte nicht lediglich zu Projektbeginn thematisiert und in der späteren Entwicklung berücksichtigt werden, sondern dass sie – die Forschung am System begleitend – zu unterschiedlichen Zeitpunkten auf der aktuellen Verständnisbasis wieder aufgegriffen werden. Eine solche Iteration ist aus drei Gründen erforderlich: Zum einen aufgrund eines Sachverhaltes, den man mit Hans-Georg Gadamer als „hermeneutischen Zirkel“ bezeichnen kann [17]: Je mehr man beim Rezipieren neuer Information schon weiß oder verstanden hat, desto mehr versteht und entdeckt man im weiteren Verlauf; dies fügt sich zur bisherigen Wissens- und Verständnisbasis hinzu und ermöglicht weiteres Verstehen und Ideenfindung. Zum anderen verändert sich im Projektverlauf häufig die Informationslage: Technische Machbarkeit, die sich mit der Zeit herausstellt, neue Ideen oder zufällig, etwa durch einen neuen Gesprächspartner im Endanwenderkontext, hinzugekommene Informationen zeigen die

Dynamik der Informationslage, die auch die ethischen Fragestellungen immer wieder in einem neuen Licht erscheinen lässt. Aus beiden Gründen heraus müssen Ziele und Risikoformulierungen kontinuierlich nachgeschärft werden.

Auch die österreichische Projektseite hat eine sogenannte GSK-Beteiligung (d. h. Geistes-, Sozial- und Kulturwissenschaften) mit eher sozialwissenschaftlichem und rechtlichem Schwerpunkt. Auch dort wird kontinuierlich Rücksprache mit Endanwendern gehalten, in diesem Fall spezialisiert auf den Außenbereich. Die ethische Begleitforschung auf deutscher Seite basiert nicht auf einem sozialwissenschaftlichen oder empirischen Vorgehen wie etwa der methodischen Auswertung von Befragungen, sondern auf einem kreativen Diskurs zwischen den Beteiligten, der mit theoretischen, d. h. technikethischen bzw. technikphilosophischen Überlegungen angereichert wird und es ermöglicht, Problembewusstsein für mögliche Technikfolgen zu schaffen und gemeinsam Lösungswege aufzuzeigen und zu diskutieren. Beide Zugänge – der eher theoretische und der eher empirische – fließen in der regelmäßigen Zusammenarbeit zwischen deutscher und ethischer GSK-Forschung in Berichte und Veröffentlichungen ein, so dass das Projekt von der methodischen Vielfalt profitiert.

1.4 Ziel des Papers

In unserem Paper konzentrieren wir uns auf einen Aspekt der Mensch-Maschine-Interaktion in UAV-Rescue: der Umgang mit einer KI-gestützten Entscheidungshilfe vor dem Hintergrund der bisherigen Einsatzpraxis beim THW und der Verwendung von radarbasierter, KI-gestützter Lebenszeichendetektion in Verbindung mit einem Innenraum-Lagebild. Dieser Aspekt wird voraussichtlich auch in weiteren Projekten relevant sein, in denen KI-generierte Information in Form einer 2D- oder 3D-Karte als Entscheidungshilfe bei Search-and-Rescue-Einsätzen zur Verfügung gestellt wird. Um die Problemlage zu skizzieren, erläutern wir zunächst, wie das System aus Sensordaten eine Kartenansicht mit lokalisierten Lebenszeichen erstellt (Abschnitt 2). Mit dem so erstellten Lagebild stellt das System eine Entscheidungshilfe dar, die bisherige Suchmethoden ergänzen soll. Um dies zu kontextualisieren, stellen wir anschließend die bisherigen Suchroutinen und -methoden beim THW vor (Abschnitt 3). Anschließend fragen wir nach den Herausforderungen, vor die das System seinen verantwortungsvollen Einsatz stellt (Abschnitt 4.1), und stellen abschließend Überlegungen zur praktischen Bewältigung dieser Herausforderung vor (Abschnitt 4.2). Da es sich um ein laufendes Projekt handelt, ist die konkrete Gestaltung des System work in progress, aber liefert dennoch eine gute Grundlage für Überlegungen bezüglich zu erwartender Risiken und ihrer möglichen Minimierung.

2 VORGEHEN UND RISIKEN BEI DER INTERPRETATION VON KI-GENERierter INFORMATION IM EINSATZKONTEXT

Im ausgewählten Einsatzszenario fliegt das UAV teleoperiert in einen Raum, der beispielsweise aufgrund seiner Lage in einem höheren Stockwerk oder spezieller Risiken wie Einsturzgefahr oder vermutetem Gasaustritt nicht begehbar ist. Radar- und Lidar-Sensorik ermöglichen eine Erfassung der Szene, und dies auch bei schlechten Sichtbedingungen aufgrund von Dunkelheit oder Rauch. Auf

der Basis dieser Sensordaten sowie mit Unterstützung von KI wird eine 3D-Lagekarte erstellt, die Wände und Objekte im Raum als Punktwolken darstellt und gegebenenfalls lokalisierte Quellen von Lebenszeichen anzeigt. Durch den Einsatz von KI-basierten Auswertemethoden zur Identifizierung von Lebenszeichen aus den durch die UAV-getragenen Sensoren erhofft man sich eine robustere und zuverlässigere Erkennung von zu rettenden Personen, als dies über eine komplexe, stark an Expertenwissen in der Auswertung geknüpfte, explizit programmierte analytische Lösung der Fall wäre. Um einen Zugang zur besonderen Art datenbasierten KI-generierten Wissens zu erschließen, möchten wir im Folgenden einen Überblick darüber geben, wie die KI-generierte Information zur Lebenszeichendetektion in UAV-Rescue zustande kommt und inwiefern diese Information problembewusst, also mit Rücksicht auf mögliche Fehldetektionen, zu bewerten ist – also einen Blick hinter die Kulissen der KI-gestützten Informationsgewinnung in UAV-Rescue werfen.

Die Daten für das Training und Testen der KI entstehen durch dezidierte Messungen (stationär, rein mit dem Radar und UAV-getragen) sowie durch Simulation. Als Sensor bei den Messungen wird ein MIMO-Radar verwendet, das auch durch dünnere Materialschichten wie etwa Kleidung hindurch Bewegung erkennt: Die Radardaten werden dergestalt vorverarbeitet, dass der Raum über eine gewisse Zeitspanne mit Radarsignalen abgetastet wird, aus der Masse an Radardatenpunkten diejenigen mit einer zeitlichen Veränderung (Varianz) herausgepickt und für diese eine Slowtime-Analyse durchgeführt, also die Veränderungen der gemessenen Phase über die Zeit extrahiert wird. So werden gezielt Raumbereiche ausgewählt, an denen systematisch regelmäßige Bewegungen detektiert werden. Für diese Messpunkte wird dann das Slowtime-Signal an die KI zur genaueren Analyse weitergegeben.

