Abschlussbericht

Projekt TraumAInterfaces

Kapitel 1504 Titel 68605

Eingereicht von:
Institut für Technologie und Innovationsmanagement (TIM), RWTH Aachen University (Übergeordnete Projektleitung)

Zur Vorlage bei:
Deutsches Zentrum für Luft- und Raumfahrt e. V. (DLR) Projektträger, Bundesverwaltungsamt, ZM I 5 - Zuwendungen im Bereich BMG

Beteiligte Projektpartner:

Gefördert durch:
Bundesministerium für Gesundheit
aufgrund eines Beschlusses des Deutschen Bundestages

Inhaltsverzeichnis

I Abbildungsverzeichnis 3

II Tabellenverzeichnis 3

1 Projektverantwortliche 4

3 Einleitung 7

4 Erhebungs- und Auswertungsmethodik 7

5 Durchführung, Arbeits- und Zeitplan 8

6 Ergebnisse 17

7 Gender Mainstreaming Aspekte 25

8 Diskussion der Ergebnisse, Gesamtbeurteilung 26

9 Verbreitung und Öffentlichkeitsarbeit der Projektergebnisse 28

10 Verwertung der Projektergebnisse (Nachhaltigkeit/Transferpotenzial) 30

10.1 Erfahrungen und Ergebnistransfer 30

10.2 Erkenntnisse für die Gesundheitsversorgung/Pflege 31

10.3 Nutzung der aufgebauten Projektstrukturen 31

10.4. Weitere Implikationen der Ergebnisse 32

I Abbildungsverzeichnis

Abbildung 1: Arbeits- und Zeitplan des Projektes 9

Abbildung 2: Prototyp des TraumaAgenten für die Schockraumversorgung 18

Abbildung 3: Prototyp des Traumaregisterbogen-Assistenten 19

Abbildung 4: Eye-Tracking Aufnahme in der Schockraumsimulation 21

Abbildung 5: Prototyp des TraumaAgenten für die Schockraumversorgung mit laufender Echtzeitspracherkennung 24

II Tabellenverzeichnis

Tabelle 1: Eckdaten des Projektes 4

Tabelle 2: Informationen zur Projektleitung 4

Tabelle 3: Übersicht über die Verbundpartner 5

Tabelle 4: Übersicht der Meilensteine (2020-2024) 10

Tabelle 5: Übersicht der veröffentlichten Publikationen 28

1 Projektverantwortliche

In dem vom Bundesministerium geförderten TraumAInterfaces Projekt waren verschiedene Akteure und Akteurinnen der Disziplinen: Medizin, Technologieentwicklung, Psychologie und Innovationsmanagement involviert. Das Projekt wurde entgegen dem planmäßigen Ende im Oktober 2023 bis Januar 2024 verlängert. Mit dieser Verlängerung wurde die Entwicklung des Künstliche Intelligenz (KI)-Systems unterstützt und mögliche Anwendungsszenarien geschärft. Das Projekt TraumAInterfaces umfasst folgende Eckdaten:

Tabelle 1: Eckdaten des Projektes

Titel des Vorhabens	TraumAInterfaces: Entwicklung und Erprobung eines KI-basierten Spracherkennungssystems für die verbale Kommunikation in der Polytraumaversorgung
Förderkennzeichen	ZMVI1
Kapitel	1504
Titel	68605
Förderungszeitraum	01.10.2020 – 31.01.2024

Die übergeordnete Projektleitung oblag dem Institut für Technologie- und Innovationsmanagement (TIM):

Tabelle 2: Informationen zur Projektleitung

Projektleiter	Univ. Prof. Torsten-Oliver Salge, PhD Institut für Technologie- und Innovationsmanagement (TIM), Lehrstuhl für Innovation, Strategie und Organisation
Kontaktadresse der ausführenden Stelle	Adresse: Kackertstraße 7, 52072 Aachen Telefon: 0241 8099196 E-Mail: salge@time.rwth.aachen.de Webseite: www.time.rwth-aachen.de/iso
Rechtsverbindlicher Name	Rheinisch-Westfälische Technische Hochschule Aachen Lehrstuhl für Innovation, Strategie und Organisation
Kontaktadresse der rechtsverbindlichen Stelle	Adresse: Templergraben 55, 52062 Aachen Webseite: www.rwth-aachen.de

Der Verbund wurde von der Rheinisch-Westfälischen Technischen Hochschule Aachen koordiniert und bestand zusätzlich aus vier weiteren Partnern: Fraunhofer Gesellschaft (Fraunhofer-Institut für Intelligente Analyse- und Informationssysteme), Rheinisch Friedrich-Wilhelms-Universität Bonn, Private Universität Witten/Herdecke und dem Universitätsklinikum Aachen. Die Rheinisch Friedrich-Wilhelms-Universität Bonn schied Ende 2022 als Partner aus. Die Aufgaben wurden von der Fraunhofer Gesellschaft übernommen. Die nachstehende Tabelle zeigt eine Übersicht der involvierten Partner und Partnerinnen:

Tabelle 3: Übersicht über die Verbundpartner

Verbundkoordination	Rheinisch-Westfälische Technische Hochschule Aachen
Kooperationspartner 1	Fraunhofer Gesellschaft (Fraunhofer-Institut für Intelligente Analyse- und Informationssysteme)
Kooperationspartner 2	Rheinische Friedrich-Wilhelms-Universität Bonn **(Ausgeschieden Ende 2022)**
Kooperationspartner 3	Private Universität Witten/Herdecke
Kooperationspartner 4	Universitätsklinikum Aachen, AIXTRA

2 Zusammenfassung

Das Projekt TraumAInterfaces befasste sich mit der Frühphase der multidisziplinären Behandlung polytraumatisierter, d.h. mehrfach und lebensbedrohlich verletzter Patienten und Patientinnen, im Schockraum. Die Behandlung polytraumatisierter Patienten und Patientinnen erfordert präzise und schnelle Entscheidungen. Die verbale Kommunikation spielt hierbei eine zentrale Rolle. Sie ist entscheidend für die Verständigung und den Informationsaustausch zwischen den behandelnden Personen, sei es bei der Anmeldung, der Übergabe von der notärztlichen Versorgung an das Schockraumteam oder während der Behandlung und der Übergabe an die Intensivstation. Gerade an den Schnittstellen der Polytraumaversorgungskette kann es zu erheblichen Informationsverlusten, vor allem bei der verbalen Kommunikation, kommen.

Das Projekt TraumAInterfaces zielte darauf ab, diese Schwachstelle aufzudecken und mit Hilfe von KI-Systemen diese Schnittstelle, sowie den Schockraumversorgungsprozess zu unterstützten. Im Rahmen des Projektes wurde ein KI-basiertes Spracherkennungssystem entwickelt, welches die verbale Kommunikation erfasst, transkribiert und strukturiert. Im

Projekt wurden hierzu zwei KI-Use Cases entwickelt, um sowohl den Behandlungsprozess im Schockraum (siehe Abb. 2, 5) als auch die automatische Dokumentation und Protokollierung des Behandlungsprozesses (siehe Abb. 3) zu unterstützen. Um die Effizienz und Genauigkeit der Informationsübertragung zu verbessern und somit die Qualität der Patienten- und der Patientinnenversorgung zu erhöhen, wurde ein iterativer Entwicklungsansatz genutzt, der es erlaubte, generierte Erkenntnisse direkt in den Entwicklungsprozess einzubinden und die ersten Prototypen iterativ zu verbessern.

Insgesamt liefert das TraumAInterfaces Projekt Best Practice Ansätze für die Medizin KI-Systeme auch in dynamischen Anwendungsumgebungen systematisch entwickelt und in die klinische Umgebung implementiert werden können. Die Projektergebnisse zeigen, dass insbesondere Large Language Models (LLMs) große Potenziale bieten, den Behandlungsprozess im Schockraum nachhaltig zu verbessern und medizinisches Personal zu entlasten. Um solche Systeme flächendeckend nutzen zu können, müssen insbesondere die damit verbundenen legalen und datenschutzrechtlichen Fragestellungen zeitnah sowohl in der Medizin als auch in einem gesamtgesellschaftlichen Diskurs geklärt und durch die Politik in einen belastbaren gesetzlichen Rahmen überführt werden.

Der nachfolgende Bericht fasst den Projektverlauf zusammen, beleuchtet die verschiedenen Arbeitspakete und Meilensteine, legt eine Gesamtbeurteilung der Ergebnisse vor und liefert Ideen und Ansätze zur (zukünftigen) Verwertung der Projektergebnisse.

3 Einleitung

Das Verbundprojekt TraumAInterfaces war ein vom Bundesministerium für Gesundheit (BMG) gefördertes Forschungs- und Entwicklungsprojekt zum Einsatz von KI-Anwendungen in der Polytraumaversorgung. Im Vordergrund stand die realitätsnahe Erprobung und Optimierung eines KI-basierten Spracherkennungssystem zur Erfassung, Transkription und Strukturierung der verbalen Kommunikation während der Patientenversorgung im Schockraum. Im Verlaufe des Projektes wurden folgende übergeordnete Forschungsfrage beantwortet:

*Wie können **intelligente Spracherkennungs- und Verarbeitungssysteme***

> *(1) die **Entscheidungsträger und Entscheidungsträgerinnen** der Schockraumversorgung bestmöglich **unterstützen***
>
> *(2) die **Informationsweitergabe** auch **über Schnittstellen** hinweg **optimieren** und helfen, (3) den individuellen Patienten und die individuelle Patientin in den Kontext der verfügbaren medizinisch wissenschaftlichen Evidenz und Registerdaten einzuordnen, so dass die **Qualität** und die **Sicherheit** des **Versorgungsprozesses** polytraumatisierter Patienten und Patientinnen **verbessert** werden?*

Die nachfolgenden Kapitel beschreiben den Projektverlauf unter besonderer Berücksichtigung der durch die COVID-19 Pandemie und die Veröffentlichung des LLMs Chat GPT 3 bedingten Veränderungen des Projektverlaufes.

