

Kurzbericht

zum Verbundvorhaben

„Führungsausbildung für Einsatzkräfte mittels intelligenter virtueller Realitäten (feir)“



Teilvorhaben:

„Automatisierte Analyse und Bewertung von komplexen Handlungsabläufen in virtuellen Umgebungen“

Förderkennzeichen: 13N16413



FZI Forschungszentrum Informatik

Haid-und-Neu Str. 10-14

76131 Karlsruhe

Autoren: Dennis Birkenmaier, Lara Schweickart, Christina Erler

Telefon: +49 721 9654-762, +49 721 9654-760, +49 721 9654-195

E-Mail: birkenmaier@fzi.de, schweickart@fzi.de, erler@fzi.de

1. Ausgangssituation und Zielsetzung des Vorhabens

Ziel des Vorhabens FEIR war die Konzeption, Entwicklung und Evaluation eines VR-basierten Ausbildungssystems zur realitätsnahen Schulung von Einsatz- und Führungskräften im Bevölkerungsschutz. Ausgangspunkt des Projekts war der bestehende wissenschaftlich-technische Stand in den Bereichen simulationsbasierter Ausbildung, Virtual Reality sowie automatisierter Leistungsbewertung. Insbesondere sollte der bisher eingeschränkten Möglichkeit zur objektiven Bewertung von Verhalten, Entscheidungsfindung und Stressreaktionen in Ausbildungsszenarien begegnet werden.

Das Vorhaben zielte darauf ab, einsatztaktisch relevante Szenarien automatisiert zu generieren, die Durchführung in einer immersiven VR-Umgebung zu ermöglichen und das Verhalten der Nutzenden anhand definierter Metriken automatisiert auszuwerten. Ergänzend sollte eine objektivierbare Erfassung von Stressreaktionen integriert werden, um die Leistungsfähigkeit unter Belastung bewerten zu können.

2. Ablauf des Vorhabens

Zu Beginn des Projekts erfolgte eine strukturierte Bedarfsermittlung. Hierzu wurden das notwendige Domänenwissen sowie der Ist-Zustand bestehender Ausbildungsformate analysiert. Auf Basis von Anwender*inneninterviews und Befragungen wurden die funktionalen und fachlichen Anforderungen an das VR-Ausbildungssystem erhoben und in einem Anforderungskatalog dokumentiert.

Darauf aufbauend wurde ein einsatztaktisches Gesamtkonzept entwickelt und eine modulare Systemarchitektur definiert. Diese umfasst ein Backend zur Szenariengenerierung und zum Datenaustausch, ein Frontend zur Bedienung und Darstellung von Trainingsfortschritten, eine VR-Anwendung zur Durchführung der Trainingsszenarien sowie ein Auswertungsmodul zur automatisierten Leistungsbewertung. Die Kommunikation zwischen den Komponenten erfolgt über standardisierte REST-Schnittstellen unter Verwendung strukturierter Datenformate.

Ein wesentlicher Bestandteil des Vorhabens war die Entwicklung eines Konzepts zur automatisierten Szenariengenerierung. Die Übungsszenarien werden schrittweise aufgebaut, wobei Entscheidungen getroffen werden, die den Schwierigkeitsgrad beeinflussen. Für jede Szenariovariante wird ein Schwierigkeitswert ermittelt, sodass Szenarien gezielt entsprechend definierter Anforderungen erzeugt werden können.

Parallel wurde ein umfassendes Auswertungs- und Bewertungskonzept erarbeitet. Hierzu wurden relevante Metriken, unter anderem zu Zeitbedarf, Laufweg, Erfassung von Kontrollmarkierungen, Aufmerksamkeit, Sicherheit, Erinnerungsleistung und Unsicherheit, definiert und mit konkreten Messparametern hinterlegt. Ergänzend wurde eine automatische Handlungserkennung implementiert, die das Verhalten der Nutzenden fachlich einordnet und bewertet.

Zur erweiterten Analyse des Nutzendenverhaltens wurde zusätzlich eine Stressmessung integriert. Hierzu kamen externe EKG-Sensoren zum Einsatz, deren Daten über Parameter der Herzratenvariabilität ausgewertet wurden. Da diese Parameter stark personenabhängig sind, wurde ergänzend ein KI-basiertes Verfahren zur Stressbewertung entwickelt und validiert.

Im weiteren Projektverlauf wurde das Auswertungsmodul in das Gesamtsystem integriert und die Kommunikation zwischen den Komponenten umfassend getestet. Begleitend wurden Workshops, Nutzendenevaluationen und Praxistests mit Einsatzkräften durchgeführt.

3. Wesentliche Ergebnisse und Zusammenarbeit

Im Ergebnis wurde ein funktionsfähiges, modular aufgebautes VR-Ausbildungssystem entwickelt, das die automatisierte Generierung realitätsnaher Einsatzszenarien sowie eine objektive, automatisierte Auswertung des Nutzendenverhaltens ermöglicht. Die definierten Metriken und die

implementierte Handlungserkennung erlauben eine differenzierte Bewertung von Effizienz, Handlungssicherheit und taktischem Vorgehen. Die ergänzende Stressanalyse liefert zusätzliche Informationen zur Leistungsfähigkeit unter Belastung.

Die durchgeführten Evaluationen bestätigten die fachliche Korrektheit sowie den didaktischen Mehrwert des Systems. Die Rückmeldungen der Anwender*innen waren überwiegend positiv und lieferten wertvolle Hinweise für mögliche Weiterentwicklungen.

Das Vorhaben wurde in enger Zusammenarbeit zwischen den beteiligten Forschungseinrichtungen und Praxispartnern durchgeführt, insbesondere zwischen dem Fraunhofer IAO, dem FZI sowie weiteren Projektpartnern. Die interdisziplinäre Zusammenarbeit trug maßgeblich zur erfolgreichen Umsetzung der Projektziele bei.

Abschlussbericht

zum Verbundvorhaben

„Führungsausbildung für Einsatzkräfte mittels intelligenter virtueller Realitäten (feir)“



Teilvorhaben:

„Automatisierte Analyse und Bewertung von komplexen Handlungsabläufen in virtuellen Umgebungen“

Förderkennzeichen: 13N16413



FZI Forschungszentrum Informatik

Haid-und-Neu Str. 10-14

76131 Karlsruhe

Autoren: Dennis Birkenmaier, Lara Schweickart, Christina Erler

Telefon: +49 721 9654-762, +49 721 9654-760, +49 721 9654-195

E-Mail: birkenmaier@fzi.de, schweickart@fzi.de, erler@fzi.de

1 Eingehende Ergebnisdarstellung

AP1: Bedarfsermittlung

TAP 1.1: Domänenwissen & Ist-Zustand bei Führungsausbildung

Das FZI beteiligte sich maßgeblich an der Erarbeitung des projektrelevanten Domänenwissens und brachte seine technische sowie methodische Expertise bei der strukturierten Erstellung der Wissenssammlung ein. Dabei wurden einsatztaktische Grundlagen, typische Handlungsabläufe, Entscheidungslogiken sowie relevante Bewertungskriterien systematisch erfasst, analysiert und in eine formal nutzbare Struktur überführt. Ziel war es, implizites Erfahrungswissen aus der Praxis in explizite, technisch verwertbare Modelle zu überführen, sodass dieses sowohl für die Szenariogenerierung als auch für die automatisierte Bewertung nutzbar gemacht werden kann.

Ein besonderer Fokus lag auf der Übersetzung fachlicher Anforderungen in technische Parameter und Entscheidungsregeln. Hierbei unterstützte das FZI insbesondere bei der Identifikation messbarer Indikatoren für fachlich korrektes oder verbesserungswürdiges Verhalten sowie bei der Strukturierung von Einflussfaktoren auf Schwierigkeitsgrad und Performanz.

Darüber hinaus unterstützte das FZI federführend bei der Konzeption und Ausarbeitung der Fragebögen zur Befragung der Nutzenden. Ziel war es, die relevanten Aspekte für Szenariogenerierung, didaktische Ausgestaltung und Bewertungslogik systematisch zu erfassen. Die Fragebögen wurden so gestaltet, dass sowohl qualitative Einschätzungen (z. B. wahrgenommene Realitätsnähe, Verständlichkeit der Auswertung, subjektive Belastung) als auch quantitative Bewertungen erhoben werden können.

Dabei wurde insbesondere darauf geachtet, dass die abgefragten Inhalte direkt auf die Weiterentwicklung der Szenarienlogik und der Bewertungsmetriken einzahlen. Dies umfasst beispielsweise Einschätzungen zur Angemessenheit der Schwierigkeit, zur Nachvollziehbarkeit der Bewertung, zur Relevanz einzelner Metriken sowie zum didaktischen Mehrwert der Rückmeldungen.

Durch diese strukturierte Einbindung des Domänenwissens und der Nutzendenperspektive wurde eine fundierte Grundlage geschaffen, um das System sowohl fachlich valide als auch praxisnah weiterzuentwickeln.