Diese analytische Datenvorauswahl und –vorverarbeitung ist notwendig, da das KI-System zunächst auf der Bodenstation laufen soll und die Datenübertragungsraten vom UAV-System mit der Messsensorik zur Bodenstation potenziell begrenzt sind. Die vom Radarsystem aufgezeichneten Daten werden bereits auf dem UAV um die durch die Eigenbewegung des UAV verursachten Störungen rechnerisch weitestgehend bereinigt. Zur Lebenszeichenklassifizierung wird dann ein künstliches Neuronales Netz (ANN) eingesetzt, welches datenbasiert und überwacht trainiert wird, um zunächst entscheiden zu können, ob es sich bei den Bewegungsänderungen um Lebenszeichen, also Atem und Herzschlagbewegungen z.B. eines menschlichen Brustkorbs handelt, oder ob die Bewegung Nicht-Lebewesen bezogenen Ursprungs ist, wie z.B. einem sich in einem Luftzug bewegenden Vorhang oder Pflanze. In weiteren Entwicklungsschritten soll zudem untersucht werden, ob sich klassifizieren lässt, ob das Lebewesen wirklich ein Mensch ist, oder z. B. ein Tier und ob man aus den Daten schon erste Schlussfolgerungen auf den Gesundheitszustand extrahieren kann – dies wäre die ideale Leistung des Systems.

Die Wahl des Einsatzes eines ANN liegt in verschiedenen Punkten begründet: Neuronale Netze haben die Fähigkeit, nicht-lineare und komplexe Zusammenhänge zu lernen und zu modellieren; Sie können verallgemeinern, d. h. nachdem sie aus den ursprünglichen Eingaben und ihren Beziehungen gelernt haben, können sie auch aus ungesehenen Daten ungesehene, aber ähnliche Beziehungen ableiten, so dass das Modell verallgemeinert werden und Vorhersagen

für ungesehene Daten treffen kann; ANNs erlegen den Eingabevariablen zunächst keine direkten Beschränkungen auf. Sie sind in der Lage, verborgene Beziehungen in den Daten zu erlernen, ohne zuvor explizit feste Beziehungen in den Daten vorzuschreiben, sondern das Netz muss die relevanten Merkmale aus den gegebenen Daten durch den Lernprozess „selbst“ herausfinden. Dies geschieht durch die iterative Korrektur und Optimierung der Netzparameter (weights and bias), welche in komplexeren Problemstellungen sehr mannigfaltig sein können. Schon allein deshalb braucht man für einen ausgewogenen Trainingsdatensatz und eine sinnvolle Optimierung des Systems viele aussagekräftige Daten. Aus diesem Grund wird hier ein Simulationstool verwendet, um randomisierte Datenpunkte mit Rauschen und anderen Signalstörungen mit Signalverläufen im Frequenzbereich der menschlichen Atmung sowie der Herzfrequenz zu erzeugen. Es ermöglicht uns auch, den Datensatz zu erweitern, um die Vielfalt der Datenpunkte zu vergrößern, insbesondere im Bereich der Vitalzeichen, um auch verschiedene Gesundheits- (und Körper-)Zustände einzubeziehen.

Da ein KI-System ein zumeist hochkomplexes und dabei nicht explizit programmiertes Softwaretool ist, somit nicht einfach durch Lesen des Quellcode nachvollziehbar ist, wie es sich in bestimmten Fällen insbesondere in Grenzbereichen verhält, ist weiterhin der aussagekräftigste Weg, es auf seine Leistungsfähigkeit zu prüfen, das dezidierte Testen durch explizit gebaute Testdatensätze, die alle relevanten Fälle abdecken sollten. Hier gibt es eine Methode (counterfactual method) [18] dieses Verfahren neben der Prüfung der Zuverlässigkeit auch zum Verständnis des eigentlichen Entscheidungsprozesses (Explainable AI) zu nutzen, indem einzelne Parameter bzw. Eigenschaften der Inputdaten gezielt verändert werden und dann beobachtet wird, wie sich das Ergebnis verändert.

Mögliche Quellen von Fehldetektion speisen sich aus aleatorischer oder epistemischer Unsicherheit oder Überanpassung (Overfitting) und können mit unterschiedlichen Maßnahmen verringert werden. Generell lässt sich die Unsicherheit der Entscheidung des KI-Systems in zwei Bereiche abgrenzen, der Unsicherheit die bereits innerhalb des Datensatzes vorhanden ist und der durch das Modell selbst verursachten Ungenauigkeit in den Berechnungsprozessen. Erster umfasst sowohl Ungenauigkeiten und Störungen innerhalb eines Datenpunktes (in unserem Fall z. B. Rauschen, die Eigenbewegung des UAV und die Ungenauigkeiten und Artefakte die im Pre-Processing der Daten auftreten) wie auch die innerhalb der Datengesamtheit (z.B. Fehler/Ungenauigkeiten beim Labeling, Ausgewogenheit der Daten, Bias in der Datenzusammenstellung). Zweiter ergibt sich aus Ungenauigkeiten durch die Verarbeitung in einem Computersystem (z.B. Rundung oder Memoryprobleme) und die Approximationen die mathematisch-algorithmischen Umsetzungen in der Regel inhärent sind, sowie der praktischen Umsetzung des Trainingsprozesses (Abbruchkriterium für die Beendigung des iterativen Prozesses des Trainings, Overfitting, Underfitting). Um die Stabilität eines KI-Modells zu validieren, braucht man eine gewisse Sicherheit, dass das Modell die meisten Muster aus den Daten richtig erkannt hat und nicht zu viel Rauschen aufnimmt, oder anders gesagt, dass es wenig Verzerrungen und Varianz aufweist.

Es gibt eine ganze Reihe von Ansätzen um diesen Problemquellen zu begegnen, wie z.B. probabilistischen Ansatz für Deep Learning um Modellunsicherheiten zu verringern oder Regularisierung um Overfitting zu vermeiden. Es bleibt aber immer ein gewisser Rest an

Ungenauigkeit bzw. Potential an Fehlentscheidungen bestehen, die eine hundertprozentige Korrektheit einer durch eine KI getroffenen Entscheidung quasi ausschließen. In der Praxis macht dies eine zuverlässige Einschätzung dieser Lücken und damit des „Wissens“ der KI notwendig.

Technisch gesehen ergibt sich eine mögliche Einschätzung des Wissens einer KI als Kombination aus der Qualitätsabschätzung des Trainingsdatensatzes, der Bewertung der Trainingsgüte (z. B. durch eine Konfusionsmatrix oder die Anwendung von Monte-Carlo-Dropout-Methoden [48]) und verschiedenen Ansätzen zur Einschätzung der Zuverlässigkeit des Systems.

Generell ist gerade diese Problematik ein andauerndes Forschungsthema in allen drei aufgeführten Punkten.