4 Erhebungs- und Auswertungsmethodik

Im Kern des Projekts und damit der methodischen Vorgehensweise stand die realitätsnahe Erprobung verschiedener Iterationen des Prototyps eines Spracherkennungssystems zur automatisierten Erfassung, Transkription und Strukturierung der verbalen Kommunikation während der Patientenversorgung im Schockraum. Dazu wurden die Prinzipien der agilen Entwicklung genutzt, welche sich am Build-Measure-Learn-Zyklus orientieren und dadurch eine iterative System-Weiterentwicklung und Verbesserung ermöglichen. Dafür wurde ein erprobbarer Prototyp bestehend aus Spracherkennung zur Dokumentationserfassung, Nutzerschnittstellen und Output-Templates erstellt (**Build**). Dieser wurde dann realitätsnah erprobt und hinsichtlich Performanz, Akzeptanz und klinischer Wirkung gemessen (**Measure**). Daraus wurden Erkenntnisse zur Weiterentwicklung des Systems, zu notwendigen organisationalen Gelingensbedingungen für die wirtschaftliche Tragfähigkeit und Ideen für eine flächendeckende Implementierung in der klinischen Praxis abgeleitet (**Learn**).

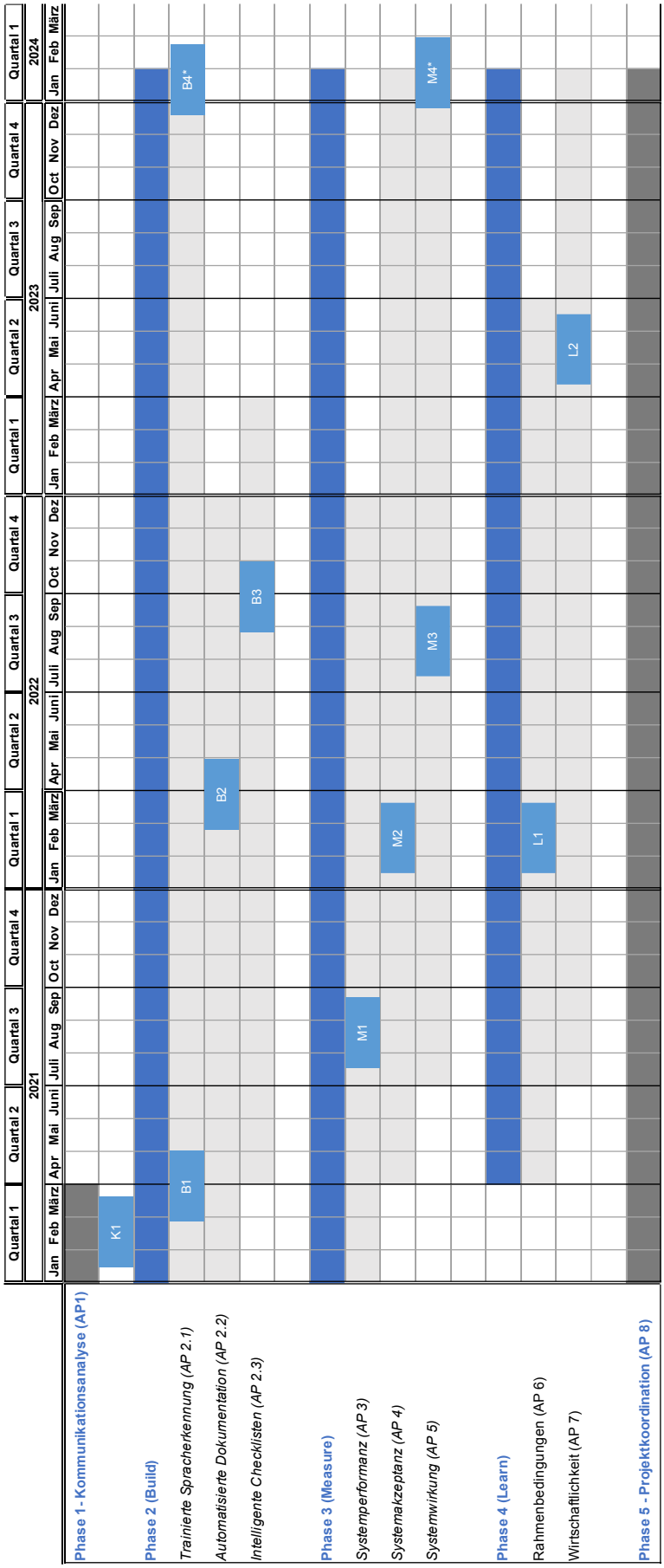
Um das Projektziel und die kontinuierliche Weiterentwicklung und Verbesserung des Systems sicherzustellen, wurden verschiedene Arbeitspakete und Meilensteine formuliert, die im

Rahmen des Projektes erreicht werden mussten. Diese Meilensteine dienten als Indikatoren zur Messung der Zielerreichung.

5 Durchführung, Arbeits- und Zeitplan

Die Durchführung des Projektes orientierte sich an dem vorgestellten Build-Measure-Learn-Ansatz. Der zugrundeliegende Arbeits- und Zeitplan wird in der nachstehenden Abbildung zusammengefasst.

Abbildung1: Arbeits- und Zeitplan des Projektes



Das Projekt wurde in fünf Phasen unterteilt. Die Phasen gliederten sich in verschiedene Arbeitspakete. Insgesamt wurden im Rahmen des Projektes alle Arbeitspakete erfolgreich bearbeitet. Jedes dieser Arbeitspakete umfasste eine oder mehrere Meilensteine, die im Projektverlauf erreicht wurden. Die Arbeitspakete und die dazugehörigen Meilensteine sind in Tabelle 4 dargestellt. Zur Erreichung der verschiedenen Meilensteine wurde verschiedene Methoden und Daten genutzt.

Tabelle 4: Übersicht der Meilensteine (2020-2024)

Arbeitspaket	Meilenstein	Beschreibung	Methodik und Daten
Arbeitspaket 1	Meilenstein K1	Identifizierung von kritischen Schnittstellen im Kommunikationsprozess und Quantifizierung von Informationsverluste	Literaturrecherche Nicht teilnehmende Beobachtung Interviews Fragebogen
Arbeitspaket 2	Meilenstein B1	Training des Spracherkennungssystem mit geeigneten klinischen Sprachsamples	Sprachtranskripte Sprachsamples Versorgungsprotokolle
Arbeitspaket 2	Meilenstein B2	Klinische Informationsextraktion aus Sprachtranskripten zur automatisierten Dokumentation	Trainingssimulationen (offline und online Wortwolken Prototypische Versorgungsabläufe GPT3.5, GPT4
Arbeitspaket 2	Meilenstein B3	Einsatzbereites System für die Generierung intelligenter Checklisten zur Testmöglichkeit in Live-Anwendung im PoC	Trainingssimulationen Befragungen
Arbeitspaket 3	Meilenstein M1	Untersuchung der Performanz der Spracherkennung im Vergleich zum Menschen in Experimenten im Computerlabor	Kognitionspsychologische Online-Experimente mit Experten Vergleich mit der Word Error Rate (WER) des Prototypens
Arbeitspaket 4	Meilenstein M2	Untersuchung und Quantifizierung der Akzeptanz des klinischen Spracherkennungssystems aus Sicht klinischer Nutzer und Nutzerinnen	Behandlungssimulation im Schockraum Experimente Fragebogen Interviews
Arbeitspaket 5	Meilenstein M3	Quantifizierung der Wirkung des Spracherkennungssystems auf für die Qualität des Schockraumprozesses relevante Parameter	Simulationen Manuelle Datengenerierung durch mediz. Experten und Expertinnen Befragungen GPT3.5, GPT4

Arbeitspaket	Meilenstein	Beschreibung	Methodik und Daten
Arbeitspaket 6	Meilenstein L1	Identifizierung der Anforderungen an die erforderlichen Rahmenbedingungen zum Einsatz des klinischen Spracherkennungsystems	Workshop
Arbeitspaket 7	Meilenstein L2	Ableitung der Implikationen für die wirtschaftliche Tragfähigkeit sowie Verbreitung und Implementierung des Systems in Traumazentren	Literaturrecherche Business Modell Workshop
Projekt-verlängerung	Meilenstein B4*	Systemweiterentwicklung und -machbarkeit in der Praxis	Lokale Large Language Models (LLMs)
Projekt-verlängerung	Meilenstein M4*	Praktische Anwendung und Systemimplementierung	Echzeittranskripte

Insgesamt ist festzustellen, dass alle skizzierten Meilensteine mit Ausnahme der Meilensteine M1 und M3 im angegebenen Zeitraum erreicht worden sind. Durch das verspätete Ethikvotum sowie die anhaltende Corona Pandemie war eine Verschiebung des Meilensteins M1 erforderlich. Der Meilenstein M1 wurde zu einem späteren Zeitpunkt im Dezember 2021 erreicht.

Der Meilenstein M3 konnte nicht im ursprünglich geplanten Umfang erreicht werden, da aufgrund der o.g. Einschränkungen vor allem infolge der Corona Pandemie nicht der vollständige Versorgungsgprozess abgebildet werden konnte und mithin die Wirkung des Spracherkennungssystems auf die Qualität des Schockraumprozesses nicht abgebildet werden konnte. Um zufriedenstellende Projektergebnisse generieren zu können und das KI-System zu verbessern, wurde das Projekt im Oktober 2023 um weitere 4 Monate verlängert. Hierzu wurden zwei weitere Meilensteine B4* und M4* formuliert.