TAP 1.2: Erhebung und Fixierung der Bedarfe an ein VR-Ausbildungssystem

Das FZI beteiligte sich aktiv an der Erstellung, Strukturierung und Dokumentation eines umfassenden Anforderungskataloges für das VR-Ausbildungssystem sowie für das übergeordnete Gesamtsystem, in das dieses integriert wurde. Ziel war es, sowohl technische als auch fachlich-didaktische Anforderungen systematisch zu erfassen und priorisiert zu dokumentieren. Dabei wurde besonderer Wert auf eine klare Trennung zwischen funktionalen Anforderungen (z. B. Szenariogenerierung, Bewertung, Visualisierung) und nicht-funktionalen Anforderungen (z. B. Skalierbarkeit, Latenz, Wartbarkeit, Erweiterbarkeit) gelegt. Zur Priorisierung der Anforderungen wurde die MoSCoW-Methode angewendet, um Transparenz hinsichtlich Umsetzungsrelevanz und Projektumfang zu schaffen.

MUST(Zwingend erforderlich)

- Das System muss modular aufgebaut sein (Backend, Frontend, VR-Anwendung, Auswertungsmodul).
- Das Backend muss zentrale Datenhaltung und -verwaltung übernehmen.
- Die VR-Anwendung muss echtzeitfähig auf einem VR-System lauffähig sein.
- Das Auswertungsmodul muss direkt an die VR-Anwendung angebunden sein (geringe Latenz).

- Die Kommunikation muss über eine REST-Schnittstelle mit JSON-Format erfolgen.
- Das System muss erweiterbar und skalierbar konzipiert sein.
- Jede Bewertungsmetrik muss durch einen eindeutig definierten Messparameter abgebildet sein.
- Szenarien müssen automatisiert generierbar sein.
- Der Schwierigkeitsgrad muss deterministisch berechnet werden.
- Innerhalb einer Schwierigkeitsstufe müssen Varianten erzeugbar sein.
- Die Bewertung muss grafisch dargestellt werden.
- Messparameter müssen in Metriken überführt und visualisiert werden.
- Die Ergebnisse müssen unmittelbar nach dem Szenario verfügbar sein.
- Das Gesamtsystem muss evaluiert werden (Workshops & Nutzendentests).
- Die fachliche Korrektheit der Bewertung muss überprüfbar sein.

SHOULD (sollte umgesetzt werden)

- Automatische Anpassung des Schwierigkeitsgrades basierend auf Leistungsprofil
- Replay-Funktion mit zeitlicher Aufschlüsselung von Schlüsselereignissen

COULD (optional, Mehrwert)

- Adaptive Szenarien in Echtzeit
- Integration zusätzlicher Sensorik (EDA, BVP)
- Benchmarking zwischen verschiedenen Einsatzkräften
- Kompetenzentwicklungs-Tracking über mehrere Trainings
- Cloudbasierte Auswertung
- Mobile Analyse-App

WON'T (im aktuellen Projektumfang nicht vorgesehen)

- Monolithische Systemarchitektur
- Lokale Speicherung der Bewertungsdaten im Auswertungsmodul
- Vollständige KI-basierte Bewertung ohne regelbasierte Metriken
- Echtzeit-Neutrainierung der KI-Modelle im Produktivsystem

Durch die strukturierte Priorisierung nach der MoSCoW-Methode konnte eine klare Entwicklungsstrategie definiert werden. Kritische Kernanforderungen wurden verbindlich umgesetzt, während weiterführende Funktionen als perspektivische Erweiterungen konzipiert wurden. Dies schuf Planungssicherheit, technische Klarheit und eine transparente Entscheidungsgrundlage für alle Projektbeteiligten.

AP2: Gesamtkonzept & Einsatztaktik

TAP 2.1: Einsatztaktisches Gesamtkonzept

Aus den Erkenntnissen von AP1 entstand die Systemarchitektur, welche in vier Komponenten gegliedert wurde. Bei der Konzeption wurde bewusst eine modulare Architektur gewählt, um eine klare Trennung der Verantwortlichkeiten (Separation of Concerns) zu gewährleisten. Ziel war es, einzelne Systembestandteile unabhängig voneinander entwickeln, testen und perspektivisch erweitern zu können. Alternativ wäre eine stärker integrierte oder monolithische Architektur möglich gewesen, bei der beispielsweise Auswertung und VR-Anwendung enger im Backend verankert worden wären. Dies hätte jedoch die Flexibilität reduziert und Anpassungen oder Erweiterungen einzelner Module erschwert. Die Entscheidung für die gewählte Architektur basiert daher insbesondere auf folgenden Argumenten:

- Klare Verantwortungsbereiche je Modul
- Geringe gegenseitige Abhängigkeiten
- Einfache Weiterentwicklung einzelner Komponenten

- Flexible Deployment-Möglichkeiten
- Skalierbarkeit und Austauschbarkeit einzelner Module.

Die Systemarchitektur gliedert sich in folgende vier Komponenten:

1. **Feir-Backend:**
Hier erfolgen alle Verarbeitungsschritte bezüglich der Szenariengenerierung sowie der Datenaustausch zwischen den restlichen Modulen. Das Backend fungiert als zentrale Kommunikations- und Verwaltungsinstanz und wird durch Vomatec betrieben. Es übernimmt die dauerhafte Speicherung und Verwaltung aller relevanten Trainingsdaten.
2. **Feir-Frontend:**
Diese Komponente ist für die Darstellung des Trainingsfortschrittes sowie für die Bedienung des VR-Ausbildungssystems außerhalb der VR-Anwendung verantwortlich. Das Frontend greift auf die im Backend gespeicherten Daten zu und stellt diese nutzerfreundlich dar.
3. **VR-Anwendung:**
Die VR-Anwendung bezeichnet das System zur Durchführung der Trainingsszenarien unter Einsatz eines VR-Headsets. Sie läuft direkt auf der VR-Brille bzw. dem angeschlossenen System und ist auf eine performante, latenzarme Ausführung optimiert. Aus diesem Grund wurde darauf geachtet, rechenintensive oder nicht zeitkritische Prozesse aus der VR-Anwendung auszulagern.
4. **Auswertungsmodul:**
Das Auswertungsmodul empfängt die während der Simulation erfassten Messparameter und berechnet daraus konkrete Bewertungen für die einzelnen Metriken.

Das Auswertungsmodul ist direkt an die VR-Anwendung angebunden und nicht als Teil des Backends umgesetzt. Diese Entscheidung wurde getroffen, um eine unmittelbare Verarbeitung der Trainingsdaten zu ermöglichen. Dadurch können relevante Kennzahlen ohne zusätzliche Umwege berechnet und bei Bedarf direkt im Anschluss an ein Szenario bereitgestellt werden.

Eine Integration in das Backend hätte zusätzliche Kommunikationsschritte und potenzielle Latenzen verursacht. Da die Auswertung eng mit den während der Simulation entstehenden Rohdaten verknüpft ist, erwies sich eine unmittelbare Anbindung als technisch effizienter und architektonisch klarer.

Im Rahmen der Systemkonzeption wurde ebenfalls berücksichtigt, auf welchen Systemen die einzelnen Komponenten ausgeführt werden sollen. Die VR-Anwendung läuft auf der VR-Brille bzw. dem zugehörigen Ausführungssystem und ist auf Echtzeitfähigkeit optimiert. Das Backend wird serverseitig betrieben, wodurch zentrale Datenhaltung, Wartung und Skalierung ermöglicht werden. Das Frontend kann auf Standard-Arbeitsplätzen genutzt werden und greift über definierte Schnittstellen auf das Backend zu.

Durch diese bewusste Verteilung der Komponenten wird sichergestellt, dass:

- rechenintensive Echtzeitprozesse lokal in der VR-Umgebung stattfinden,
- persistente Datenhaltung zentral erfolgt,
- Wartung und Weiterentwicklung einzelner Komponenten unabhängig voneinander möglich bleiben.

Die definierte Systemarchitektur ist in Abbildung 1 schematisch visualisiert.

Durch die direkte Weiterverarbeitung der Daten vor und nach dem Auswertungsmodul war eine zusätzliche Speicherung der verarbeiteten Werte seitens des FZI nicht notwendig. Ein initial angedachtes Datenbankkonzept als Teil des Auswertungsservice entfällt daher. Die Daten werden ausschließlich im von Vomatec betriebenen Backend gespeichert und verwaltet.

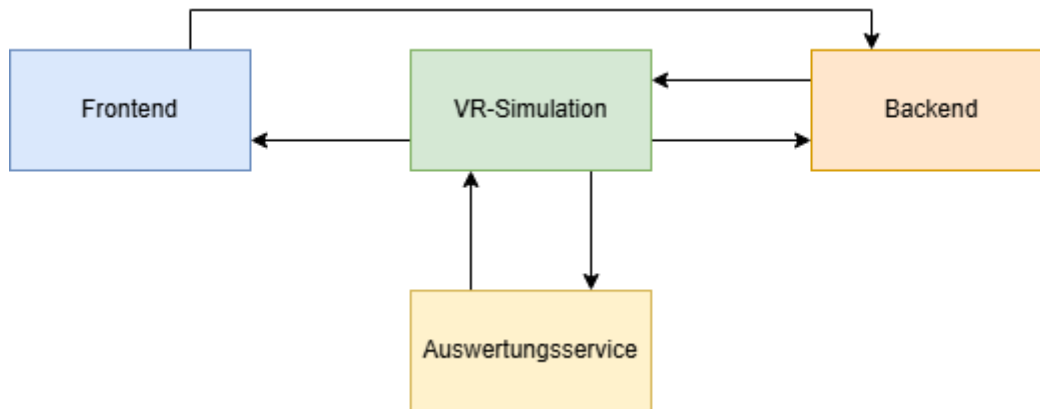


Abbildung 1: feir-Systemarchitektur

Parallel zur technischen Systemkonzeption wurde ein domänenspezifisches Kompetenzmodell für Feuerwehrführungskräfte erarbeitet.