Gerade für den letzten Punkt spielt z.B. auch das Grundverständnis des Entscheidungsprozesses der KI eine nicht unerhebliche Rolle. In den letzten Jahren wurden, unter dem Begriff Explainable AI zusammengefasst, eine ganze Reihe an Methoden zum besseren Verständnis des „Denkprozesses“ eines Neuronalen Netzes entwickelt, wie z.B. *Heat Maps* oder *Layer-wise relevance propagation* [19, 20]. Insbesondere für dem menschlichen Gehirn nicht nativ zugängliche Daten, wie die hier im Projekt auszuwertenden Radardaten die ein hohes Expertenwissen brauchen um überhaupt für den Menschen verständlich zu sein (z.B. im Gegensatz zu Bilddaten, welche unser Gehirn tagtäglich verarbeitet) sind aber auch die Ergebnisse von solchen Erklärungsansätzen häufig nur bedingt nachvollziehbar und damit insbesondere für die Erklärung gegenüber reinen Anwendern eines solchen KI-Systems nicht verständlich darstellbar.

In UAV-Rescue spielen, zusätzlich zur Unsicherheit des KI-Systems, für das finale Ergebnis natürlich alle anderen Komponenten des Datenflusses von der Datenakquise bis zur Darstellung ebenfalls eine Rolle, einschließlich der Genauigkeit des Messsystems und die Ungenauigkeiten und Fehlerquellen des Pre- und Post-Processings. Ersteres beinhaltet die Vorauswahl der relevanten Datenpunkten aus der Menge der insgesamt in einer Messung aufgenommenen Radardaten, die Slowtime-Data-Extraktion, mögliche Kompressionen zur Datenübertragung. Zweiteres enthält z.B. die (möglicherweise vereinfachte und damit auf eine Art komprimierte) Darstellungsform des Ergebnisses über die GUI.

Aus dieser Darstellung ergibt sich, wie komplex die Genese der Information gestaltet ist, die das entwickelte UAS den Einsatzkräften zur Verfügung stellt. Zwar basiert diese Information auf Messdaten, doch dennoch darf diese Art von Objektivität nicht fälschlich auf die KI übertragen werden. Auch wenn die genannten Maßnahmen zur Steigerung der Zuverlässigkeit und Transparenz sorgfältig ergriffen werden, basiert die KI-generierte Information auf Mustern, Korrelationen und Wahrscheinlichkeiten. Dies bedeutet: Durch den Einsatz des Systems ist die Informationslage über schwer zugängliche und mit bisherigen Suchmethoden nicht oder schwierig zu erkundenden Innenräumen deutlich verbessert. Dennoch ist zu beachten, dass hier ein datenbasiertes Modell vorliegt, also keine absolut sichere Abbildung des Raumes. Aus diesem Grund ist das System als Entscheidungshilfe zu betrachten, als Ergänzung der bisherigen Suchmethoden und nicht etwa als Ersatz für solche. Hier ist die Expertise der Endanwender im Umgang mit Unsicherheit gefragt.

3 DAS UAS IM EINSATZ: ERFORDERNISSE DER EINBETTUNG VON KI-GESTÜTZTER TECHNOLOGIE IN DIE PRAXIS

Die Nutzung von technischen Suchinstrumenten beim THW findet bisher in der Form der Bedienung eines Geräts statt, noch nicht im Sinne einer Kommunikation zwischen menschlichen und künstlichen Akteuren. Das gilt auch für das System, an dem in UAV-Rescue geforscht wird: Auch wenn hier KI eingesetzt wird, ist das System als ein Werkzeug zu verstehen und nicht als „Partner“ etwa im Sinne eines kollaborierenden Roboters („Cobot“), der unter entsprechender Führung eigenständig agiert [21] und besondere kognitive Leistungen aufweisen muss [22]. Damit ähnelt das UAS in seiner Nutzung den bisher gängigen Einsatzmitteln in den Trupps der THW-Ortung, die entweder technischer oder biologischer Natur sind. Die technische Ortung nutzt Wärmebildkameras, [23] optische Kameras (sogenannte „Searchcams“) [24] und auch akustische Verfahren (Bodenhörgeräte) [25]. Die biologische Ortung hingegen nutzt Hunde [26]. Da auch die Trupps „Unbemannte Luftfahrzeuge“ (Trupp UL) Aufgaben aus dem Bereich der Ortung wahrnehmen, können auch UAVs [27] zum Einsatz kommen. Es sind aktuell weit weniger Trupps UL im THW disloziert als Fachgruppen der Ortung; auch die Nutzung von UAVs in Innenräumen wie in UAV-Rescue ist eine neue Entwicklung. Für alle Suchmethoden gilt der Grundsatz, dass Einsatzmittel in Ergänzung zueinander eingesetzt werden, und jedes Ergebnis mit einem anderen Einsatzmittel verifiziert wird [28]. Dies ist für UAV-Rescue von besonderer Bedeutung.

Die Einsatzpraxis ist stark von Routinen geprägt. Diese Routinen bestehen zum einen in einer sehr klaren, hierarchischen Rollen- und Aufgabenverteilung, zum anderen in den Prozessen selbst. Im konkreten Fall des THW bedeutet dies: Einsatzabläufe werden durch eine Vielzahl von Gesetzen und Dienstvorschriften sowie durch Standards und Empfehlungen geregelt [29]. Abgesehen vom THW-Führungsvorgang, der die Weitergabe von Informationen zur Entscheidungsfindung regelt, prägt auch das Konzept der fünf Phasen der Bergung bzw. Rettung den Einsatzablauf maßgeblich [30]. Die Ortungs- und Bergungstrupps des THW strukturieren bei Inlandseinsätzen ihre Arbeit basierend auf diesen fünf Phasen, die konkret aus dem Erkunden, dem Absuchen, dem Durchforschen, dem Eindringen und Orten sowie aus dem Beräumen bestehen. International wurde das Konzept durch die Guidelines der International Search and Rescue Advisory Group (INSARAG) in ähnlicher Form standardisiert. Hier werden diese Phasen als ASR (Assessment, Search and Rescue) Levels 1 bis 5 bezeichnet [31]. Das in UAV-Rescue zu entwickelnde System bietet insbesondere für Phase 3 (Erforschen) bzw. ASR Level 2 („*Worksite Triage Assessment*“) Unterstützung. Basierend darauf, welche dieser Phasen Ortungs- und Bergungsteams bewältigen können, werden diese von INSARAG auch zertifiziert; ein solches Team ist die Schnelleinsatzeinheit Bergung Ausland (SEEBA) des THW [32]. Jedoch hat die SEEBA bisher noch in keinem Einsatz ein UAV eingesetzt, noch verfügt sie über ein solches [33]. Während eines solchen Einsatzes sehen die Regeln der Triage vor, dass zunächst dort gearbeitet wird, wo sicherlich Personen verschüttet sind, die innerhalb von 12 Stunden gerettet werden können. Zweite Priorität haben die Verschütteten, die in mehr als 12 Stunden gerettet werden können. Danach wird an jenen Stellen gearbeitet, wo Personen nur vermutet werden. Gleichzeitig

gilt der Grundsatz, dass dort, wo die größte Anzahl an Personen bekannt ist oder vermutet wird, vorrangig gearbeitet wird. Diese Regeln der Triage finden somit auch bei den Auslandseinsätzen der THW-SEEBA Anwendung [34] und veranschaulichen, warum die Einsatzkräfte in der Lage sein müssen, die vom System erhobenen Messdaten zuverlässig einordnen zu können.