Nachfolgende werden die verschiedenen Meilensteine kurz skizziert und die entsprechenden Outcomes und Herausforderungen näher erläutert.

Meilenstein K1

Im Rahmen des ersten Meilensteins wurden eine Vielzahl an Datenerhebungsstrategien genutzt, um den Ist- und den Soll-Zustand im Schockraum möglichst genau abzubilden. Dazu wurden eine Literaturrecherche und eine Pilotstudie durchgeführt, um den Kommunikationsprozess im Schockraum darzustellen und Informationsverluste sowie Stress-

und Störfaktoren zu identifizieren. Des weiteren wurden diese Erkenntnisse durch Interviews und Befragungen (Fragebögen) analysiert und validiert. Hierbei zeigte sich, dass Informationsverluste sowie Stör- und Stressfaktoren den Behandlungsprozess negativ beeinflussen. Die generierten Informationen unterstützten den nächsten Schritt im Projekt und gaben erste Inputs wie der Informationsfluss im Schockraum verbessert werden kann und wie komplexe Informationen entsprechend den Bedürfnissen der User besser dargestellt werden können.

Meilenstein B1

Ziel des Meilensteins war es, einen ersten Prototyp des Spracherkennungssystems zu entwickeln und iterativ zu verbessern. Das System wurde zunächst mit einem umfangreichen Trainingskorpus aus der Rundfunkdomäne trainiert. Zur Anpassung an die medizinische Domäne wurden das Sprachmodell sowie das Aussprache- und Akustikmodell mit medizinischen Fachbegriffen, Texten und Vorlagen aus Schockraum-Simulationen weiter trainiert. Zusätzlich wurde medizinisches Fachvokabular aus der Polytraumaversorgung von ärztlichem Personal des Klinikums Köln-Merheim aufgezeichnet. Diese Aufzeichnungen schwieriger Fachvokabeln wurden phonetisch annotiert, um das Aussprachemodell zu verbessern. Der Trainingskorpus wurde durch diese Lautbeschreibungen erweitert, um die Erkennung schwieriger Fachwörter zu verbessern. Das bereits trainierte Akustikmodell wurde mit Audiodateien aus simulierten Schockraumszenarien weiter adaptiert.

Meilenstein M1

Das Ziel des Meilenstein M1 war es, die Performanz des Spracherkennungsassistenten gegen die menschliche Benchmark zu testen. Dazu wurde die Genauigkeit der Spracherkennung mit unterschiedlichen Sprechern und Umgebungsgeräuschen im Vergleich zum Menschen überprüft, um somit die Leistungsfähigkeit zu quantifizieren und Prioritäten zur Optimierung und Weiterentwicklung der nächsten Iteration des Prototypens bestimmen zu können. Die Performanz des trainierten Spracherkennungssystems wurde anhand von Sprachsamples aus den simulierten Schockraumszenarien getestet. Einzelne Abschnitte der Tonaufnahmen wurden mit dem trainierten Spracherkennungssystem transkribiert. Anschließend wurden die von dem Spracherkennungssystem generierten Transkripte mit den erstellten Ground-truth-Transkripten der jeweiligen simulierten Schockraumszenarien verglichen und die Wortfehlerrate (WER%) erfasst. Dieses Verfahren wurde ebenfalls mit menschlichen Benchmarks durchgeführt. In kognitionspsychologischen Online-Experimenten haben

Experten (medizinisches Fachpersonal mit Schockraumerfahrung) einzelne Abschnitte der Tonaufnahmen von verschiedenen simulierten Schockraumszenarien transkribiert. Die Daten wurden mit der gleichen Methode ausgewertet und die Wortfehlerrate (WER%) bestimmt, um einen bestmöglichen Vergleich von Mensch zu Spracherkennungssystem zu gewährleisten.

Meilenstein B2

Das Ziel dieses Meilensteins war die Entwicklung eines Systems zur klinischen Informationsextraktion aus Sprachtranskripten, das zur automatisierten Dokumentation genutzt werden kann und für eine realitätsnahe Evaluation bereit ist. Eine große Herausforderung bei der Entwicklung dieses Systems und des Systems aus Meilenstein B3 war die Verfügbarkeit von Trainingsdaten. Pandemiebedingt (Versammlungsverbot) konnten nur wenige Schockraumsimulationen durchgeführt werden. Damit stand nur eine geringe Zahl an Sprachtranskripten zur automatisierten Befüllung des Traumaregisterbogens zur Verfügung. Dazu wurde der Traumaregisterbogen als HTML-basiertes Formular umgesetzt, der per REST-Calls befüllt werden kann. Initial wurden aufgrund der unzureichenden Datenbasis regelbasierte Verfahren eingesetzt (bspw. Durch prototypische Abläufe o.Ä.), um eine Automatisierung der Formular-Befüllung zu erreichen. Später im Projekt wurden große Sprachmodelle wie GPT3.5 und GPT4 öffentlich verfügbar, woraufhin diese von unserem technischen Partner des Fraunhofer IAIS eingebunden und in den Use Cases evaluiert wurden.

Meilenstein B3

Das Ziel des Meilensteins B3 war die Entwicklung eines Systems zur Ableitung intelligenter Checklisten, das einsatzbereit ist und in einer Live-Anwendung im PoC getestet werden kann. Im Rahmen dieses Meilensteins wurde mit den Konsortialpartnern, insbesondere mit den klinischen Partnern, eine Live-Anzeige (graphisches User Interface, GUI) für den Schockraum entworfen. Zusätzlich wurde ein Steuerungsbildschirm entworfen, über den die Inhalte des GUI, die die Nutzer und Nutzerinnen sehen, manuell eingegeben werden können. Im Sinne eines reversed-engeneering Ansatzes war es so möglich den Zeitpunkt, an dem eine Information auf dem GUI eingespielt wurde mit dem Text, d.h. mit der verbalen Kommunikation, zu synchronisieren. Zur Entwicklung eines solchen Systems wurden vortrainierte große Sprachmodelle wie GPT3.5 und GPT4 eingesetzt. Dazu wurden sogenannte Agenten entwickelt, welche eigenständig auf Basis des Transkripts den Steuerungsbildschirm steuern und damit die Live-Anzeige beeinflussen. Diese Agenten

müssen autonom verstehen, welche Informationen für das Personal im Schockraum relevant sein könnten. Darüber hinaus müssen sie robust gegenüber Transkriptionsfehlern sein und entscheiden ob, Informationen überholt sind und/oder gelöscht oder geändert werden müssen.

Meilenstein M2

Im Rahmen von Schockraumsimulationen und Experimenten wurde die Akzeptanz des Spracherkennungssystem aus Sicht der Nutzer und Nutzerinnen untersucht. Hierzu wurden sowohl unterschiedliche medizinische Szenarien implementiert als auch die Gruppengröße der Schockraumteilnehmenden (Zweier- und Dreiergruppen variiert. Um die Interaktion zwischen Menschen und Maschinen besser zu verstehen, wurden die Entscheidungsträger und Entscheidungsträgerinnen (Trauma Leader, Anästhesie) mit Eye-Tracking Brillen ausgestattet. Diese Daten erlauben es die Wahrnehmung der durch das KI-basierte Spracherkennungssystem generierten Informationen zu erfassen. In diesen Studien konnten Unterschiede in der Praktikabilität und Gründe für die Nutzung (z.B. Reduzierung von Behandlungsunsicherheiten sowie Nicht-Nutzung (z.B. Ungewohnte Handhabung) erfasst werden. Zusätzliche Daten wurden im Rahmen von Interviews und Fragebögen erhoben, um Feedback der Probanden und Probandinnen im Hinblick auf die Systempraktikabilität, das User Design und der dargestellten Informationen zu erfassen. Die generierten Daten bilden die Grundlage für die systematische Anpassung und Verbesserung des KI-basierten Spracherkennungssystems.

Meilenstein M3

Das Ziel dieses Meilensteines war die Quantifizierung der Wirkung des Spracherkennungssystems auf für die Qualität des Behandlungsprozesses im Schockraum relevanten Parameter. Hierzu wurden in enger Kooperation mit dem Innovationstechnologie- und Psychologiepartner im Rahmen von Befragungen von im Schockraum des Klinikums Köln Merheim regelhaft tätigen Ärzte und Ärztinnen und Schockraumbeobachtungen zunächst mögliche Angriffspunkte für eine qualitätsverbessernde Intervention identifiziert.

Darüber hinaus wurden im Rahmen von Schockraumsimulationen, in denen das durch Spracherkennung augmentierte Display eingesetzt wurde, das visuelle Nutzungsverhalten der KI durch die Schockraumprotagonistinnen mit Eyetracking Brillen beobachtet. Da die Entwicklung des Spracherkennungssystems in Verbindung mit dem LLM zu diesem Zeitpunkt noch nicht abgeschlossen war wurde auf einen manuell gesteuerten Prototyp der KI (“Fake KI”) zurückgegriffen, mit der im Rahmen der simulierten Szenarien die Informationen der KI in das Szenario eingespielt wurden.

Meilenstein L1

Ziel des Meilensteins war es, Anforderungen an die erforderlichen organisationalen Rahmenbedingungen zum bestmöglichen Einsatz des klinischen Spracherkennungssystems zu identifizieren. Zur Realisierung dieses Meilensteins fand ein zweistündiger Online-Workshop statt. Ziel war es, essenzielle Anforderungen an die erforderlichen Rahmenbedingungen zum Einsatz des klinischen Spracherkennungsystems zu identifizieren. Der Workshop wurde vom Team des AIXTRA geleitet. Die Teilnehmenden kamen aus verschiedenen Bereichen der Uniklinik RWTH Aachen, die vorab als relevante Institutionen für das betreffende Thema identifiziert wurden (Ethikkommission, Datenschutzbeauftragter, Rechtsabteilung, Stabsstelle Qualitätssicherung, Informationssicherheitsbeauftragte). Zur besseren Vorbereitung erhielten alle Teilnehmenden vorab den Prüfplan des Projekts. Die Ergebnisse wurden in Form einer interaktiven Mindmap festgehalten und visualisiert. Die identifizierten Anforderungen stellen zentral wichtige Punkte dar, die bei der Implementierung und beim Einsatz des klinischen Spracherkennungssystems zu beachten sind.