Die Ermittlung der Kompetenzen erfolgte auf Basis von:

- Literaturrecherche zu Führung im Einsatz- und Katastrophenschutz,
- Analyse bestehender Ausbildungskonzepte,
- strukturierten Interviews mit erfahrenen Feuerwehrführungskräften,
- Workshops mit Praxisexpert*innen,
- iterativen Abstimmungen im Projektkonsortium.

Ziel war es, nicht nur beobachtbare Handlungen, sondern zugrunde liegende Fähigkeiten systematisch zu erfassen und operationalisierbar zu machen. Dabei wurde besonderer Wert darauf gelegt, dass jede Kompetenz durch messbare Indikatoren in der Simulation abbildbar ist.

Die identifizierten Kompetenzbereiche umfassen unter anderem:

1. Zeit- und Ressourceneffizienz
→ Operationalisiert durch Szenariozeit und Wegoptimierung.
2. Strukturierte Lageerkundung
→ Abgebildet über Kontrollmarkierungen, gesichtete Bereiche und getestete Zugänge.
3. Gefahrenerkennung und Aufmerksamkeit
→ Gemessen durch Reaktionszeit auf Gefahren und Markierungsgenauigkeit.
4. Sicherheitsbewusstsein
→ Bewertet über übertretene Gefahrenbereiche (Damage Events).
5. Informationsverarbeitung und Erinnerungsleistung
→ Operationalisiert über Nachbefragung (Endquiz) und Frontalansicht.
6. Entscheidungssicherheit unter Zeitdruck
→ Erfasst durch Bearbeitungszeit und Fehlerquote.
7. Stressregulation und Belastungssteuerung
→ Analysiert über Herzratenvariabilität und KI-basiertes Stress-Scoring.

Die in Abbildung 2 dargestellte Zuordnung von Kompetenzen zu konkreten Metriken verdeutlicht den Brückenschlag zwischen pädagogisch-didaktischem Anspruch und technischer Messbarkeit. Jede Kompetenz ist mindestens einer quantifizierbaren Metrik zugeordnet, wodurch eine objektivierte, nachvollziehbare und reproduzierbare Leistungsbewertung möglich wird.

Durch diese systematische Operationalisierung entsteht ein Bewertungsrahmen, der:

- fachlich fundiert,
- didaktisch begründet,
- technisch messbar,
- und automatisiert auswertbar

ist.

Somit bildet das Kompetenzmodell die inhaltliche Grundlage für die automatisierte

Handlungsbewertung und stellt sicher, dass das System nicht lediglich Verhaltensdaten sammelt, sondern einsatzrelevante Führungsqualitäten strukturiert bewertet.

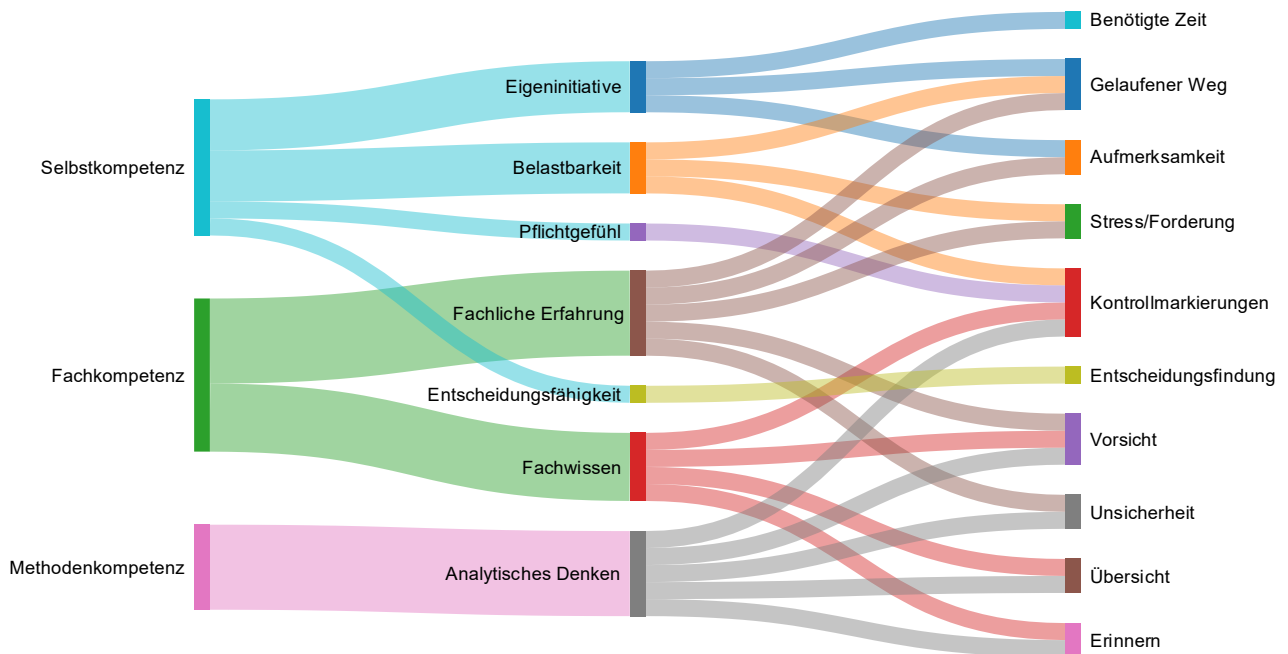


Abbildung 2: Kompetenzen und zugehörige Metriken

AP 3: Gesamtplattform & Kommunikation der Komponenten

TAP 3.1: Schnittstellendefinition

Das FZI beteiligte sich am Entwurf der Schnittstellen zwischen dem Gesamtsystem und des Auswertungssystems. Die Schnittstelle wurde in Form einer REST-Schnittstelle entworfen, bei der Daten in der JavaScript Object Notation (JSON) vorliegen. Das JSON-Format ermöglicht eine einfache Erweiterung der übertragenen Werte und den zugehörigen Wertebereichen, um auch nach Projektende eine einfache Weiterentwicklung zu gewährleisten. Eine einfache Weiterentwicklung sollte gewährleistet werden, um langfristige Stabilität, Anpassungsfähigkeit und Zukunftssicherheit sicherzustellen. Systeme, Prozesse oder Konzepte, die flexibel erweiterbar sind, können effizient an neue Anforderungen, technologische Veränderungen oder organisatorische Entwicklungen angepasst werden. Dadurch wird vermieden, dass bei jeder Änderung grundlegende Strukturen neu geschaffen werden müssen.

Der Mehrwert einer solchen Weiterentwicklung liegt insbesondere in der Nachhaltigkeit und Wirtschaftlichkeit. Bestehende Lösungen können schrittweise optimiert werden, wodurch Ressourcen geschont und Kosten reduziert werden. Gleichzeitig bleibt die Handhabung übersichtlich, da auf bekannten Strukturen aufgebaut wird. Dies erleichtert sowohl die Implementierung neuer Funktionen als auch die Einarbeitung beteiligter Personen.

Darüber hinaus fördert eine einfache Weiterentwicklung Innovationsfähigkeit. Neue Ideen oder Verbesserungen können schneller integriert und getestet werden, ohne umfangreiche Umstellungen vornehmen zu müssen. Insgesamt trägt dies zu einer höheren Effizienz, geringeren Risiken und einer besseren langfristigen Planbarkeit bei.

Die Ein- und Ausgabewerte des Auswertungsmoduls sind jeweils in Tabelle 1 und Tabelle 2 beschrieben.

Tabelle 1: Eingabewerte des Auswertungsmodul

Metrik	Identifizier	Beschreibung
Benötigte Gesamtzeit	Scenario_time	Benötigte Zeit zur Durchführung des Szenarios
Gelaufener Weg	User_path_optimal	Wegstrecke des algorithmisch bestimmten optimalen Pfades
	User_path_actual	Wegstrecke des vom Nutzenden gewählten Pfad in der VR-Anwendung
Erfassung Kontrollmarkierungen	Checkpoints_max	Maximale Anzahl der erfassbaren Kontrollmarkierungen im aktuellen Szenario
	Checkpoints_actual	Anzahl der tatsächlich erfassten Kontrollmarkierungen
Aufmerksamkeit	Danger_react_time	Zeit, in der eine Gefahr sich im Sichtkegel befindet, bevor sie als Gefahr markiert wird
	Danger_clicks	Anzahl der Versuche eine Gefahr zu markieren, obwohl sich keine Gefahr im Sichtkegel befindet
Vorsicht/ Sicherheit/ Übertretene Gefahren	Damage_events	Anzahl der übertretenen Gefahrenbereiche
Erinnern	Endquiz_max	Anzahl der maximalen korrekten Einträge in der Befragung nach dem Szenario
	Endquiz_actual	Anzahl der erreichten korrekten Einträge in der Befragung nach dem Szenario
Unsicherheit	Endquiz_time	Benötigte Zeit zum Ausfüllen der Befragung nach dem Szenario
Frontalansicht	Frontalview_max	Anzahl der maximalen korrekten Einträge in der Befragung nach der Frontalansicht
	Frontalview_actual	Anzahl der erreichten korrekten Einträge in der Befragung nach der Frontalansicht

Tabelle 2: Ausgabewerte des Auswertungsmoduls

Metrik	Identifizier
Benötigte Gesamtzeit	Time
Gelaufener Weg	User_path
Erfassung Kontrollmarkierungen	Checkpoints
Aufmerksamkeit	Attention
Vorsicht/Sicherheit	Safety
Erinnern	Memory

Darüber hinaus diente die Mock-Schnittstelle der Unterstützung projektbeteiligter Partner, deren Komponenten von der Anbindung an das Auswertungsmodul abhängig waren. Durch die frühzeitige Bereitstellung einer funktionsfähigen, wenn auch vereinfachten Schnittstelle konnten diese Partner ihre eigenen Entwicklungen parallel vorantreiben, testen und anpassen. Dies reduzierte Integrationsrisiken und ermöglichte eine iterative, abgestimmte Weiterentwicklung der Systemarchitektur.