Auch bezüglich der Arbeitsteilung ist der Einsatz von UAS beim THW klar geregelt. Die UAV des THW kommen im Trupp Unbemannte Luftsysteme (Tr UL) [35] in den Einsatz. Ein Trupp UL besteht aus vier Einsatzkräften und seine Kernaufgaben sind die Erkundung aus der Luft sowie die Ortung. Da die UAV des THW meist mit einem optischen Sensor ausgestattet sind, werden diese derzeit beispielsweise für die Personensuche in der Fläche genutzt. Die gewonnenen Daten werden sowohl lokal genutzt als auch den Führungsstellen zur Verfügung gestellt. Geflogen werden die UAV von ausgebildeten Luftfahrzeugfernführern, die auch die Flugpläne erstellen, etwaige erforderliche Freigaben einholen und für die Überprüfung der Flugausstattung auf Verwendungsfähigkeit und Vollständigkeit verantwortlich sind. Jedoch ist die Kommunikation mit anderen Einheiten, bzw. innerhalb der vorgegebenen Führungsstruktur, Aufgabe des Truppführers UL, der bei Bedarf Unterstützung von einem Sprechfunker erhält [36]. Somit werden die Informationen, die auf der Bodenstation eingehen, sowohl vom Truppführer UL als auch von anderen THW-Einheiten genutzt und ausgewertet. Die vom Trupp UL generierten Informationen können aber auch an THW-externe Stellen weitergegeben werden, und zwar über den sog. Führungskreislauf. Häufig liegt die Leitung des gesamten Einsatzes nicht beim THW, sondern bei der Feuerwehr oder anderen Bedarfsträgern.

Unabhängig davon, wohin die Informationen fließen bzw. auf welche Hierarchieebene sie gelangen, gliedern sich diese in den sog. Führungsvorgang ein. Seine Elemente sind die Lagefeststellung (Erkundung bzw. Kontrolle), Planung mit Beurteilung der Lage und Entschluss und anschließend die Befehlsgebung [37]. Insbesondere die Lagefeststellung kann von den Informationen, die von einem Trupp UL zur Verfügung gestellt werden, profitieren, da kontinuierlich Informationen erhoben werden. Die höheren bzw. externen Hierarchieebenen sehen hierbei jedoch auch wiederum eine Art Ausschnitt der tatsächlichen Daten, da die Granularität mit der Weitervermittlung zunehmend, aber geregelt, abnimmt. Die Routinen einer Einsatzorganisation wie dem THW haben eine deutliche positive Auswirkung hinsichtlich der Einbettung von KI-gestützten Hilfsmitteln in die bestehende Praxis: Erstens liefern sie bereits Explikationen von Prozessen, Situationen, Objekten und Akteuren, die in der im System verwendeten Modellierung der Situation aufgegriffen werden können. Insbesondere die einsatzrelevante Qualitätssicherung und Optimierung von Prozessen ist bereits Gegenstand der Forschung im Kontext von Rettungsrobotik [38]. Wovon die Routinen jedoch nicht schützen, ist eine unsachgemäße Handhabung des Systems im Sinne unangemessener Erwartungen – nämlich bei der Interpretation der KI-generierten Information und darauf aufbauend einer unzulässigen Basis von Annahmen bei der Priorisierung der Rettung. Die Herausforderung besteht dabei in einer sachgemäßen Darstellung (GUI) und Interpretation (Mensch) der KI-generierten Informationen.

4 USABILITY UND SACHGERECHTE INTERPRETATION VON KI-GENERierter INFORMATION IM EINSATZKONTEXT

In den vorigen beiden Abschnitten wurde gezeigt, wie das System Lebenszeichen detektiert und diese Information dann in eine mittels LiDAR erzeugte 3D-Lagekarte einfügt, so dass Überlebende im Einsatzszenario lokalisiert werden können. Außerdem wurde die bisherige, von Routinen und einem professionellen Umgang mit Informationsunsicherheit geprägte Einsatzpraxis beim THW vorgestellt, in die das System aus dem Projekt UAV-Rescue perspektivisch eingebettet werden soll. Im Folgenden werden die Herausforderungen skizziert, vor die das System sowohl die Endanwender als auch jene Technikpartner stellt, die im Projekt und perspektivisch darüber hinaus für die Mensch-Maschine-Schnittstelle des Systems verantwortlich zeichnen. Dabei geht es weniger um die Steuerung des Systems – angedacht ist hier ein Controller in Ergänzung zur Bodenstation, so dass zwei Personen das System bedienen – als um die Anzeige der Information und ihre Modifikationsmöglichkeiten auf dem Bildschirm der Bodenstation. Konkret handelt es sich um die Darstellung des beflogenen Innenraums mit den lokalisierten Lebenszeichen.

4.1 Risikoszenarien beim Einsatz der KI-gestützten Entscheidungshilfe

Beim Einsatz einer KI-gestützten Entscheidungshilfe, wie sie in UAV-Rescue erforscht wird, sind aus ethischer Perspektive insbesondere zwei Szenarien zu beachten, in denen Fehleinschätzungen möglich sind und in denen die Information daher überprüft werden muss, um – möglicherweise fatale – Fehlentscheidungen in der Rettungspriorisierung zu vermeiden. Anhand dieser zwei Szenarien werden im Folgenden zwei ethisch relevante Besonderheiten des hier behandelten UAS dargestellt.

Das erste Szenario betrifft die Möglichkeit, dass in einem kartierten Raum kein Lebenszeichen lokalisiert wird, obgleich sich eine überlebende Person im Raum befindet. Dies kann unterschiedliche Ursachen haben, etwa ein sehr schwaches Vitalsignal oder eine unvollständige Detektion des Raumes. Eine solche Nicht-Detektion ist auch bei anderen Suchmethoden denkbar, etwa wenn ein Suchhund nicht anschlägt oder eine akustische Suchmethode oder eine Searchcam keine Hinweise auf die Anwesenheit einer Person erfasst. Das System bringt jedoch durch die Visualisierung des Innenraums in einer 3D-Karte eine neue Herausforderung mit sich. Bereits bei einer 2D-Karte wirken sich sowohl die Informationsdarstellung in der Karte einerseits als auch die Interpretation der Karte andererseits maßgeblich auf die Lagebeurteilung im Einsatz aus [39]. Bei einer 3D-Ansicht kommt hinzu, dass eine Raumeinsicht gewährt wird, die allerdings keine Wiedergabe des Sichtbaren im Sinne eines Kamerabildes ist, sondern ein „errechnetes Bild“. Zum einen hängt die Korrektheit des Angezeigten von der dahinterstehenden Rechentechnologie ab, zum anderen ist die Interpretationsleistung bei der Interpretation des 3D-Bildes anders als bei der – gegebenenfalls als „abstrakter“ empfundenen – „Draufsicht“ der 2D-Karte. So kann sich hier ein Effekt verstärken, den bereits 2D-Karten haben, nämlich die Darstellung eines vollständig erscheinenden Kontextes auch bei unvollständiger Detektion. Im Fall des Systems aus UAV-Rescue steht dank der Punktwolkendarstellung aller im Raum

befindlichen Objekte voraussichtlich nicht zu befürchten, dass eine Stelle, an der die radar- und KI-gestützte Lebenszeichendetektion keine Person lokalisiert, leer erscheint. Dennoch sollte dieses Risiko bei diesem und ähnlichen Systemen berücksichtigt werden, auch weil die Punktwolke errechnet ist und damit selbst bei einer hohen Zuverlässigkeit ein gewisses Fallibilitätsrisiko übrigbleibt.