Meilenstein L2

Um Implikationen für die wirtschaftliche Tragfähigkeit und möglichst flächendeckende Verbreitung und Implementierung des Systems in Traumazentren abzuleiten wurde auf verschiedene Datenquellen zurückgegriffen. Zum einen wurden relevante Kosten und Nutzenfaktoren, die ein solches System mit sich bringt, auf Basis von umfassender Literaturrecherche und Diskussionen mit Technologieexperten und -expertinnen definiert. Darüber hinaus wurde im Rahmen eines Workshops geeignete Geschäftsmodelle entwickelt, die Möglichkeiten für eine flächendeckende Implementierung des im Rahmen von TraumAInterfaces skizzierten Systems aufzeigen. Die Betrachtung und Validierung von Geschäftsmodellen bietet das Potenzial Wertversprechen, Zielgruppen, notwendige Ressourcen und Partner sowie Kostenstrukturen und Monetarisierungsoptionen systematisch zu entwickeln und anzupassen.

Meilenstein B4*

Für die klinische Implementierung des Systems ist die Berücksichtigung datenschutzrechtlicher Aspekte von besonderer Relevanz. Daher sollte in diesem Meilenstein überprüft werden inwiefern on-premise LLMs für die Steuerung der Live-Anzeige im Schockraum (B3) und die Ausfüllung des Traumaregisterbogens grundsätzlich geeignet sind (B2). Hierzu wurden verschiedene lokale Sprachmodelle wie Mistral, Mixtral und Llama2 analysiert. Ebenfalls wurde die Übertragbarkeit auf andere Sprachen als Deutsch untersucht. Die Ergebnisse werden im nächsten Kapitel näher erläutert.

Meilenstein M4*

Um das System für die reale Anwendung testbar zu machen, musste eine Live-Spracherkennung implementiert werden. In den vorherigen Meilensteinen wurde ein fertiges Sprachtranskript an das System gegeben, welches nicht den reellen Testbedingungen entspricht. Um die Testbarkeit in Realszenarien zu ermöglichen, wurde das System so weiterentwickelt, dass Transkriptionen in Echtzeit erstellt und an die Sprachmodelle weitergegeben werden können.

6 Ergebnisse

Im Nachfolgenden werden die Projektergebnisse entsprechend der verschiedenen Arbeitspakete und Meilensteine dargestellt.

Meilenstein K1: Prozessanalyse der Traumaversorgung

Die Ergebnisse der zuvor skizzierten Arbeitspakete unterstreichen das umfassende Potenzial von KI-basierten Spracherkennungssystemen in der Traumaversorgung. Im Rahmen einer detaillierten Prozessanalyse mittels Interviews, Umfragen und Beobachtungen konnte festgestellt werden, dass insbesondere entlang der Schnittstellen Informationsverluste und Kommunikationsprobleme auftreten können. Beispielsweise sind patientenspezifische Daten und aktuelle Gesundheitsparameter die am häufigsten nachgefragten Informationen. Für den Informations- und Kommunikationsfluss spielen Stress- und Störfaktoren eine relevante Rolle. Diese wurden im Rahmen des Projektes systematisch erhoben. Etwa gaben Akteure und Akteurinnen des Schockraumversorgungsprozesses an, dass es verschiedene exogene (bspw. interdisziplinäre Interaktion, Patient/Patientin) und endogene Faktoren (bspw. Erfahrungslevel, Aufregung) gibt, die den Versorgungsprozess beeinflussen bzw. sie sich durch diese Faktoren gestresst fühlen.

Meilensteine B1, B2, B3: Prototypenentwicklung

In Anlehnung an den Projektantrag wurden zwei verschiedene Anwendungsfälle für den KI-basierten Spracherkenner entwickelt: Der TraumaAgent und der Traumaregisterbogen-Assistent. Im Rahmen des Projektes wurde das Fraunhofer IAIS-Spracherkennungssystem als Grundlage genutzt und systematisch mit Hilfe von selbsterhobenen Sprachdaten, Wortwolken etc. weiterentwickelt und getestet. Die Ergebnisse zeigen, dass das System in der frühen Entwicklungsphase relevante Fachbegriffe erkennen kann, jedoch nicht fehlerfrei. Für ein fehlerfreies System muss das Modell auf einer großen Menge an Daten und variablen Gesprächskontexten trainiert werden. Aufgrund der COVID-19-Pandemie war es jedoch nicht möglich, eine ausreichend große Anzahl von Simulationen aufzuzeichnen. Um die Aufnahmequalität zu verbessern, wurden im Projektverlauf verschiedene Mikrofone getestet.

Für die weiteren Simulationen im Meilenstein M1 wurden die SHURE SLXD14DE-Mikrofone durch AUDAC CMX 826-Mikrofone ersetzt, was die Aufnahmequalität durch Reduktion von Hintergrundgeräuschen und Rückkopplungen verbesserte.

Im Rahmen des TraumAInterfaces Projektes wurden zwei KI-Use Cases umgesetzt: die Live-Anzeige wichtiger Informationen im Schockraum und die Dokumentation im Traumaregisterbogen. Der ursprüngliche Projektplan sah vor, durch eine ausreichend große Menge an Simulationen Trainingsdaten für KI-Modelle aufzubauen. Dazu wurden die Simulationsteams mit Mikrofonen ausgestattet und die Kommunikation während der Behandlungssimulationen aufgezeichnet. Zusätzlich wurde über ein speziell entwickeltes Tool erfasst, welche für die Behandlung relevanten Informationen genannt wurden und zu welchem Zeitpunkt dies geschah. Außerdem wurde das korrekte Ausfüllen des Traumaregisterbogens für den jeweiligen Fall der Simulation erfasst. Da aufgrund der COVID-19-Pandemie keine ausreichende Anzahl an Simulationen durchgeführt werden konnte, ist das Entwicklungsteam in beiden Use Cases auf die Nutzung von großen Sprachmodellen ausgewichen.

Use Case 1: Agentengesteuerte Live-Anzeige im Schockraum

Im ersten Use Case wurde eine Live-Anzeige für die Behandlungsunterstützung im Schockraum entwickelt. Das, in mehreren Iterationen entwickelte, Frontend (GUI) für die Schockraumanzeige (siehe Abb. 2, Abb. 5) zeigt Informationen wie das ABCDE-Schema, die Notarztübergabe und generelle Patienten und Patientinnen- und Behandlungsinformationen an. Ein zusätzlicher "Control Screen" ermöglichte die manuelle Erfassung von Informationen und das Sammeln von Trainingsdaten. Die aufgezeichneten Dialoge wurden per Spracherkennung transkribiert. Diese Dialogtranskripte konnten dann über Zeitstempel mit den Trainingsdaten verknüpft werden. Zu diesem Zweck wurden Whisper-Modelle verwendet, da diese in der größten Variante eine gute Erkennungsqualität erzielten. Unter Zuhilfenahme des großen Sprachmodells GPT-4 war es möglich, die Problematik der geringen Anzahl an Sprachaufnahmen auszugleichen. Auf Basis von GPT-4 und langchain wurde ein LLM-Agent entwickelt, der als Umgebung den aktuellen Zustand des Frontends und die Ausgaben aus der Sprachtranskription erhält. Der Agent erhält Zugriff auf die REST-Schnittstelle des Schockraum-Frontends. Zu jedem Zeitpunkt kann der Agent sich entscheiden keine Aktion durchzuführen, neue Informationen hinzuzufügen, vorhandene Informationen zu verändern oder diese zu löschen.

Dabei muss der Agent entscheiden, welche Informationen an welcher Stelle im Screen platziert werden sollen und wie diese zusammengefasst werden.

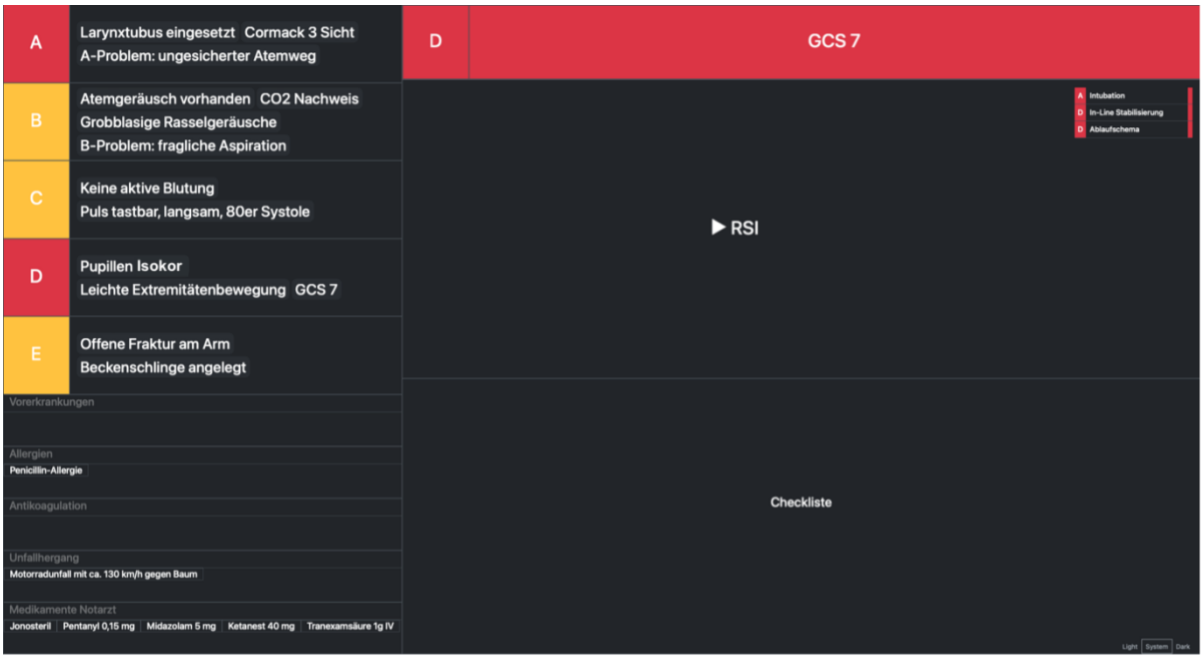


Abbildung 2: Prototyp des TraumaAgenten für die Schockraumversorgung

Use Case 2: Traumaregisterbogen-Assistent zum automatischen Ausfüllen des Traumaregisterbogens