Nach erfolgreicher Validierung der Kommunikationsstruktur wurde die Mock-Implementierung schrittweise durch die produktive Bewertungslogik ersetzt. Dadurch konnte eine kontinuierliche Integration gewährleistet werden, ohne die bereits angebotenen Systemkomponenten zu beeinträchtigen.

AP4: Simulationsumgebung

TAP 4.3: Automatisierte Szenariengenerierung

Das FZI entwickelte ein Konzept zur automatisierten Erzeugung von Übungsszenarien in der Simulation mit unterschiedlichen Schwierigkeitsstufen (sehr leicht, leicht, mittel, schwer, sehr schwer). Bereits beim Konzeptentwurf wurde darauf geachtet, dass nicht nur eine klare Abgrenzung zwischen den Schwierigkeitsstufen besteht, sondern auch innerhalb einer Schwierigkeitsstufe eine ausreichende Varianz an Szenarien generiert werden kann. Dadurch wird verhindert, dass sich Trainingssituationen wiederholen oder vorhersehbar werden. Das Konzept zur automatisierten Szenariengenerierung ist in Abbildung 4 schematisch dargestellt.

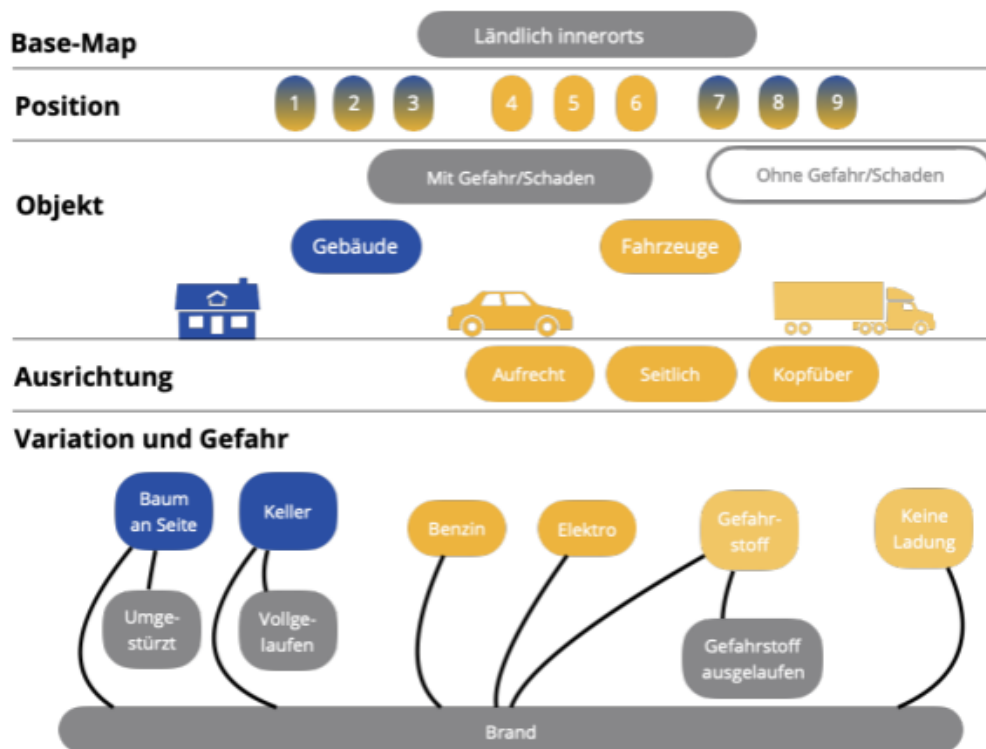


Abbildung 4: Konzept zur automatisierten Szenariengenerierung

In dem erarbeiteten Konzept wird das Übungsszenario schrittweise und modular aufgebaut. Die Szenarienkonstruktion erfolgt entlang einer definierten Entscheidungsstruktur, in der in jedem Schritt Parameter festgelegt werden, die maßgeblich die Charakteristik und Schwierigkeit des Szenarios beeinflussen. Hierzu zählen beispielsweise die Positionierung der Gefahrenstelle, die Art und Dynamik der Gefahr, Sicht- und Umweltbedingungen, zeitliche Randbedingungen oder das Verhalten beteiligter Akteure.

Jede getroffene Entscheidung wirkt sich auf einen kumulativen Schwierigkeitswert aus, der während des Durchlaufens der Entscheidungsstruktur berechnet wird. Dieser Wert entsteht durch die gewichtete Aggregation einzelner Einflussfaktoren. Komplexere, unerwartetere oder zeitkritischere Konfigurationen erhöhen dabei den Schwierigkeitsgrad stärker als einfache, klar strukturierte Situationen.

Da die zugrunde liegende Entscheidungsstruktur statisch definiert ist, können die resultierenden Schwierigkeitswerte für alle möglichen Entscheidungskombinationen im Vorfeld berechnet und in einer Referenzmatrix hinterlegt werden. Dies ermöglicht eine gezielte Auswahl von Szenarien mit exakt definierter Schwierigkeit. Zur Generierung eines Szenarios einer bestimmten Schwierigkeitsstufe wird folglich aus einer geeigneten Teilmenge von Entscheidungsoptionen gewählt, deren kombinierter Schwierigkeitswert innerhalb des gewünschten Wertebereichs liegt. Ein wesentlicher Vorteil dieses Ansatzes liegt in der Nachvollziehbarkeit und Reproduzierbarkeit der Szenarien. Da jede Konfiguration eindeutig über ihre Entscheidungsparameter beschrieben ist, können Szenarien exakt erneut erzeugt oder systematisch variiert werden. Darüber hinaus erlaubt das Konzept eine spätere Erweiterung, indem zusätzliche Entscheidungsdimensionen oder Gewichtungsfaktoren ergänzt werden, ohne die Grundstruktur verändern zu müssen.

Im Rahmen dieses Arbeitspaketes wurde außerdem die Simulationsumgebung von FZI-Seite bei jeder neu verfügbaren Iteration systematisch getestet. Dies umfasste eine strukturierte Funktionsprüfung, gezielte Tests definierter Szenarienkombinationen sowie eine explizite Suche nach Softwarefehlern. Auffälligkeiten im Systemverhalten wurden analysiert, dokumentiert und hinsichtlich ihrer Auswirkungen auf die Szenariengenerierung bewertet.

Für identifizierte Softwarebugs oder unerwartetes Systemverhalten wurde jeweils eine detaillierte Beschreibung inklusive reproduzierbarer Testschritte erstellt. Diese enthielt Angaben zur Systemversion, zu den gewählten Szenarioparametern, zum erwarteten Verhalten sowie zum tatsächlich beobachteten Verhalten. Dadurch konnte eine effiziente Fehlerbehebung unterstützt und die Stabilität sowie Qualität der Simulationsumgebung iterativ verbessert werden.

AP5: Automatisierte Auswertung & Handlungserkennung

TAP 5.1: Datenextraktion & -analyse

Zu den in TAP 2.1 definierten Metriken wurde jeweils bei jeder Metrik ein Messparameter definiert, d.h. eine konkrete messbare Größe, die die Metrik abbildet. Die Metriken und Messparameter sind in Tabelle 3: Metriken und zugehörige Messparameter dargestellt.