Das zweite Szenario ist die Fehldetektion, d. h. fälschlicherweise werden Bewegungen im Innenraum als Lebenszeichen eingeordnet, die nicht von einer Person stammen. In diesem Fall besteht das Risiko in einer zeit- und ressourcenaufwändigen Rettungsaktion in einem gefährlichen Umfeld, die nicht zur Rettung einer Person führt, während noch nicht erfasste Überlebende dringend Hilfe benötigen. Die Zuverlässigkeit der KI trägt hier maßgeblich zur Minimierung dieses Risikos bei; doch gerade hinsichtlich der Unterscheidung von Mensch und Tier ist hier, gemäß dem THW-Grundsatz, eine weitere Überprüfung angezeigt. Auch dieses denkbare Szenario zeigt, dass die Möglichkeit einer Nicht- oder Fehldetektion nicht prinzipiell ausgeschlossen ist und bei der Gestaltung von System und Einsatzpraxis berücksichtigt werden muss, denn es besteht trotz verbesserter Informationslage noch Unsicherheit.

4.2 Lösungsweg in der Gestaltung: ethical by design

Da das System in schwer zugängliche Räume einfliegt, für die sich die meisten anderen Suchmethoden weniger gut eignen, bietet sich hier eine technische Lösung zur Gegenprüfung an. So kann die Überprüfung des ersten Lagebildes etwa durch das Zuschalten zusätzlicher Information geschehen. Umschaltmöglichkeiten etwa von der Punktwolkenansicht in eine – bei entsprechender Ausstattung des UAVs – (Infrarot-)Kameraansicht oder in einen Modus mit weiteren Informationen zu den verarbeiteten Sensordaten erhalten hier eine besondere Bedeutung. Zugleich stellen sie sowohl das Graphische User Interface (GUI) als auch die Nutzungspraxis vor spezielle Herausforderungen, die im Folgenden skizziert und kontextualisiert werden sollen.

Die Herausforderung einer Benutzeroberfläche, die ihre anspruchsvollen Funktionen auch in einer Stresssituation und gegebenenfalls auf einem Tablet übersichtlich darstellt, ist speziell. Natürlich muss jede Benutzeroberfläche die verfügbaren Optionen übersichtlich und den Ansprüchen der Nutzergruppe gemäß darstellen, etwa in einer plausiblen Menüanordnung. Auf einer ersten Ansicht ist die Präsentation der Funktionen dann im Vergleich zu ihrem Gesamtumfang in der Software reduziert [40]. Wie stark diese Reduktion vorgenommen werden kann und sollte, hängt im Allgemeinen davon ab, ob es sich bei der Zielgruppe um Laien oder um *Professionals* handelt. Sehr intuitiv bedienbare Anwendungen für den Alltagsgebrauch unterscheiden sich – ungeachtet der Probleme, die dies mit sich bringen kann – in ihrem Anspruch, aber eben auch in ihren Möglichkeiten stark von Programmen für den professionellen Gebrauch; man kennt die umfangreichen Menüs von Photoshop für die Bildbearbeitung, Ableton für die Musikproduktion oder QGIS für die Erstellung und Bearbeitung von Karten. Bei diesen professionellen Anwendungen kann jedoch vorausgesetzt werden, dass die Nutzerin für ihre Arbeit einen stationären Rechner in einem hinreichend konzentrationsfreundlichen Umfeld zur Verfügung hat; es handelt sich nicht um das spezielle Mobile-Work-Szenario im

Einsatzkontext [41], das seine eigenen Anforderungen an Nutzerfreundlichkeit und *Situational Awareness* [42] mit sich bringt. Eine Anwendung im Einsatzkontext muss auf kleinem Raum die wichtigsten Informationen parat halten und dabei auf umfangreiche Menüs verzichten.

Zugleich fordern Umschaltoptionen die Aufmerksamkeit der unmittelbaren Systemnutzer heraus: Man schaltet nicht nur eine Funktion um, sondern den eigenen Fokus. Man verlässt die Ansicht, in der man sich gerade gedanklich bewegt. Das gedankliche Navigieren durch ein 3D-Raummodell ist eine andere intellektuelle Herausforderung als das Auslesen von zusätzlicher Information oder die Betrachtung eines Kamerabildes. Die psychologische Problematik der „Wechselkosten“ [43] ist auch bei einem solchen Perspektivenwechsel zu erwarten; hier erscheinen empirische Studien zur Vertiefung und Verifizierung dieser Einschätzung lohnend. Aus einer anderen Perspektive, nämlich einer technikphilosophischen oder medienwissenschaftlichen, kann dieses Phänomen dadurch erklärt werden, dass Werkzeuge – ob einfache oder komplexe, analoge oder digitale – die Tendenz haben, in ihrem Gebrauch zu „verschwinden“, unsichtbar zu werden; die Aufmerksamkeit richtet sich nicht auf das Werkzeug, sondern auf die Tätigkeit, in deren Kontext das Werkzeug verwendet wird. Das „Sichtbarwerden“ des Werkzeugs wird daher häufig als kontraintuitiv oder irritierend empfunden. Es ist denkbar, dass im Verlassen der Kartenansicht genau diese Irritation entsteht, auch wenn dies nicht bewusst wird; dass die Systemnutzenden insbesondere in der Stresssituation des Einsatzes unbewusst dazu neigen, in der Kartenansicht zu verharren. Dieses Phänomen kann als *cognitive bias* beschrieben werden und stellt im Kontext des KI-gestützter Entscheidungsfindung ein besonderes Problem dar [44]. *Cognitive biases* sind nicht zwingend auf Nachlässigkeit oder Naivität zurückzuführen, sondern mit der Charakteristik der Aufmerksamkeit, dass man, wenn man sich auf eine Sache konzentriert, anderes ausblenden muss. Hier lohnt es sich medientheoretische [45] oder phänomenologische [46] Überlegungen in den Kontext der KI-Nutzung zu übertragen, um ein vertieftes Verständnis für ihre Besonderheiten zu entwickeln.