Für die Dokumentation von Informationen im Traumaregisterbogen wurde ein weiterer LLM-Agent entwickelt. Dieser nutzt das vollständige Dialogtranskript aus dem Schockraum in Verbindung mit dem Traumaregisterbogen, um relevante Informationen zu extrahieren (siehe Abb. 3). Dazu wurde eine „Prompt-Template-Engine“ entwickelt, mit deren Hilfe Single-Choice, Multiple-Choice und Freitextfelder auf den Formularen abgebildet werden können. Diese Engine erlaubt es, das Formular in ein für das Modell verständliches Format zu überführen.

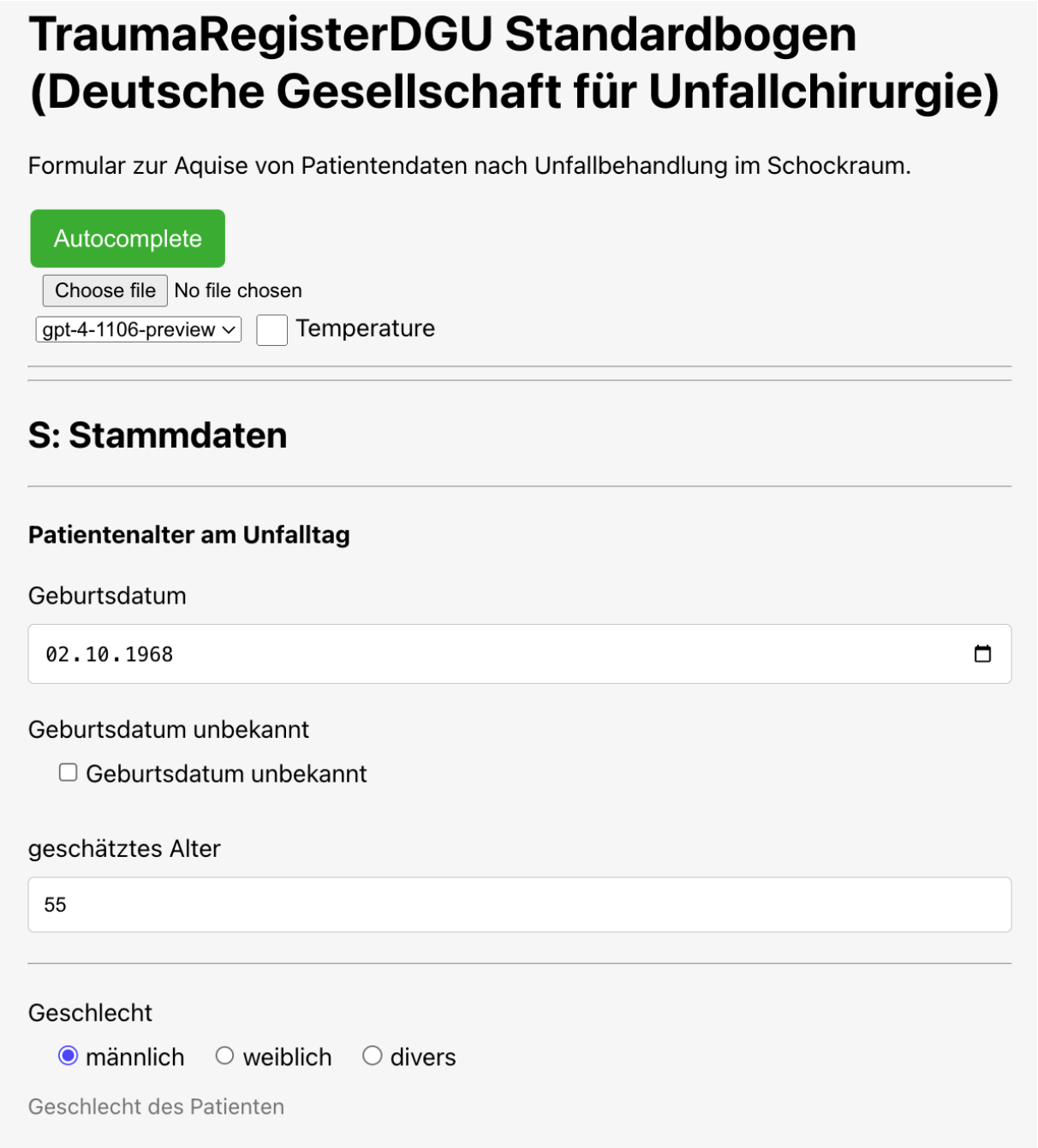


Abbildung 3: Prototyp des Traumaregisterbogen-Assistenten

Meilenstein M1: Systemperformanz

Die Ergebnisse zeigten eine bessere Performanz der menschlichen Benchmarkt im Vergleich zum Prototyp. Um die Performanz und die Qualität des Spracherkennungssystems zu steigern, wurden im weiteren Projektverlauf die drei Module des Spracherkennungssystems weiter trainiert und verbessert. Zusätzliche Simulationen wurden mit den neuen AUDAC Mikrofonen durchgeführt und generiert, um Hintergrundgeräusche und Rückkopplungen zu unterdrücken.

Meilenstein M2: Systemakzeptanz

Die umfassende Testung des KI-Systems im Feld zeigt, dass Endnutzer und Endnutzerinnen das KI-Spracherkennungssystem als grundsätzlich hilfreiches Unterstützungssystem ansehen. Allerdings ist zu diskutieren, in welcher Art und Weise das System den Behandlungsprozess am besten unterstützten kann. Hier lässt sich ein Spannungsfeld zwischen Augmentierung und Automatisierung von Aufgaben (beispielsweise Zusammenfassung Versorgungskomponenten vs. konkrete Handlungsempfehlungen) feststellen. Zudem ist festzuhalten, dass insbesondere das Design und die Anzahl der dargestellten Informationen optimiert werden sollte. Die Auswertungen zeigen, dass insbesondere Informationszusammenfassungen und Gedächtnisstützen als vielversprechend wahrgenommen werden. Diese Erkenntnisse dienten zur Weiterentwicklung und Verbesserung des Systems, um insbesondere zielgerichtete Trainings und Schulungen für die Anwendung und Interaktion mit KI-Systemen entwickeln zu können.

Meilenstein M3: Systemwirkung

Im Rahmen der Befragungen zeigte sich, dass die Versorgung Schwerverletzter mit einem Cognitive (over)load verbunden ist. Sowohl endogene als auch exogene Faktoren spielen hier eine Rolle. Von den Befragten wurden sämtliche abgefragte Faktoren auf einer Skala von 1-6 mit > 4 bewertet. Diese Beobachtung ist dahingehend von großer Relevanz, als dass sie zeigt, dass ein großer Unterstützungsbedarf besteht. Allerdings lassen sich aus den Antworten der Befragten keine eindeutigen Ansatzpunkte für KI basierte Interventionen ableiten, weil sämtliche abgefragten Bereiche (von der Unterstützung durch Protokollierung von Informationen, bis hin zu tatsächlichem Decision Support) von den Probanden und Probandinnen als relevant eingestuft wurden. Die ursprüngliche Annahme, dass sich eindeutige Ansatzpunkte für eine KI-Intervention identifizieren lassen, die bei adäquater

Abbildung in der KI zu messbaren Verbesserungen der Prozessqualität führen muss, deshalb als unterkomplex verworfen werden.

Das mit einem Eye-Tracking System (Tobii Glasses) protokollierte Nutzungsverhalten des KI-Monitors hat allerdings gezeigt, dass die an der Simulation Teilnehmenden den Monitor regelhaft visuell aufgesucht haben, um Informationen zu erhalten, so dass eine grundsätzliche Akzeptanz für derartige Informationen angenommen werden kann.

Abbildung 4: Eye-Tracking Aufnahme in der Schockraumsimulation

Neben den subjektiven Angaben der Befragten zeigte sich im Rahmen der Schockraumbeobachtungen, dass Informationsverluste bzw. das Erfragen/Nachfragen von Informationen (z.B. Vorerkrankungen) ein relevantes Problem im Versorgungsprozess darstellen (Tjardes et al. 2023, eigene Publikation). Diese objektiven Beobachtungen fügen

den o.g. subjektiven Angaben der Befragten eine weitere Dimension hinzu, die potenzielle Ansatzpunkte für KI basierte Interventionen in den Behandlungs- und Entscheidungsfindungsprozess im Schockraum aufzeigt.