Tabelle 3: Metriken und zugehörige Messparameter

Metrik	Untermetriken	Beschreibung	Messung
Benötigte Gesamtzeit		Wie effizient nutzt der Nutzende seine verfügbare Zeit?	Benötigte Zeit in Sekunden
Gelaufener Weg		Wie effizient war der vom Nutzenden gewählte Weg?	Vergleich mit dem optimalen Pfad
Erfassung Kontrollmarkierungen		Wie viele Kontrollmarkierungen hat der Nutzende erfasst bzw. behandelt?	
	Abgelaufene Zugänge	Wie viele Zugänge zum Gebäude hat der Nutzende überprüft?	Anzahl der abgelaufenen Zugänge
	Getestete Zugänge	Bei wie vielen Zugängen zum Gebäude hat der Nutzende geprüft, ob sie begehbar sind?	Anzahl der getesteten Zugänge
	Erkannte Gefahren	Wie viele Gefahrenstellen hat	Anzahl der

		der Nutzende erkennt bzw. markiert?	erkannten Gefahren
	Gesichtete Bereiche	Wie groß war der vom Nutzenden erkundete Gesamtbereich?	Anzahl der gesichteten Bereiche
	Befragung	Wurde die Befragung des Zivilisten-NPCs durchgeführt?	Befragung Ja/Nein
Aufmerksamkeit		Wie viel Zeit hat der Nutzende im Durchschnitt benötigt, um eine Gefahr zu erkennen?	Verweildauer des Sichtkegels auf der Gefahr bevor sie markiert wird in Sekunden
Vorsicht/ Sicherheit/ Übertretene Gefahren		Wie viele Gefahrenüberschreitungen (z.B. Betreten des Gefahrenbereichs) sind während der Durchführung der Übung aufgetreten?	Anzahl der Schadensevents
Erinnern		Wie viele Einträge in der Gefahrenmatrix hat der Nutzende korrekt ausgefüllt?	Anzahl korrekte Einträge
Unsicherheit		Wie viel Zeit hat der Nutzende benötigt, um die Gefahrenmatrix auszufüllen?	Benötigte Zeit zum Ausfüllen der Matrix in Sekunden
Frontalansicht		Bewertung der Durchführung einer Lage auf Sicht zu Beginn des Szenarios	
	Frontalansicht Korrektheit	Wurde die Frontalansicht korrekt durchgeführt?	Anzahl korrekte Einträge
	Frontalansicht Zeit	Wie schnell wurde die Frontalansicht durchgeführt?	Dauer des Ausfüllens der Matrix in Sekunden
Stress/Forderung		Wie verhält sich das Stresslevel während dem Szenario?	Stressmessung über externe Sensorik

Für die Auswertung der Metriken wurden zunächst empirische Schwellwerte festgelegt, um die erfassten Messparameter in drei qualitative Stufen – gut, mäßig und schlecht – einordnen zu können. Diese Kategorisierung erlaubt eine schnelle und verständliche Interpretation der Ergebnisse und erleichtert Vergleiche zwischen unterschiedlichen Testbedingungen oder Probandengruppen. Die Festlegung der Schwellwerte erfolgte auf Basis vorläufiger Datenanalysen und Erfahrungswerte aus ersten Testläufen. Während des Projektverlaufs wurden die Schwellwerte kontinuierlich überprüft und im Rahmen interner Testungen angepasst, um die Aussagekraft und Relevanz der Bewertungen zu erhöhen. Dabei flossen sowohl statistische Kennzahlen wie Mittelwerte, Varianz und Extremwerte als auch qualitative Beobachtungen der Testpersonen ein. Durch diesen iterativen Anpassungsprozess konnte sichergestellt werden, dass die Einstufungen die realen Belastungssituationen und Leistungseinschätzungen der Einsatzkräfte möglichst genau widerspiegeln. Gleichzeitig erlaubt die abgestufte Bewertung, Abweichungen frühzeitig zu erkennen und gegebenenfalls gezielte Maßnahmen zur Optimierung von Trainings- oder

Interventionsstrategien abzuleiten.

Optional könnte man auch eine grafische Visualisierung der Schwellenwerte und ihrer Anpassungen einfügen, um die Nachvollziehbarkeit des Bewertungsprozesses zu erhöhen und die Transparenz für die Projektbeteiligten zu verbessern.

TAP 5.2: Handlungserkennung in der Simulation

Im Rahmen der Handlungserkennung werden verschiedene Metriken systematisch erfasst und ausgewertet. Hierzu zählen insbesondere die erreichten Kontrollmarkierungen, übertretene oder ignorierte Gefahren, der gewählte Laufweg, die benötigte Zeit sowie das beobachtete Stressverhalten. Diese Metriken ermöglichen eine differenzierte Analyse des konkreten Verhaltens während der Durchführung der Übung und bilden die Grundlage für eine objektivierte, automatisierte Bewertung.

Die Reihenfolge der erfüllten Kontrollmarkierungen in Kombination mit dem Laufweg gibt detaillierten Aufschluss darüber, wie sich die Person innerhalb des Szenarios orientiert und bewegt hat. Daraus lässt sich ableiten, ob ein strukturiertes, einsatztaktisch sinnvolles Vorgehen gewählt wurde oder ob die Abarbeitung eher zufällig und unsystematisch erfolgte. Beispielsweise kann ein kurzer, effizienter Laufweg mit logisch aufgebauter und fachlich korrekter Reihenfolge der Aufgaben auf ein zielgerichtetes Vorgehen hinweisen, während ein stark verzweigter oder redundanter Laufweg mit unlogischer Reihenfolge auf Unsicherheit oder mangelnde Struktur schließen lässt. Auf Basis dieser Einordnung kann eine fachliche Bewertung des taktischen Vorgehens erfolgen.

Die Betrachtung des Laufweges ist in Abbildung 5 schematisch dargestellt. In der vereinfachten Variante besteht das Ziel der Person (dargestellt durch den dunkelblauen Punkt) darin, sich zu mehreren Brandstellen (rote Quadrate) zu begeben und dabei alle vier Seiten der Gebäude (graue Quadrate) abzulaufen, um eine vollständige Erkundung zu simulieren. Während der Durchführung wird kontinuierlich der optimale Weg zum jeweils nächsten relevanten Zielpunkt berechnet (hellblauer Pfad) und mit dem tatsächlich gewählten Laufweg abgeglichen.

Als Bewertungsgrundlage dient hierbei primär die Länge des optimalen Pfades im Vergleich zur tatsächlich zurückgelegten Strecke. Je größer die Abweichung, desto stärker wird die Bewertungsmetrik negativ beeinflusst. Eine signifikante Verlängerung des Weges deutet auf ineffiziente Bewegung oder fehlende Priorisierung hin. Entscheidend ist dabei nicht der exakt berechnete optimale Pfadverlauf, sondern dessen minimale Länge, da der zugrunde liegende Algorithmus nicht zwingend alle möglichen optimalen Pfade identifiziert. Die Bewertung basiert somit auf einer robusten Längenapproximation statt auf einer exakten Routenübereinstimmung.

Bewertungsmodell. Dieses erlaubt eine konsistente, reproduzierbare und objektivierbare Einschätzung der im Szenario durchgeführten Handlungen.

Somit ermöglicht das entwickelte Modul eine automatische, nachvollziehbare und fachlich fundierte Handlungsbewertung innerhalb der Simulationsumgebung.

TAP 5.3: Einbezug zusätzlicher Sensorik

Auch das numerisch erfasste Stressverhalten ermöglicht einen differenzierten Einblick in das Verhalten während der Übungsdurchführung. Es liefert Hinweise darauf, ob eine Person ruhig, strukturiert und reflektiert agiert oder ob sie unter erhöhter psychophysiologischer Belastung steht und möglicherweise hektisch oder impulsiv handelt. Ziel ist dabei nicht die isolierte Bewertung eines „hohen“ oder „niedrigen“ Stressniveaus, sondern die Einordnung des Aktivierungszustandes im Kontext der Performanz und der situativen Anforderungen.

Zur objektiven Erfassung des Stressniveaus wurde ein externer EKG-Sensor eingesetzt. Sensoren zur Messung der elektrodermalen Aktivität (EDA) sowie des Blutvolumenpulses (BVP) wurden nicht berücksichtigt, da diese im Feldeinsatz anfälliger für Bewegungsartefakte sind und in dynamischen Einsatzszenarien eine geringere Signalstabilität aufweisen. Das EKG hingegen erlaubt eine robuste Erfassung der kardialen Aktivität auch unter körperlicher Belastung und liefert eine verlässliche Grundlage zur Analyse autonomer Regulationsmechanismen.

Stress manifestiert sich physiologisch primär über das autonome Nervensystem (ANS), das aus zwei funktionell gegensätzlichen Anteilen besteht¹:

- **Sympathikus:** Aktiviert bei Belastung, Gefahr oder erhöhter Anforderung („Fight-or-Flight“-Reaktion).
- **Parasympathikus:** Fördert Regeneration, Ruhe und Erholung.

Unter akuter Stressbelastung steigt die sympathische Aktivität, während die parasympathische Aktivität reduziert wird. Dies führt unter anderem zu einer Beschleunigung der Herzfrequenz, einer Erhöhung des Blutdrucks sowie zu einer Reduktion der Herzratenvariabilität (HRV).

Die Herzratenvariabilität beschreibt die zeitliche Variation zwischen zwei aufeinanderfolgenden Herzschlägen (RR-Intervalle). Entgegen einer intuitiven Annahme schlägt ein gesundes Herz nicht metronomisch gleichmäßig, sondern weist eine kontinuierliche Variabilität auf. Diese Variabilität ist Ausdruck einer adaptiven, flexiblen autonomen Regulation.

- **Hohe HRV:** Starke parasympathische Regulation, hohe Anpassungsfähigkeit, geringere akute Stressbelastung.
- **Niedrige HRV:** Dominanz sympathischer Aktivität, reduzierte vagale Modulation, erhöhte Stressbelastung.

Zur Stressbestimmung wurden etablierte HRV-Parameter herangezogen, die in der wissenschaftlichen Literatur als valide Marker psychophysiologischer Belastung gelten. Hierzu zählen beispielsweise zeitdomänenbasierte Kennwerte (z. B. RMSSD) sowie frequenzdomänenbasierte Parameter, die die Balance zwischen sympathischer und parasympathischer Aktivität widerspiegeln.

Da die Stressmessung über HRV-Parameter erfolgt, gilt: Ein niedriger HRV-Wert entspricht einer erhöhten Stressbelastung. In der nachfolgenden Darstellung bedeutet daher ein niedriger numerischer HRV-Wert einen hohen Stresszustand.