Hier soll mit der Überlegung geschlossen werden, dass gerade die stark routinisierte Einsatzpraxis des THW prädestiniert ist, dieses Problem aufzufangen. Wie oben gezeigt, sind Sensibilisierung in Schulungen und die explizite Gestaltung von Prozessen in Regelwerken wesentliche Standards bei der standortübergreifenden, hochleistungsfähigen Einsatzpraxis des THW. Diese Standards können helfen, die nötige „Awareness“ beim Gebrauch KI-basierter Technologien in Routinen zu überführen und somit das mentale Umschalten in einen „bewusst skeptischen“ Modus gewohnheitsmäßig zu motivieren und damit zu erleichtern. Zugleich können auf diese Weise die Risiken uneinlösbarer Erwartungen an „gesichertes Wissen“ in der Informationsvermittlung des Entscheidungsunterstützungssystems minimiert werden. Erwartungen können sich aus bisherigen Wahrnehmungsgewohnheiten oder medialer Darstellung von Künstlicher Intelligenz ergeben; ebenso ist es denkbar, dass bei einem System mit deutlich präsenter Sensorik und naturgemäß unsichtbarer Künstlicher Intelligenz jene Objektivität, die Sensoren in der Regel zugeschrieben wird, auch dem Gesamtsystem zugeschrieben wird. Solche Erwartungen können ebenfalls dazu verführen, die Entscheidungsunterstützung als Ersetzung und nicht als Ergänzung bisheriger Suchmethoden zu betrachten. Auch ein

Grundverständnis des Systems trägt daher zur Sensibilisierung und somit zum verantwortungsvollen Umgang mit der leistungsfähigen, neuartigen Technologie bei.

5 ZUSAMMENFASSUNG

Bei der technischen Betrachtung der Sensor- und KI-gestützten Erstellung von Information über einen schwer zugänglichen Innenraum, die das System bereitstellt, hat sich gezeigt, dass auch bei sehr guter Performanz des Systems eine prinzipielle Fallibilität nicht ausgeschlossen werden kann und daher die beim THW übliche Zweitprüfung bzw. das Einbeziehen zusätzlicher Informationen unabdingbar sind. Vor allem aus diesem Grund ist das System nicht als Ersetzung, sondern als Ergänzung bisheriger Suchmethoden zu verstehen. Sichtbar geworden ist dabei, dass die verbleibende Unsicherheit über die Situation und das Erfordernis eines problembewussten Umgangs mit einer solchen KI-gestützten Entscheidungshilfe ihren Wert für die Einsatzpraxis nicht mindert. Das System erweitert die Informationslage über schwer zugängliche Innenräume enorm, und die besonderen Herausforderungen, die durch die Genese der Information sowie die Darstellung in einer Innenraumkarte ergeben, lassen sich gut durch die besondere Gestaltung des Anwendungsfeldes durch Routinen auffangen und mittels Standards und Schulungen minimieren. Dies ist nicht zuletzt ein wichtiger Aspekt beim Thema Akzeptanz, da hohe Erwartungen naturgemäß auch zu nachhaltigen Enttäuschungen und somit Akzeptanzverlust führen können.

Wie die Entwicklung des UAS selbst, sind auch die hier vorgestellten Überlegungen der ethischen Begleitforschung *work in progress*. Sie werden kontinuierlich ausgearbeitet, mit Technikpartnern und Endanwendern auf deutscher und österreichischer Seite diskutiert und so im Sinne des eingangs dargestellten iterativen Verfahrens in den Entwicklungs- und Gestaltungsprozess zurückgespielt. Darüber hinaus sind sie für einen breiteren Diskurs im Bereich sicherheitskritische Systeme und Unterstützung schwerwiegender Entscheidungen anschlussfähig. Gerade Lagebildgebende Systeme wie die Entscheidungsunterstützung aus UAV-Rescue haben ein hohes Potenzial; ihre Entwicklung und Verbreitung ist daher zukünftig vermehrt zu erwarten. UAV-Rescue verspricht gutes Material für eine Case Study zu diesem Thema, weil das System gleich zwei Komponenten erhält, die im Vergleich zu bisherigen Suchmethoden eine neuartige Interpretationsleistung der übermittelten Information erfordern, nämlich die Ausgabe einer 3D-Innenraumkarte und die Lebenszeichendetektion. Ein besonderer Wert technikethischer Überlegungen besteht darin, dass sie nicht an Legitimität verlieren, wenn sie über den aktuellen Fall hinausgehen, da die aktuellen Forschungen und Entwicklungen in den Bereichen Künstliche Intelligenz und Visualisierung in so hohem Tempo voranschreiten, dass es sich empfiehlt, „auf Vorrat zu denken“ [47], d. h. aktuelle Forschungsprojekte als Anlass zu nehmen, sich in den anschlussfähigen Fragestellungen zu orientieren, Risiken abzuwägen, mögliche Positionen auszuloten und einen Diskurs zu etablieren, der nicht die konkreten Projekte, sondern das gesellschaftliche Großunternehmen der digitalen Transformation im Sinne einer aktiven Widerlegung des Collingridge-Dilemmas durch kontinuierliche Technikfolgenabschätzung begleitet.

FÖRDERHINWEIS

Das Projekt “UAV-Rescue” wird im Rahmen der Zivilen Sicherheitsforschung vom Bundesministerium für Bildung und Forschung gefördert.