In der Zusammenschau konnte der Meilenstein M3, im Sinne von nachweisbaren Verbesserungen der Prozessqualität durch den Einsatz spracherkennungsbasierten KI-Interventionen nicht erreicht werden. Die bis zu diesem Meilenstein erhobenen Daten, in Verbindung mit den durchgeführten Simulationsszenarien, haben allerdings erstmals sowohl qualitativ als auch quantitativ belegt, dass wirkungsvolle KI-Tools und Interventionen nicht anhand medizinischer Prozessparameter (z. B. Leitlinienadhärenz, Algorithmusadhärenz, Time to...) entwickelt werden können. Vielmehr können KI-basierte Tools nur dann wirkungsvoll sein, wenn sie den Unterstützungsbedarf des einzelnen Anwenders adressieren. Das bedeutet, dass es zur Verbesserung eines spezifischen Prozessparameters u.U. verschiedener Tools bedarf, derer sich der einzelne Anwender entsprechend seines persönlichen Unterstützungsbedarfes bedienen kann.

Meilenstein L1: Systemanforderungen

Der Workshop zu Systemanforderungen behandelte die folgenden als relevant erachteten Themenbereiche:

1. Hürden bei der Einführung des klinischen Spracherkennungssystems;
2. Vorteile durch die Einführung des klinischen Spracherkennungssystems;
3. Unterstützungswünsche und -möglichkeiten für die bestmögliche Nutzung;
4. In die Umsetzung einzubeziehende Personen(gruppen).

Bei 1) wurde als Hürde insbesondere die Erhebung, Verarbeitung und Speicherung personenbezogener Daten angesehen. Hierbei wurde zwischen Patienten und Patientinnen und Mitarbeitenden differenziert und unterschiedliche Möglichkeiten der Lösung diskutiert (Rechtsnormen, Einwilligung nach vorheriger Aufklärung etc.). Ferner wurde der Themenkomplex Medizinprodukt-Entwicklung und Zulassung (inklusive Voraussetzungen) diskutiert. Themenbereich 2) brachte interessante neue rechtliche Aspekte hinsichtlich der Behandlung zu Tage. Die Themenbereiche 3) und 4) ergaben eine Reihe von Institutionen und Personengruppen, die bei der Umsetzung von Relevanz sind, sowie Möglichkeiten und Strategien zu deren Einbindung. Besonders relevant erschien hierbei der Aspekt der Einbeziehung von Mitarbeitenden, die mit dem Spracherkennungssystem arbeiten sollen, und deren Aufklärungsbedarf.

Meilenstein L2: Wirtschaftlichkeit

Die Kosten von KI-basierten Sprachassistenten in der Schockraumversorgung umfassen hohe initiale Investitionen für Hardware, Software und Integration sowie laufende Kosten für Wartung und Schulung des Personals. Hierbei stellt die Wartung des Systems einen großen Kostentreiber dar. Der Nutzen eines KI-basierten Spracherkennungssystems zeigt sich unter anderem in einer gesteigerten Effizienz durch Automatisierung der Dokumentation, strukturierter Informationsfluss, verbesserter Entscheidungsfindung durch schnelle Datenanalyse und erhöhter Behandlungsqualität durch Fehlerreduktion. Eine Quantifizierung der Nutzenfaktoren wiederum ist zu diesem Zeitpunkt schwierig zu realisieren. Hierzu bedarf es klarer praktischer Tests im Feld.

Die Ergebnisse der durchgeführten Workshops und Expertendiskussionen zeigen, dass insbesondere geeignete Anreizstrukturen entwickelt werden müssen, um die Anschaffung und Nutzung solcher Systeme zu gewährleisten. Mögliche Ansätze könnten hierbei empfohlene/verpflichtende Technologieeinsätze durch Fachgesellschaften, Krankenkassen oder auch Berufsgenossenschaften sein. Dabei stellt sich insbesondere die Frage, ob das System als Medizinprodukt oder lediglich als Informationsspeicher fungieren soll. Dies hat wiederum Auswirkungen auf die Implementierungsstrategie.

Meilensteine B4, M4: Projektverlängerung

Im Rahmen der Projektverlängerung wurden verschiedene Open-Source Sprachmodelle wie Llama 2 und Mistral-7B getestet. Ziel war es, eine kosteneffiziente und datenschutzfreundliche Alternative zu GPT-4 zu finden. Die Tests ergaben jedoch, dass die Leistung dieser Modelle nicht an die von GPT-4 heranreichte. Die Erkennungsgenauigkeit und die Fähigkeit zur Kontextualisierung waren bei den Open-Source-Modellen deutlich geringer, was zu weniger zuverlässigen Ergebnissen führte. Trotzdem gibt es auf dem Markt mittlerweile vielversprechende Entwicklungen. Neue Modelle wie Llama 3 und Mixtral8x7B zeigen erhebliches Potenzial und könnten in zukünftigen Projekten als Alternativen in Betracht gezogen werden.

Ein bedeutender Fortschritt wurde mit der Entwicklung einer Live-Anwendung erzielt, die die Spracherkennung und das Sprachmodell in Echtzeit miteinander verknüpft. Diese Anwendung ermöglicht es, Informationen unmittelbar auf dem Schockraum-Screen anzuzeigen. Durch die Echtzeitverarbeitung können wichtige Behandlungsinformationen sofort visualisiert werden, Die Live-Anwendung verwendet den Azure Speech-Recognition Service zur Spracherkennung, während GPT-4 die Sprachverarbeitung übernimmt.

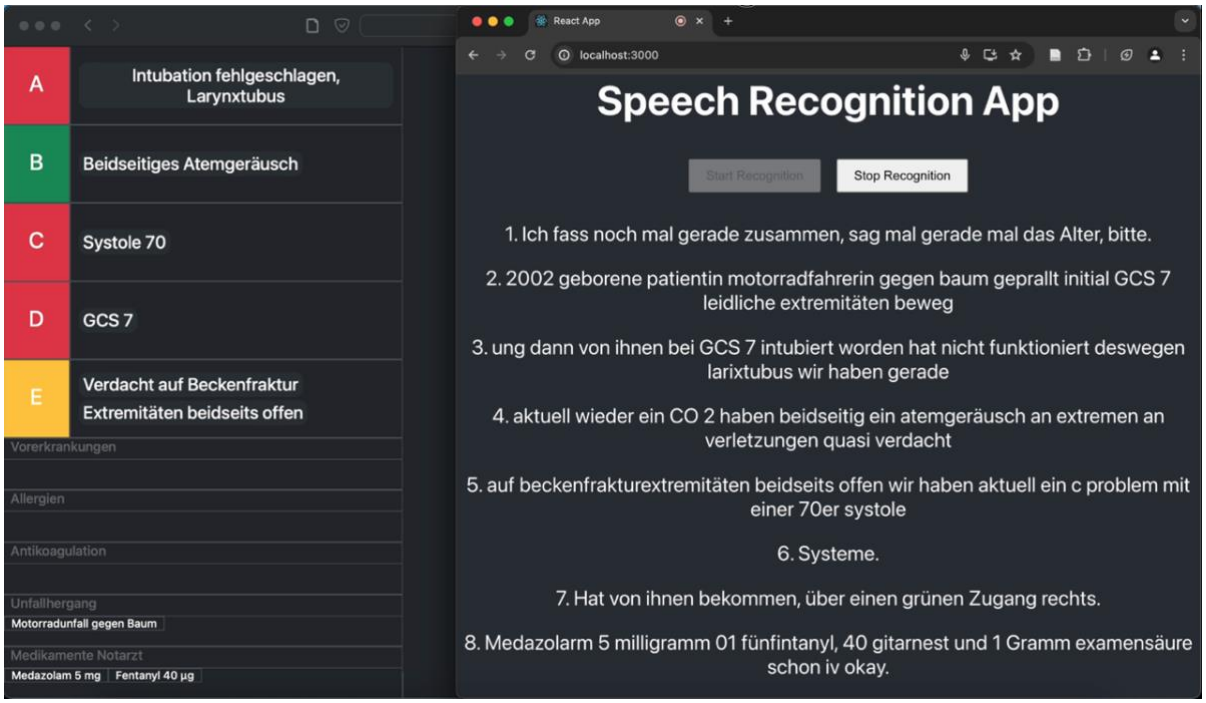


Abbildung 5: Prototyp des TraumaAgenten für die Schockraumversorgung mit laufender Echtzeitspracherkennung

7 Gender Mainstreaming Aspekte

Zur Vermeidung von potenziellen geschlechterspezifischen Verzerrungen des entwickelten KI-Spracherkennungssystems, wurden im Rahmen des Projektes Aspekte des Gender Mainstreamings berücksichtigt. Allen voran soll insbesondere die Spracherkennung für alle Stimmen gleich gut funktionieren. Um Geschlechterunterschiede besser abbilden zu können und die Entwicklung eines universellen Spracherkennungssystems sicherzustellen, wurde ein möglichst heterogener Datenpool als Grundlage für die Prototypenentwicklung genutzt.

Darüber hinaus wurde sowohl für die Schockraumstudie, kognitionspsychologischen Experimente als auch für die Zusammensetzung der Experten- und Expertinnen-Interviews stets auf Geschlechterparität geachtet, sodass Daten von verschiedenen Geschlechtern erhoben worden sind. Diese Erkenntnisse wurden gleichermaßen für den Ausbau und die Weiterentwicklung des KI-Systems berücksichtigt.

Insgesamt ist festzuhalten, dass für die Entwicklung eines in der Praxis anwendbaren und integrierbaren Systems ein möglichst breiter Datenpool unabhängiger spezifischer Endnutzergruppen unerlässlich ist, sodass wir in den gesamten Probandenstudien, Testungen und Weiterentwicklungen Probanden von verschiedenen Geschlechtern, Abteilungen und Qualifikationen berücksichtigt haben.

8 Diskussion der Ergebnisse, Gesamtbeurteilung

Die Behandlung von schwerstverletzten Patienten und Patientinnen im Schockraum ist ein medizinisch und kommunikativ komplexer und prognostisch für den Patienten und die Patientinnen höchst relevanter Abschnitt im Patienten- und Patientinnen Journey nach einer schweren Verletzung.