Ein Beispiel einer solchen Stressmessung ist im Folgenden anhand eines realen Einsatzes exemplarisch dargestellt. Während des Einsatzverlaufs wurden zwei Gefahrenlagen identifiziert, deren zeitliche Einordnung in Abbildung 6 beschrieben ist. Parallel dazu sind die HRV-Werte im Minutenmittel aufgetragen.

Typischerweise zeigt sich bei Annäherung an eine kritische Situation ein deutlicher Abfall der HRV, was auf eine erhöhte sympathische Aktivierung und damit auf eine gesteigerte Stressreaktion

¹ Task Force of the European Society of Cardiology and the North American Society of Pacing and Electrophysiology (1996). *Heart Rate Variability: Standards of Measurement, Physiological Interpretation, and Clinical Use*. *Circulation*, 93(5), 1043–1065. <https://www.ahajournals.org/doi/10.1161/01.cir.93.5.1043>

hinweist. Nach Bewältigung der Gefahrenlage ist – bei ausreichender Regulationsfähigkeit – wieder ein Anstieg der HRV zu beobachten, der die parasympathische Reaktivierung widerspiegelt. Durch diese zeitlich aufgelöste Betrachtung kann nicht nur das generelle Stressniveau beurteilt werden, sondern auch die dynamische Stressreaktion im Kontext konkreter Einsatzereignisse. Dies erlaubt eine differenzierte Analyse:

- Wie stark reagiert die Person auf kritische Situationen?
- Wie schnell erfolgt eine physiologische Erholung nach Belastung?
- Besteht ein dauerhaft erhöhtes Aktivierungsniveau?

In Kombination mit den leistungsbezogenen Metriken entsteht somit ein integratives Bild aus Verhalten und physiologischer Beanspruchung. Dies ermöglicht eine fundierte Bewertung der Belastungsreaktion im Kontext der Aufgabenerfüllung und liefert wertvolle Hinweise für Trainingsanpassungen, Szenarioskalierung und individuelle Kompetenzentwicklung.

Ergebnisse Stressmessung Realeinsatz

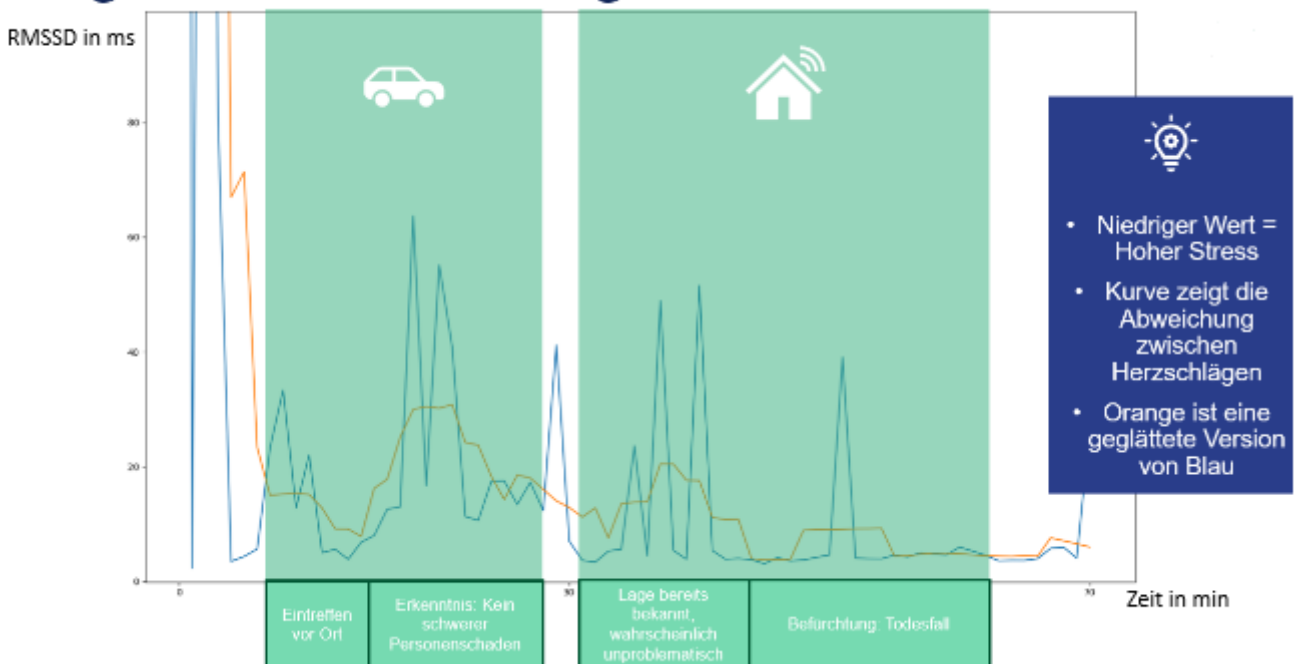


Abbildung 6: Gemessene Herzratenvariabilität im Verlauf eines Einsatzes

In der Grafik ist deutlich erkennbar, dass das Stresslevel eng mit der Beschreibung der Einsatzlage korreliert. So zeigt sich beispielsweise, dass der Stress beim Eintreffen an Lage 1 zunächst hoch ist und nachlässt, sobald die Erkenntnis gewonnen wird, dass keine Schwerverletzten vor Ort sind. Bei Lage 2 beginnt der Stress aufgrund der Vertrautheit mit der Örtlichkeit zunächst auf niedrigem Niveau, steigt jedoch stark an, sobald der Verdacht auf einen Todesfall aufkommt.

Da HRV-Parameter stark individuell und personenabhängig variieren, ist eine automatisierte Stressbewertung auf Basis der HRV-Daten alleine nicht zuverlässig möglich. Um dennoch ein belastbares Stress-Scoring zu ermöglichen, wurde ein KI-basiertes Verfahren entwickelt. Für das Training wurden öffentliche EKG-Datensätze mit Stresswerten aus Selbsteinschätzungen herangezogen, namentlich der SWELL-KW-Datensatz² und der MCD-rPPG-Datensatz³. Alle Datensamples wurden zunächst in ein einheitliches Format überführt und auf Plausibilität geprüft. Anschließend wurden die zuvor festgelegten HRV-Parameter für jedes Sample extrahiert. Für die KI-Modellierung wurden verschiedene Methoden auf Basis aktueller Forschungsarbeiten ausgewählt, darunter:

² <https://cs.ru.nl/~skoldijk/SWELL-KW/Dataset.html>

³ https://huggingface.co/datasets/kyegorov/mcd_rppg/tree/main

- Neuronale Netze
- XGBoost
- Support Vector Machine (SVM)
- K-nearest Neighbor (KNN)

Die Hyperparameter der Modelle wurden systematisch über ein Grid-Search-Verfahren⁴ optimiert. Dabei wurden für jede Modellarchitektur unterschiedliche Parameterkombinationen getestet, um diejenige Konfiguration zu identifizieren, die die beste Vorhersageleistung liefert. Die jeweils „besten“ Modelle beziehen sich also direkt auf die Ergebnisse dieser Grid-Search-Optimierung: Es handelt sich um die Modell- und Hyperparameterkombination, die auf den Trainings- und Validierungsdaten die höchste Genauigkeit und Plausibilität erzielte.

Zur Validierung wurden die Daten einerseits zufällig in Trainings- und Testsets im Verhältnis 80:20 aufgeteilt, andererseits wurde eine Plausibilitätsprüfung anhand der zur Verfügung stehenden EKG-Daten aus der Bedienung der VR-Simulation durchgeführt. Die Ergebnisse der jeweils besten Modelle, also derjenigen Konfigurationen, die nach der Grid-Search die optimale Vorhersageleistung zeigten, sind in Tabelle 4 dargestellt. Diese Vorgehensweise stellt sicher, dass die Modelle sowohl auf standardisierten Datensätzen als auch auf praxisnahen VR-Szenarien belastbare Vorhersagen liefern.

Tabelle 4: Ergebnisse der Modell-Validierung

Model	Genauigkeit
SVM	50.8%
XGB	90.7%
MLP	45.7%
KNN	19.3%

Abschließend wurde im Rahmen der Benutzbarkeit eine Begleitanwendung entwickelt, die eine grafische Oberfläche für die Verwendung des KI-Systems bietet, so dass der Nutzende die Auswertung unabhängig von der VR-Simulation einsehen kann. Die Anwendung ist in Abbildung 7 dargestellt. Mit der Anwendung kann der Nutzende EKG-Daten einlesen und analysieren lassen.

TAP 5.4: Datenaufbereitung und Darstellungskonzept

Die Aufbereitung der erhobenen Daten wurde in enger Zusammenarbeit mit der Northdocks GmbH umgesetzt und ist direkt in die Simulationsumgebung integriert. Dadurch erfolgt die Datenerfassung unmittelbar während der Szenariodurchführung, ohne dass eine nachgelagerte manuelle Aufbereitung erforderlich ist.

Erfasst werden die in TAP 5.1 definierten Messparameter. Diese umfassen sowohl positions- und zeitbezogene Daten als auch Ereignis- und Interaktionsinformationen, die während der Übung generiert werden. Die Rohdaten werden zunächst strukturiert gespeichert, zeitlich synchronisiert und hinsichtlich Vollständigkeit sowie Plausibilität geprüft. Anschließend erfolgt eine Vorverarbeitung, bei der die Daten normalisiert, aggregiert oder in geeignete Datenstrukturen überführt werden, um eine konsistente Weiterverarbeitung im Auswertungsmodul zu gewährleisten.