LITERATUR

- [1] United Nations Office for the Coordination of Humanitarian Affairs OCHA 2022. International Search and Rescue Advisory Group (INSARAG). Aufgerufen am 19.7.22 unter <https://www.unocha.org/our-work/coordination/international-search-and-rescue-advisory-group-insarag>
- [2] SIFO 2021. Projektumriss UAV-Rescue. Aufgerufen am 6.6.22 unter https://www.sifo.de/sifo/shareddocs/Downloads/files/projektumriss_uav-rescue.html
- [3] R. Herschel. 2020. Aktuelle Entwicklungen in der Verschüttetendetektion: Verschüttete schnell und metergenau aufspüren. Crisis Prevention - Fachportal für Gefahrenabwehr, Innere Sicherheit und Katastrophenhilfe. Aufgerufen am 19.7.22 unter <https://crisis-prevention.de/katastrophenschutz/aktuelle-entwicklungen-in-der-verschuettetendetektion.html#>
- [4] SIFO 2022. Projektumriss I-LOV. Abgerufen am 19.7.22 unter https://www.sifo.de/sifo/de/projekte/schutz-und-rettung-von-menschen/integrierte-schutzsysteme/i-lov/i-lov_node.html
- [5] de Greeff, Joachim; Hindriks, Koen; Neerincx, Mark A.; Kruijff-Korbayova, Ivana. 2015. Human-Robot Teamwork in USAR Environments. In: Adams, Julie A.; Smart, William; Mutlu, Bilge; Takayama, Leila (Hg.): Proceedings of the Tenth Annual ACM/IEEE International Conference on Human-Robot Interaction Extended Abstracts. HRI '15: ACM/IEEE International Conference on Human-Robot Interaction. Portland Oregon USA, 02 03 2015 05 03 2015. New York, NY, USA: ACM, 151–152.
- [6] Kruijff-Korbayová, Ivana; Colas, Francis; Gianni, Mario; Pirri, Fiora; de Greeff, Joachim; Hindriks, Koen; Neerincx, Mark; Ögren, Petter; Svoboda, Tomáš; Worst, Rainer. 2015. TRADR Project: Long-Term Human-Robot Teaming for Robot Assisted Disaster Response. In: Künstliche Intelligenz 29 (2), 193–201. DOI: 10.1007/s13218-015-0352-5.
- [7] Wietfeld, Christian; Daniel, Kai. 2015. Cognitive Networking for UAV Swarms. In: Kimon P. Valavanis und George J. Vachtsevanos (Hg.): Handbook of Unmanned Aerial Vehicles. Dordrecht: Springer Netherlands, 749–780.
- [8] SIFO 2022. Projektumriss Falke. Abgerufen am 19.7.22 unter https://www.sifo.de/sifo/de/projekte/querschnittsthemen-und-aktivitaeten/praxistransfer-und-kompetenzaufbau/kmu-innovativ/falke/falke_node.html
- [9] Ethikkommission Automatisiertes und Vernetztes Fahren. 2017. Bericht. Abgerufen am 19.7.22 unter https://bmdv.bund.de/SharedDocs/DE/Publikationen/DG/bericht-der-ethikkommission.pdf?__blob=publicationFile
- [10] Dusseldorp, Marc. 2021. Technikfolgenabschätzung. In: Grunwald, Armin; Hillerbrand, Rafaela (Hg.): Handbuch Technikethik. 2. Aufl., Stuttgart: J.B. Metzler, 442–446.

- [11] Böschen, Stefan; Grunwald, Armin; Krings, Bettina-Johanna; Rösch, Christine (Hg.). 2021. Technikfolgenabschätzung. Handbuch für Wissenschaft und Praxis. Baden-Baden: Nomos.
- [12] Collingridge, David. 1980. The social control of technology. New York: St. Martin's Press.
- [13] Hillerbrand, Rafaela. 2021. Value Sensitive Design. In: Armin Grunwald und Rafaela Hillerbrand (Hg.): Handbuch Technikethik. 2. Aufl., Stuttgart: J.B. Metzler, 466–471. Online verfügbar unter https://link.springer.com/chapter/10.1007/978-3-476-04901-8_89.
- [14] Hendry, David G.; Friedman, Batya; Ballard, Stephanie. 2021. Value sensitive design as a formative framework. In: Ethics Inf Technol 23 (1), 39–44. DOI: 10.1007/s10676-021-09579-x.
- [15] Mulvenna, Maurice; Boger, Jennifer; Bond, Raymond. 2017. Ethical by Design: A Manifesto. In Proceedings of the European Conference on Cognitive Ergonomics 2017 (ECCE 2017). Association for Computing Machinery, New York, NY, USA, 51–54. <https://doi.org/10.1145/3121283.3121300>
- [16] Harbers, Maaik; de Greeff, Joachim; Kruijff-Korbová, Ivana; Neerincx, Mark A.; Hindriks, Koen V. 2017. Exploring the Ethical Landscape of Robot-Assisted Search and Rescue. In: Maria Isabel Aldinhas Ferreira, Joao Silva Sequeira, Mohammad Osman Tokhi, Endre E. Kadar und Gurvinder Singh Virk (Hg.): A World with Robots, Bd. 84. Cham: Springer International Publishing (Intelligent Systems, Control and Automation: Science and Engineering), 93–107.
- [17] Gadamer, Hans-Georg. 1986. Wahrheit und Methode: Grundzüge einer philosophischen Hermeneutik. 5. Aufl., Tübingen: Mohr.
- [18] Voosen, Paul. 2017. How AI detectives are cracking open the black box of deep learning: As neural nets push into science, researchers probe back. Aufgerufen am 6.6.22 unter <https://www.science.org/content/article/how-ai-detectives-are-cracking-open-black-box-deep-learning>.
- [19] Bach, Sebastian. 2015. On Pixel-Wise Explanation for Non-Linear Classifier Decisions by Layer-Wise Relevance Propagation, Plos one Journal 2015.
- [20] Samek, Wojciech; Montavon, Grégoire; Lapuschkin, Sebastian; J. Anders, Christopher; Müller, Klaus-Robert. 2021. Explaining Deep Neural Networks and Beyond: A Review of Methods an Applications, IEEE Xplore 2021, Proceedings of the IEEE 109 (3).
- [21] Schade, Ulrich; Schwarze, Arne; Sikorski, Lukas; Trautwein, Irmtrud. 2021. Unterstützung von Brandbekämpfung durch Teams von unbemannten Systemen. In: Börteçin Ege und Adrian Paschke (Hg.): Semantische Datenintelligenz im Einsatz. Wiesbaden: Springer Fachmedien, 51–64.
- [22] Vallverdú, Jordi. 2019. Blended Cognition. The Robotic Challenge. Unter Mitarbeit von Vincent C. Müller. Cham: Springer International Publishing AG (Springer Series in Cognitive and Neural Systems Ser, v.12). Online verfügbar unter <https://ebookcentral.proquest.com/lib/kxp/detail.action?doCID=5788968>.
- [23] THW 2022. Wärmebildkamera. Aufgerufen am 1.7. 22 unter <https://www.thw.de/SharedDocs/Ausstattungen/DE/Geraete/te/waermebildkamera.html?nn=925122>.
- [24] THW 2022. Searchcam. Aufgerufen am 1.7.22 unter <https://www.thw.de/SharedDocs/Ausstattungen/DE/Geraete/searchcam.html?nn=925122>.
- [25] THW 2022. Akustisches Ortungsgerät. Aufgerufen am 1.7.22 unter https://www.thw.de/SharedDocs/Ausstattungen/DE/Geraete/akustisches_Ortungsger%C3%A4t.html?nn=925122.
- [26] THW 2022. Biologische Ortung. Aufgerufen am 1.7.22 unter <https://www.thw.de/SharedDocs/Ausstattungen/DE/Geraete/Biologische%20Ortung.html?nn=925122>.
- [27] THW 2022. Trupp UL. Aufgerufen am 1.7.22 unter <https://www.thw.de/SharedDocs/Einheiten/DE/Inland/Tr-UL.html?nn=925114>.
- [28] THW 2022. FGR Ortung. Aufgerufen am 1.7.22 unter <https://www.thw-ratingen.de/thw-ratingen/technischer-zug/fgr-ortung/>.
- [29] Bezogen auf den Einsatz von UAVs sind hier zu nennen: Luftverkehrsordnung, die „Empfehlungen für Gemeinsame Regelungen zum Einsatz von Drohnen im Bevölkerungsschutz“ des BBK, sowie die THW-internen Dokumente „Dienstvorschrift 130: Betrieb unbemannter Luftfahrtsysteme in der Bundesanstalt Technisches Hilfswerk“ und „Stärke- und Ausstattungsnachweis Trupp Unbemannte Luftfahrtsysteme (Tr UL)“.
- [30] THW 2022. 5 Phasen der Rettung. Aufgerufen am 6.6.22 unter <https://download.thw-pirna.de/5phasen.pdf>.
- [31] INSARAG 2022. INSARAG Guidelines. Aufgerufen am 6.6.22 unter <https://www.insarag.org/methodology/insarag-guidelines/>.
- [32] THW 2022. FGR SEEBA. Aufgerufen am 1.7.22 unter <https://www.thw.de/SharedDocs/Einheiten/DE/Ausland/FGR-SEEBA.html?nn=925116>.
- [33] THW-Dokument: Stärke- und Ausstattungsnachweisun Schnell Einsatz Einheit Bergung Ausland (SEEBA).
- [34] THW 2022. FGR SEEBA. Aufgerufen am 6.6.22 unter <https://www.thw.de/SharedDocs/Einheiten/DE/Ausland/FGR-SEEBA.html>.
- [35] THW 2022. Trupp UL. Aufgerufen am 6.6.22 unter <https://www.thw.de/SharedDocs/Einheiten/DE/Inland/Tr-UL.html>.
- [36] THW-Dokument: Stärke- und Ausstattungsnachweisung Trupp Unbemannte Luftfahrtsysteme (Tr UL), 3–13.
- [37] THW-Dokument: Dienstvorschrift DV 1-100, Seite 30.
- [38] Becker, Tobias. 2013. Qualitätsmanagement in der zivilen Gefahrenabwehr: Optimierung von Führungsprozessen durch interaktive SOPs. Dissertation, Fakultät für Maschinenbau, Universität Paderborn. Aufgerufen am 19.7.2022 unter <https://digital.ub.uni-paderborn.de/ubpb/urn/urn:nbn:de:hbz:466:2-12597>.
- [39] Petersen, Katrina. 2017. Visualizing Risk: Drawing Together and Pushing Apart with Sociotechnical Practices. In: Journal of Contingencies and Crisis Management 25 (1), 39–50. DOI: 10.1111/1468-5973.12144.
- [40] Wieglering, Klaus. 2021. Mensch-Maschine-Schnittstelle. In: Armin Grunwald und Rafaela Hillerbrand (Hg.): Handbuch Technikethik. 2. Aufl., Stuttgart: J.B. Metzler, 295–299. Online verfügbar unter https://link.springer.com/chapter/10.1007/978-3-476-04901-8_56.