Gegenstand des Projektes TraumAInterfaces ist die Nutzung der während des Behandlungsprozesses gesprochenen Sprache:

- als Informationsmedium hinsichtlich des medizinischen Zustandes der behandelten Patienten und Patientinnen des medizinischen Behandlungsablaufes im Schockraum
- als Abbildung des Entscheidungsfindungsprozesses der ärztlichen Protagonisten und Protagonistinnen und insbesondere der Traumaleader und Traumaleaderinnen.
- als Informations- und Dokumentationsuntertsützungssystem im Schockraum

Zielsetzung des Projektes TraumAInterfaces war es, diese bisher ungenutzte bzw. nicht nutzbaren Ressourcen für die Entscheidungsfindung, die Prozesskontrolle und die Dokumentation des Behandlungsprozesses zugänglich zu machen.

Um dieses Ziel zu erreichen war es erforderlich die gesprochene Sprache in ein nachhaltig verfügbares Format zu übertragen. Dies geschah mittels eines Spracherkennungssystems mit automatisierter Transkription. Aufgrund der umfangrechen ethischen Regulatorik wurden diese Entwicklungsschritte im einem Fullscale Schockraumsimulator durchgeführt. Die zunächst bestehenden technischen Probleme konnten gelöst werden, so dass die in der gesprochenen Sprache transportierten Informationen als Echtzeittranskript vorliegen. Damit wurde die technisch notwendige Voraussetzung geschaffen, um KI-basierte Decision Support/Awareness Guidance Funktionalitäten zu entwickeln.

Die Veröffentlichung des ersten LLM (Chat GPT 3) während der Projektlaufzeit stellte ein für das Gesamtprojekt entscheidendes Ereignis da. Die ursprünglich geplante Generierung von Trainingsdaten in Fullscale Trainingsscenarien konnte aufgrund der Covid 19 Pandemie nicht wie geplant umgesetzt werden. Diese Lücke konnte durch das LLM geschlossen werden. Die schnelle technische Integration des LLM in das laufende Projekt war der entscheidende Schritt hinsichtlich der erfolgreichen Fortführung des Gesamtprojektes, weil ex post festgestellt werden muss, dass ein Decision Support/Awareness Guidance System, das auf der Grundlage eines kontextspezifischen Sprachmodels wahrscheinlich nur einen Bruchteil der Leistungsfähigkeit erreicht hätte.

Die grundsätzlich vorhandene Fähigkeit von LLM zur Autokontextualisierung war auch bei der Erprobung des Schockraummonitors (Abb. 2, 5) zu beobachten. Neben grundlegenden

dokumentarischen Funktionalitäten (Abb. 5) konnte durch das Projekt TraumAInterafces eine belastbare technische Grundlage für KI basierte Interventionen in medizinische Prozesse mit großer inhaltlicher und zeitlicher Dynamik vorgelegt werden. Neben dem beschriebenen Use Cases der Schockraumversorgung schwerstverletzter Patienten und Patientinnen sind sämtliche präklinische Prozesse, vergleichbare Schockraumsituationen in konservativen Fächern (Innere Medizin), Versorgungsprozesse in der militärischen Versorgungskette, aber auch vergleichbar dynamische Situationen im Operationssaal, denkbare Einsatzgebiete der hier entwickelten Integration von fallbezogener Kommunikation und LLM basierter Kontextualisierung unter Echtzeitbedingungen.

Ausgehend von der Annahme, dass KI technische Methoden bereitstellt, die einzelnen Aspekte der menschlichen Intelligenz unterstützen, ergänzen oder augmentieren, wurden im Rahmen des TraumAInterfaces Projektes Umfragen und Prozessbeobachtungen durchgeführt, um die subjektiven und objektiven Unterstützungsbedarfe über das gesamte Spektrum kognitiver Funktionen (von der Erinnerung an einzelne Fakten, bis hin zu Unterstützung bei komplexen Entscheidungen) zu identifizieren. Hierbei zeigt sich, dass in sämtlichen abgefragten Items (insgesamt 11) von den Befragten ein Unterstützungsbedarf besteht. In gleicher Weise zeigten die Prozessbeobachtungen, das ebenso externe Störfaktoren und Informationsverluste im Verlauf einer Schockraumbehandlung wirksam werden. Gleichzeitig zeigt sich in den begleitend durchgeführten Experteninterviews, dass die Akzeptanz von KI im Kontext von Unterscheidungsunterstützung sehr groß ist, so dass von Seiten der potenziellen Anwender keine fundamentalen Widerstände zu erwarten sind.

In der Zusammenschau hat das Projekt TraumAInterfaces für das anspruchsvolle Themenfeld der KI basierten Decision Support/Awareness Guidance in medizinisch komplexen und zeitlich dynamischen Szenarien technisch relevante Erkenntnisse erbracht, die für die zukünftige Gestaltung von Forschungs- und Entwicklungsprojekten einen klaren Rahmen vorgeben und im Sinne von Best Practice Regeln auch auf andere Projekte im Themenfeld KI basierte Decision Support/Awareness Guidance übertragen werden können.

Die Performanz von DS/AG Systemen wird davon abhängen, dass für die individuellen Unterstützungsbedarfe der einzelnen Anwender und Anwederinnen eine adäquate Modalität bereitgestellt wird. Dies bedeutet, dass es für einen Prozessschritt unterschiedliche/mehrere Unterstützungstools geben kann. Wenn der Endanwender und die Endanwenderin und mit dem jeweiligen Tool gut trainiert ist (Fullscale Simulationen), wird sich eine Verbesserung der Benchmark Parameter messen lassen. Eine One-Size-Fits-All Lösung lässt sich für dieses Anwendungsfeld von KI im klinischen Kontext nicht entwickeln.

Das zentrale Bindeglied zwischen dem medizinischen Prozess, dem Unterstützungsbedarf des einzelnen Anwenders, dem LLM und der medizinischen Evidenz ist der Prompt mit dem

das LLM angesteuert wird. Promptmanagement ist mehr als einer Frage an ein LLM zu stellen. Deshalb werden zukünftige Entwicklungsvorhaben, unabhängig vom jeweiligen medizinischen Kontext, vor allem die Identifikation optimaler Prompts zum Ziel haben müssen. Hierbei müssen sowohl theoretische Ansätze, die z.B. bestimmte kognitive Leistungen mit bestimmen semantischen Promptkonstruktionen in Verbindung bringen, aber auch experimentelle Ansätze, die Auswirkungen bestimmter Prompts auf das klinische Verhalten der einzelnen Ärzte und Ärztinnen, aber auch die Auswirkungen auf Prozessparameter untersuchen – hierzu erscheinen Fullscale Simulationsszenarien am besten geeignet.

9 Verbreitung und Öffentlichkeitsarbeit der Projektergebnisse

Im Rahmen des Projektes wurde der Projektfortschritt und die Ergebnisse des Projekts in den jeweiligen Gemeinschaften über die Projektwebseite, Konferenzbeiträge und Fachartikel in den beteiligten Disziplinen (u.a. Unfallchirurgie und Traumaforschung, Informatik und KI, Kognitionspsychologie, Technologie- und Innovationsmanagement) breit zugänglich gemacht werden. Das Teilen der Projektergebnisse ermöglicht es Wissen zu verbreiten und Innovation zu fördern. Gleichzeitig regt es die Zusammenarbeit zwischen Forschenden und Fachleuten an.

Die nachfolgende Tabelle fasst die veröffentlichten Publikationen zusammen. Weitere Publikationen befinden sich noch in der Entwicklungsphase:

Tabelle 5: Übersicht der veröffentlichten Publikationen

Publikation	Beschreibung
Handelsblatt	Wittenbrink, J. (2024). KI-Hilfen für die Notaufnahme. Kommen Patienten in die Notaufnahme, kann jede Sekunde zählen – vor allem für unerfahrene Ärzte bedeutet das großen Stress. KI-Lösungen können Orientierung bieten. Online verfügbar unter: https://www.handelsblatt.com/inside/digital-health/akutversorgung-ki-hilfen-fuer-die-notaufnahme/100056933.html. (Interview mit medizinischen und technischen Experten und Expertinnen aus dem Projekt)
Sonderheft Die Unfallchirurgie	Tjardes T., Meyer L.M., Lotz A., Defosse J., Hensen S., Salge T., Klasen M., Stead S., & Walossek N. (2023). Anwendung von Systemen der künstlichen Intelligenz im Schockraum. Die Unfallchirurgie, 126(7), 552-558.
Current Opinion in Critical Care	Gauss, T., Perkins, Z., & Tjardes, T. (2023). Current knowledge and availability of machine learning across the spectrum of trauma science. Current Opinion in Critical Care, 29(6), 713-721.
Technological Forecasting and Social Change	Meyer, L. M., Stead, S., Salge, T. O., & Antons, D. (2024). Artificial intelligence in acute care: A systematic review, conceptual synthesis, and research agenda. Technological Forecasting and Social Change, 206, 123568.
Whitepaper	Meyer, L. M., Giesselbach, S., Antweiler, D., Defosse, J., Hensen, S., Iser, H., Salge, T.O., Stead, S., Tjardes, T., & Waloßek, N. (2024). Wie Agenten und Foundation-Modelle bei der Versorgung Schwerverletzter helfen.