Auf Basis der aufbereiteten Messparameter wird für jede zugehörige Metrik eine eigenständige Bewertung berechnet. Hierbei kommen definierte Bewertungsfunktionen zum Einsatz, die die jeweiligen Parameter in normierte Kennwerte überführen. Diese können beispielsweise als

⁴ Bergstra, J., & Bengio, Y. (2012). *Random Search for Hyper-Parameter Optimization*. Journal of Machine Learning Research, 13, 281–305. <https://jmlr.org/papers/v13/bergstra12a.html>

Punktwerte, prozentuale Erfüllungsgrade oder kategorisierte Qualitätsstufen ausgegeben werden.

Durch diese strukturierte Pipeline – bestehend aus Datenerfassung, Validierung, Vorverarbeitung und metrikenspezifischer Bewertung – wird eine transparente, reproduzierbare und technisch robuste Grundlage für die automatische Handlungsbewertung geschaffen. Gleichzeitig ermöglicht die direkte Integration in die Simulationsumgebung eine effiziente Echtzeit- oder Near-Echtzeit-Auswertung der durchgeführten Szenarien.

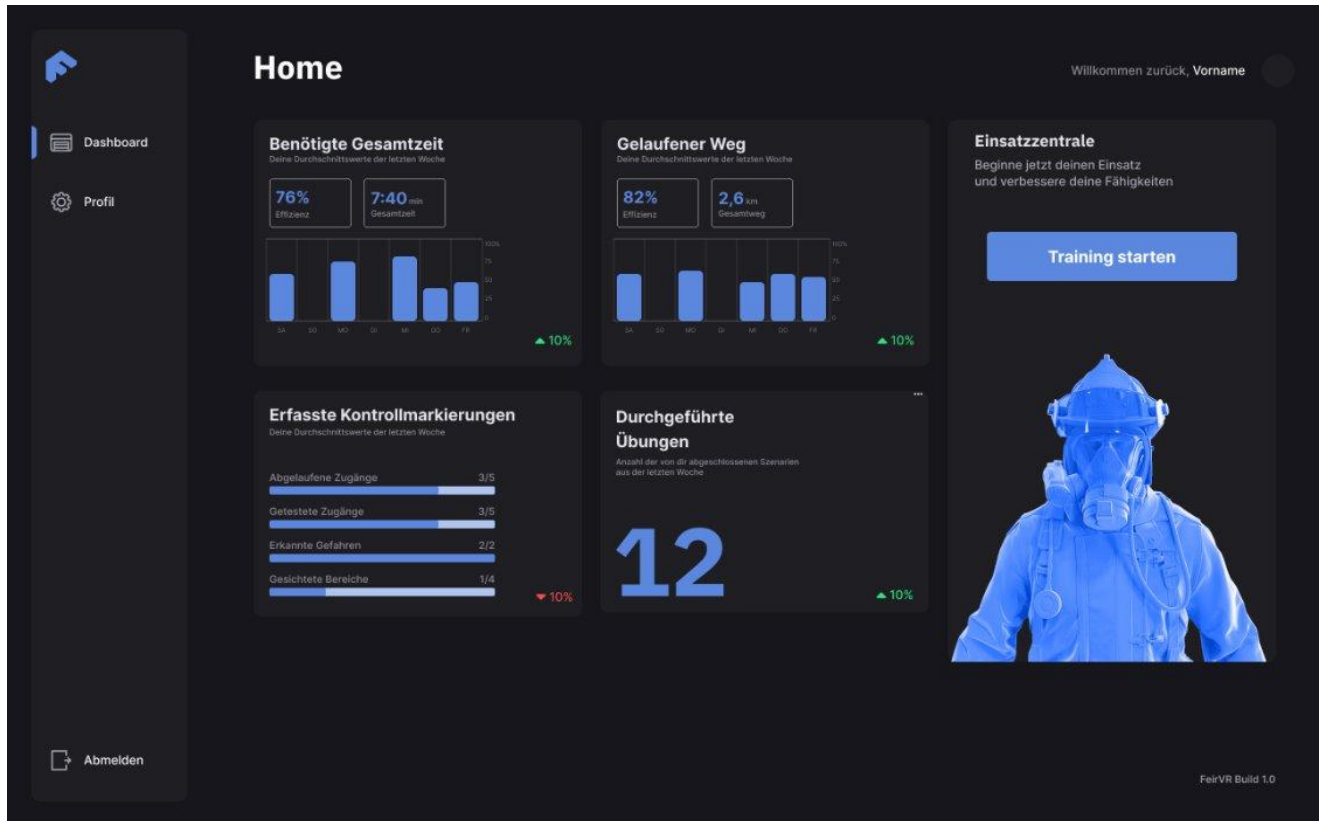


Abbildung 7: Grafische Oberfläche der Anwendung

TAP 5.5: Implementierung Auswertekonzept

Das Auswertekonzept wurde erfolgreich in einem dedizierten Auswertungsmodul in Python implementiert. Die Softwarearchitektur ist dabei modular aufgebaut, sodass jede Messgröße in einer eigenständigen Verarbeitungseinheit analysiert und bewertet wird. Dieses komponentenbasierte Design ermöglicht eine klare Trennung der einzelnen Bewertungslogiken sowie eine transparente Nachvollziehbarkeit der Berechnungsschritte.

Die jeweiligen Messparameter – beispielsweise Laufweg, Kontrollmarkierungen, Zeitbedarf oder Stressindikatoren – werden isoliert voneinander verarbeitet. Dadurch wird sichergestellt, dass die Berechnung einer Metrik nicht von der Verfügbarkeit oder Korrektheit anderer Messwerte abhängt. Im Sinne der Robustheit und Fehlertoleranz werden fehlende, unvollständige oder fehlerhafte Eingabedaten abgefangen und innerhalb der jeweiligen Komponente behandelt. Dies erfolgt etwa durch definierte Fallback-Mechanismen, Plausibilitätsprüfungen oder das Setzen neutraler Bewertungswerte.

Durch diese Entkopplung bleibt das Gesamtsystem auch dann funktionsfähig, wenn einzelne Datenquellen temporär nicht verfügbar sind oder einzelne Messwerte nicht valide berechnet werden können. Die übrigen Metriken werden weiterhin regulär ausgewertet und fließen unverändert in die Gesamtbewertung ein.

Zusätzlich erlaubt die modulare Struktur eine einfache Erweiterbarkeit des Systems. Neue Bewertungsmetriken können ergänzt werden, ohne bestehende Komponenten zu verändern. Ebenso können Gewichtungen oder Bewertungslogiken einzelner Parameter angepasst werden, ohne die Stabilität des Gesamtsystems zu beeinträchtigen.

Insgesamt gewährleistet die gewählte Implementierungsstrategie eine robuste, wartbare und skalierbare Umsetzung des Auswertekonzepts im Python-basierten Auswertungsmodul.

TAP 5.6: Integration in Grundsystem

Das Auswertungsmodul wurde über die zuvor definierte Schnittstelle vollständig in das Gesamtsystem integriert. Die Anbindung erfolgte entsprechend der spezifizierten Datenformate, Übergabeprotokolle und Triggermechanismen, sodass eine konsistente Kommunikation zwischen Simulationsumgebung, Datenerfassung und Auswertung gewährleistet ist.

Zur Validierung der Schnittstellenintegration wurde der vollständige Datenfluss systematisch getestet. Hierbei wurde der gesamte Prozess – von der Generierung eines Szenarios über die Durchführung der Simulation und die Erfassung der Rohdaten bis hin zur Übergabe an das Auswertungsmodul und der Berechnung der Bewertungsmetriken – end-to-end durchlaufen. Dadurch konnte sichergestellt werden, dass alle relevanten Daten korrekt übertragen, vollständig verarbeitet und in der erwarteten Struktur ausgegeben werden.

Neben der reinen Funktionsprüfung wurde auch die Datenkonsistenz überprüft. Dies umfasste unter anderem die Kontrolle von Zeitstempeln, eindeutigen Identifikatoren, Parameterbereichen sowie die Validierung der berechneten Ergebnisse im Vergleich zu Referenzszenarien. Auf diese Weise konnte die fehlerfreie Kommunikation zwischen den beteiligten Systemkomponenten nachgewiesen werden.

Zur Sicherstellung der praktischen Anwendbarkeit wurde ergänzend eine detaillierte Bedienungsanleitung erstellt. Diese beschreibt die Systemvoraussetzungen, die Einbindung in das Gesamtsystem, die Konfiguration relevanter Parameter sowie den Ablauf der Auswertung. Darüber hinaus werden die interne Funktionsweise des Moduls, die zugrunde liegenden Bewertungslogiken sowie typische Anwendungsfälle erläutert.

Die Dokumentation enthält außerdem Hinweise zur Fehlerdiagnose, zur Interpretation der erzeugten Bewertungsmetriken sowie zu Erweiterungs- und Anpassungsmöglichkeiten. Dadurch wird sowohl die Nutzung im operativen Betrieb als auch die Weiterentwicklung durch technische Anwenderinnen und Anwender unterstützt.