- [41] Petersen, Katrina; Büscher, Monika. 2016. Mobile work in crisis. In: *Applied Mobilities* 1 (2), 176–192. DOI: 10.1080/2380012-7.2016.1246896.
- [42] Endsley, Mica. 2008. Theoretical Underpinnings of Situation Awareness. A Critical Review. In: Mica Endsley und Daniel J. Garland (Hg.): *Situation awareness analysis and measurement*. Reprinted. Boca Raton, Fla.: CRC Press, 3–32.
- [43] Dorsch Lexikon der Psychologie. 2022. Wechselkosten. Aufgerufen am 19.7.22 unter <https://dorsch.hogrefe.com/stichwort/wechselkosten>
- [44] Rastogi, Charvi; Zhang, Yunfeng; Wei, Dennis; Varshney, Kush R.; Dhurandhar, Amit; Tomsett, Richard. 2022. Deciding Fast and Slow: The Role of Cognitive Biases in AI-assisted Decision-making. In: *Proc. ACM Hum.-Comput. Interact.* 6 (CSCW1), 1–22. DOI: 10.1145/3512930.
- [45] Simanowski, Roberto. 2018. *Stumme Medien. Vom Verschwinden der Computer in Bildung und Gesellschaft*. Berlin: Matthes & Seitz. Online verfügbar unter <https://ebookcentral.proquest.com/lib/kxp/detail.action?docID=5332774>.
- [46] Wehrle, Maren. 2014. *Horizonte der Aufmerksamkeit. Entwurf einer dynamischen Konzeption der Aufmerksamkeit aus phänomenologischer und kognitionspsychologischer Sicht*. München: Wilhelm Fink. Online verfügbar unter <https://elibrary.utb.de/doi/book/10.5555/9783846755662>.
- [47] Mukerji, Nikil; Mannino, Adriano. 2020. *Covid-19: Was in der Krise zählt. Über Philosophie in Echtzeit*. Ditzingen: Reclam.
- [48] Géron, Aurélien. 2019. *Hands-On Machine Learning with Scikit-Learn and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems*. 2nd. edition. Boston: O'Reilly.

KI im Einsatz: Interdisziplinäre Herausforderungen im Projekt UAV-Rescue

Martina Philippi*

Ruhr-Universität Bochum
Institut für Philosophie
Deutschland

martina.philippi@ruhr-uni-bochum.de

Meral Ari

Fraunhofer Institut für Kurzzeitdynamik
Ernst-Mach-Institut
Deutschland

Meral.Ari@emi.fraunhofer.de

Victoria Heusinger-Heß

Fraunhofer Institut für Kurzzeitdynamik
Ernst-Mach-Institut
Deutschland

Victoria.Heusinger@emi.fraunhofer.de

Emily Krüger

Bundesanstalt Technisches Hilfswerk
Deutschland

Emily.Krueger@thw.de

ZUSAMMENFASSUNG

Im Projekt „UAV-Rescue: UAV-getragene Sensorik zur KI-basierten Unterstützung von Rettungsmissionen“ wird an einem unbemannt fliegenden System geforscht, das in schwer zugänglichen Räumen bei der Rettungspriorisierung hilft, indem es den Einsatzkräften eine Visualisierung des Raumes mit darin lokalisierten Lebenszeichen zur Verfügung stellt. Diese neue Technologie wirft verschiedene ethische Fragestellungen auf, die unter anderem die Nutzung von Künstlicher Intelligenz zur Lebenszeichendetektion und die Einbettung der neuen Technologie in die Einsatzpraxis betreffen. Die ethische Begleitforschung des Projekts adressiert diese Fragestellungen in einem interdisziplinären und iterativen Diskurs mit Technikpartnern und Endanwendern. Einige dieser Fragestellungen

werden hier vor dem Hintergrund der eingesetzten KI-Methoden und der bestehenden Einsatzpraxis beim THW vorgestellt.

KEYWORDS

ethics by design, artificial intelligence

LITERATUR

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s).

Veröffentlicht durch die Gesellschaft für Informatik e.V.

in K. Marky, U. Grünefeld & T. Kosch (Hrsg.):

Mensch und Computer 2022 – Workshopband, 04.-07. September 2022, Darmstadt

© 2022 Copyright held by the owner/author(s).

<https://doi.org/10.18420/muc2022-mci-ws10-332>