Um ein systematisches Voranschreiten das Projekt zu gewährleisten, Ergebnisse und Entwicklungen innerhalb des Konsortiums zu kommunizieren und sich gegenseitig zu unterstützen fanden regelmäßige (3-Wochen Turnus) interne Abstimmungstreffen im Projektkonsortium statt. Daneben wurde in regelmäßigen Abständen übergeordneten Projekttreffen online sowie vor Ort durchgeführt. Unter anderem wurden diese Termine dafür genutzt, Ergebnisse in größerer Runde vorzustellen und zu diskutieren oder auch Workshops für relevante Arbeitspakete durchzuführen. Im Rahmen dieser Workshops konnten so neue Projektinhalte entwickelt und weitere Perspektiven zur Systemverbesserung gewonnen werden. Des Weiteren wurden Besuche von Konferenzen und Vernetzungsveranstaltungen genutzt, um das Projekt zu präsentieren und in einen fachlichen Austausch mit den Teilnehmern und Teilnehmerinnen zu gehen. Auf übergeordneter Ebene wurden regelmäßig Neuigkeiten zum aktuellen Projektstand auf der Healthcare Innovation Lab Seite des Tim Instituts geteilt und relevante Entwicklungen und Meilensteine über die LinkedIn Seite des TIM Instituts kommuniziert.

10 Verwertung der Projektergebnisse (Nachhaltigkeit/Transferpotenzial)

Im nachfolgenden Kapitel werden die Verwertung der Projektergebnisse näher spezifiziert und diskutiert. Zunächst werden die Erfahrungen und Ergebnisse, die sich für einen möglichen Praxis- und Gesetzgebungstransfer eignen diskutiert. Daran anschließend erörtern wir die Ergebnisse im Kontext der allgemeinen Gesundheitsversorgung und erläutern den Fortbestand bzw. die Weiterentwicklung der etablierten Strukturen. Zum Ende des Kapitels werden weitere Implikationen und mögliche zukünftige Projektausrichtungen diskutiert.

10.1 Erfahrungen und Ergebnistransfer

Das TraumAInterfaces Projekt fungiert als ein Best Practice um systematisch KI-Systeme auch in dynamischen Gesundheitsprozessen wie der Traumaversorgung zu entwickeln. Hierbei ist insbesondere der genutzte reversed-engineering Ansatz hervorzuheben. Mit diesem Ansatz war es möglich, in einem frühen Stadium das initiale System zu testen, neure relevante Daten aus dem Feld zu erheben und in die Entwicklung miteinfließen zu lassen. Im Fokus dabei stand die Entwicklung multidimensionales KI-Systems, welches gleichermaßen technische, medizinische und psychologische Aspekte gleichwertig berücksichtigt. Die wesentlichen Erkenntnisse für die Entwicklung und Anpassungen des Systems ergaben sich dabei an den Schnittstellen zwischen diesen Domänen. Die Projektergebnisse zeigen zudem, dass insbesondere LLMs große Potenziale bieten, den Behandlungsprozess im Schockraum nachhaltig zu verbessern und medizinisches Personal zu entlasten. Diese Erkenntnisse und Überlegungen sollten in zukünftigen Projekten mitberücksichtigt und weiterentwickelt werden.

Eine weitere wichtige Erkenntnis, die aus dem Projekt gezogen werden kann, ist die Relevanz des KI-Wissens bzw. der AI-Literacy. Auch wenn in den durchgeführten Experten und Expertinnen Interviews durchgehend eine große Akzeptanz und ein positives Interesse der potenziellen Nutzer und Nutzerinnen an KI-Anwendungen im Bereich Decision Support/Awareness Guidance zu beobachten war, darf nicht vergessen werden, dass bei vielen nicht direkt involvierten Personen und Gruppen u.U. erhebliche Vorbehalte und Ängste bestehen. Gegenwärtig stellt die Einbeziehung von Mitarbeitenden und deren Einwilligung in die Nutzung des Spracherkennungssystems die größte verbleibende Herausforderung dar. Die erste Hürde liegt hierbei in der sehr großen und unscharf begrenzten Personengruppe derjenigen, die potenziell im Schockraum zum Einsatz kommen können; in großen Krankenhäusern ist dies u.U. eine Gruppe von mehreren hundert Personen, die darüber hinaus eine starke Fluktuation aufweisen kann. Die zweite Hürde liegt in externem Personal,

das gar nicht zur Krankenhausbelegschaft gehört, aber trotzdem im Schockraum tätig ist; als Beispiel seien hier Rettungssanitätern und -sanitäterinnen und Notärzte und Notärztinnen genannt. Alle diese Personengruppen vorab über den Einsatz des Systems zu informieren, ihre Einwilligung einzuholen und deren Vorliegen vor Ort zu überprüfen, erscheint nicht realistisch.

Diese Thematik kann nur durch Wissen (AI Literacy) über KI im Gesundheitswesen, und Schaffen von Vertrauen in die Technologie seitens der betroffenen Personengruppen adressiert werden. Das übergeordnete Ziel muss sein KI-Anwendungen als gleichberechtigten Teil der medizinischen Diagnostik, Entscheidungsfindung und Therapie zu etablieren. Hierzu sollten von den institutionellen Protagonisten, namentlich den medizinischen Fachgesellschaften, KI-Strategien eingefordert werden, die hinsichtlich der Verbindlichkeit und dem akademischen Anspruch, aber auch hinsichtlich der Strategien zur Schaffung einer AI Literacy ein hohes Maß an Verbindlichkeit aufweisen. Idealerweise werden die fachspezifischen Konzepte dann durch konkrete Aus- und Weiterbildungsmaßnahmen durch fachlich ausgewiesene Institutionen wie z.B. der Fraunhofer Gesellschaft umgesetzt. Hierbei ist darauf zu achten, dass derartige Maßnahmen nicht nur auf dem Level von Unikliniken stattfinden, sondern das gesamte Spektrum der Versorgungsstufen abdecken.

10.2 Erkenntnisse für die Gesundheitsversorgung/Pflege

Das Projekt TraumAInterfaces hat einen Rahmen geschaffen, der es ermöglicht mit überschaubarem weiterem Aufwand ein markfähiges System eines KI-basierten Unterstützungssystems für Schockraumprozesse zu entwickeln. Eine Ausentwicklung des Projektes eröffnet die Implementierung von Funktionalitäten, die sowohl die medizinische Qualität (DS/AG), die Materialwirtschaft (z.B. Transfusionsmanagement) als auch dokumentarischen Aufgaben unterstützen und/oder übernehmen können und somit zur Unterstützung und Entlastung sämtlicher Berufsgruppen beitragen können.

10.3 Nutzung der aufgebauten Projektstrukturen

Die Zusammenarbeit zwischen den Konsortialpartnern hat sich auch über das Projekt hinaus inhaltlich verstetigt und mündet in der Planung und Initiierung weiterer gemeinsamer Projekte, die sich auf die Anwendung der neuesten Versionen von Chat GPT, insbesondere Chat GPT 4.0, konzentrieren. Diese Ansätze befinden sich derzeit noch in einer grundlagenorientierten Testphase und werden sukzessive weiterausgebaut. Zudem sind bereits weitere Projektanträge in Planung, die die bisherigen Erkenntnisse und etablierten Strukturen als Grundlage nutzen und das System in anderen Anwendungsfällen zu applizieren und weiter

auszubauen. Darüber hinaus wurden erste vielversprechende Austausche zwischen dem Konsortium und Microsoft Healthcare, Deutschland initiiert, um über Möglichkeiten und Strukturen einer cloudbasierten Weiterentwicklung des KI-Systems zu diskutieren. Der Austausch bietet eine vielversprechende Perspektive bestehenden Strukturen weiter nutzbar zu machen und zu skalieren.

10.4. Weitere Implikationen der Ergebnisse

Die generierten Projektergebnisse liefern zudem weitere Implikationen sowohl für die Schockraumversorgung als auch für übergeordnete (zukünftige) Fragestellungen:

Der im Rahmen des Projektes TraumAInterfaces entwickelte Showcase hat aufgezeigt, dass sich durch die Verbindung von Spracherkennung und LLM ein umfassendes Unterstützungssystem mit verschiedenen Funktionalitäten für Schockraumprozesse aufbauen lässt. Um das Potenzial von KI-basierten Systemen in der Schockraumversorgung vollumfassend entfalten zu können sollte in zukünftigen Forschungsprojekten zusätzliche KI-Systeme für die Schockraumversorgung entwickelt und deren Anwendung ausgebaut werden. Basierend auf den Projektergebnissen ergeben sich folgende weitere Anwendungen:

- Prozessdokumentation (sowohl ärztliche Dokumentation als auch Registerdokumentation)
- Anwenderunterstützung durch Checklistenfunktionen
- Kontextualisierung der konkreten Behandlungssituation durch Abgleich und Einspielen von Leitlinien an die Entscheidungsträger
- Hinweise auf potenzielle Fehler in der Wahrnehmung und Bewertung von Informationen (Awarenessguidance)
- Materialwirtschaft (z.B. verwendete Blutprodukte und Medikamente)
- Erstellen eines Debriefing Protokolls mit Hinweis auf kritische Situationen

Aus den generierten Erkenntnissen ergeben sich ebenso zukünftige Fragestellungen hinsichtlich eines personalisierten Prompmanagement und spezifische KI-Persönlichkeitsmerkmale. Im Rahmen des Projektes TraumAInterfaces konnte gezeigt werden, dass KI-Tools im Bereich DS/AG auf die Bedürfnisse der Nutzer und Nutzerinnen zugeschnitten sein müssen. Daraus ergibt sich grundlagenwissenschaftlicher Forschungsbedarf, der unterschiedliche KI-Tools bestimmten kognitiven Unterstützungsbedarfen zuordnet. Inwieweit dies tatsächlich im Sinne eines “diagnostischen Prozesses” möglich ist, oder ob im Rahmen der Ausbildung jeder potenzielle Schockraumprotagonist im Rahmen von Simulationsszenarien das individuell am besten

geeignete Unterstützungstool herausfinden muss, sollte ebenso Gegenstand weiterer Forschungsprojekte sein.