AP6: Validierung & Evaluation

TAP 6.1: Workshops und Nutzendenevaluation

Das FZI unterstützte im Projektverlauf mehrere Evaluations- und Transferformate zur praxisnahen Erprobung des entwickelten Gesamtsystems. Hierzu zählten Workshops am Fraunhofer IAO im Rahmen des „Runden Tisch Bevölkerungsschutz“ am 21.10.2025 sowie eine strukturierte Evaluation mit der Freiwillige Feuerwehr Plön.

Darüber hinaus führte das FZI eine eigenständige Evaluation mit der Freiwillige Feuerwehr Ubstadt-Weiher durch. Ergänzend wurden interne Kolleg*innen einbezogen, die selbst Mitglied einer Freiwilligen Feuerwehr sind und somit über praktische Einsatzerfahrung verfügen. Dadurch konnte sowohl externes als auch internes Fachfeedback in die Bewertung einfließen.

Im Rahmen der Evaluation wurde das Gesamtsystem mit mehreren Szenarien pro Person getestet, um unterschiedliche Schwierigkeitsgrade, Handlungsmuster und Reaktionsweisen abzubilden. Ziel war es, nicht nur die technische Funktionsfähigkeit, sondern insbesondere die fachliche Korrektheit der automatisierten Bewertung sowie den didaktischen Mehrwert der dargestellten Metriken zu überprüfen.

Dabei wurde unter anderem untersucht:

- ob die automatisch generierten Bewertungen mit der fachlichen Einschätzung erfahrener Einsatzkräfte übereinstimmen,
- ob die dargestellten Kennzahlen nachvollziehbar und transparent sind,
- und ob die Auswertung eine reflektierte Nachbesprechung (Debriefing) sinnvoll unterstützt.

Die Rückmeldungen der Proband*innen waren insgesamt geschlossen positiv. Besonders hervorgehoben wurden die strukturierte Darstellung der Ergebnisse sowie die Möglichkeit, eigenes taktisches Vorgehen objektiviert nachvollziehen zu können.

Als Verbesserungsvorschlag wurde mehrfach der Wunsch nach einer erweiterten Replay-Funktion geäußert. Diese sollte eine detaillierte zeitliche Aufschlüsselung zentraler Entscheidungspunkte und Schlüsselmomente enthalten, um Handlungsabläufe noch gezielter analysieren und im Rahmen der Ausbildung nachbesprechen zu können. Dieses Feedback liefert wertvolle Ansatzpunkte für die Weiterentwicklung des Systems, insbesondere im Hinblick auf die didaktische Aufbereitung und Trainingsunterstützung.

TAP 6.2: Erstellung Domänenbacklog und Anwendungsleitfaden

Ein Domänenbacklog wurde in gemeinsamer Abstimmung erstellt und wird von AntwortING gehostet. Alle Anmerkungen aus End- und Zwischenevaluationen wurden sorgfältig erfasst und korrekt abgelegt, sodass eine lückenlose Dokumentation aller Rückmeldungen gewährleistet ist. Die Struktur der einzelnen Einträge orientiert sich an der Darstellung in Abbildung 8, die exemplarisch Aufbau und Inhalt eines typischen Backlog-Eintrags zeigt. Dies erleichtert sowohl die Nachvollziehbarkeit der Änderungen als auch die spätere Weiterverarbeitung der Einträge.

Titel des Vorschlags*	Funktionslose Buttons Weboberfläche
Beschreibung*	zwei Buttons (Profil & Dashboard) auf Weboberfläche (feir.app) ohne Funktion
Kategorie* (Funktionalität, Technisch, UI/UX, Bugfixes, Systemkomponenten, Anderes)	Kategorie: Bugfixes
Priorität* (hoch, mittel, gering)	Priorität: mittel
Name der einreichenden Person*	[REDACTED]
Akzeptanzkriterien Technischer Aufwand: Forschungswert Lernerfolgswert	Erfüllt wenn: Buttonfunktion hergestellt Aufwand: 3h Forschungswert: mittel Lernerfolgswert: hoch

Abbildung 8: Struktur der Einträge im Domänenbacklog

Der Anwendungsleitfaden wurde federführend vom Fraunhofer IAO erstellt. Das FZI lieferte ergänzende Zuarbeiten, insbesondere zur Analyse der Beziehung zwischen Stress und Leistung sowie zu den Auswirkungen von chronischem Stress bei Einsatzkräften. Dies spiegelt sich insbesondere in Kapiteln wie „Stress – was passiert im Körper und Kopf“ oder „Stress als Stimulus“ wider. Durch diese Zusammenarbeit konnte sichergestellt werden, dass sowohl theoretische Grundlagen als auch praxisrelevante Erkenntnisse systematisch im Leitfaden abgebildet sind, wodurch er eine fundierte Basis für die Anwendung in realen Einsatzkontexten bietet.

2 Wichtigste Positionen des zahlenmäßigen Nachweises

Die Durchführung des Projekts erforderte vor allem Personalkosten. Dazu zählen wissenschaftliche Mitarbeitende, die zentrale Forschungs- und Analyseaufgaben übernommen haben, eine erfahrene Führungskraft für Projektkoordination und Berichtswesen sowie wissenschaftliche Hilfskräfte für unterstützende Recherche- und Implementierungstätigkeiten. Weitere bedeutende Kostenpositionen entstanden durch Reisen zu Partnerinstitutionen zum Austausch und Vernetzung sowie zur Präsentation der Ergebnisse auf einer wissenschaftlichen Konferenzen (ICBES2025/EECSS'25). Zusätzlich fielen Materialkosten für die Entwicklung des VR-Systems/-Demonstrators an, insbesondere für die Anschaffung von VR-Brille und zugehöriger Hardware. Insgesamt spiegeln diese Positionen die wesentlichen Aufwendungen wider, die für die erfolgreiche Durchführung des Projekts notwendig waren.

3 Notwendigkeit und Angemessenheit der geleisteten Projektarbeiten

Die Arbeiten des FZI im Teilvorhaben sowie im Gesamtprojekt waren notwendig und angemessen, um Verhalten und Leistungsfähigkeit von Einsatzkräften objektiv erfassen zu können. Besonders relevant waren die Entwicklung der automatisierten Szenariengenerierung mit differenzierten Schwierigkeitsstufen und Varianz innerhalb der Szenarien sowie die Integration von

Stressmessungen über HRV, um Handlungsreaktionen unter Belastung zu analysieren. Durch die Operationalisierung von Kompetenzen in messbare Metriken und die modulare Auswertung konnte eine automatisierte, reproduzierbare und fachlich valide Bewertung komplexer Handlungsabläufe realisiert werden. Die iterative Validierung mit Einsatzkräften sicherte dabei Praxisrelevanz und didaktischen Mehrwert des Systems.

4 Voraussichtlicher Nutzen, insbesondere der Verwertbarkeit des Ergebnisses im Sinne des fortgeschriebenen Verwertungsplans

Das Projekt feir liefert mit dem entwickelten VR-basierten Ausbildungssystem für Einsatz- und Führungskräfte kombiniert mit den Arbeiten des FZI im Teilvorhaben eine technisch und wissenschaftlich anschlussfähige Grundlage für die automatisierte Analyse und Bewertung komplexer Handlungsabläufe. Die Kombination aus automatisierter Szenariengenerierung, objektiver Leistungsbewertung und sensorbasierter Stressanalyse ermöglicht eine praxisnahe, skalierbare und nachvollziehbare Trainingsgestaltung. Die erzielten Ergebnisse sind sowohl für die Aus- und Weiterbildung von Einsatzkräften als auch für weiterführende Forschungs- und Transferprojekte verwertbar und bieten hohes Potenzial für die Integration in bestehende Ausbildungssysteme, die Optimierung didaktischer Konzepte und die Weiterentwicklung simulationsbasierter Trainingsmethoden.

5 Während der Durchführung des Vorhabens dem ZE bekannt gewordenen Fortschritts auf dem Gebiet des Vorhabens bei anderen Stellen

Während der Durchführung des Vorhabens sind dem Zuwendungsempfänger keine Ergebnisse oder Entwicklungen anderer Stellen bekannt geworden, die der Durchführung, Zielerreichung oder nachgelagerten Verwertung dieses Vorhabens entgegenstehen könnten. Auch konnten keine kürzlich abgeschlossenen Forschungsprojekte oder Verwertungsaktivitäten identifiziert werden, die die Zielsetzung dieses Vorhabens substituieren würden.

6 Erfolgten oder geplanten Veröffentlichungen des Ergebnisses

Während und nach Ende der Projektlaufzeit wurden verschiedene Veröffentlichungen zu den Ergebnissen und Teilergebnissen auf Fachkonferenzen und sonstigen Veranstaltungen eingereicht, publiziert und präsentiert.

Erfolgte wissenschaftliche Publikation:

- D. Birkenmaier, Wu, L. Schweickart, W. Stork, „Towards Using the Yerkes-Dodson Law to Select Optimal Training Difficulty in Firefighter Simulations: A Machine Learning Based ECG Approach“, Proceedings of the 11th World Congress on Electrical Engineering and Computer Systems and Sciences (EECSS'25), Paris, France, August, 2025.

FZI House-of-Living-Labs:

- Permanente Vorstellung aktueller Entwicklungen bei Führungen für Expert*innen und End-Anwendenden
- feir-Demonstrator
- FZI-Open House (2023, 2024, 2025, 2026)

Dissertation:

- Dissertation Dennis Birkenmaier (vorr. 2027)