



**Laufzeit:** 01.01.2022 – 31.12.2024  
Zuwendungsneutrale Laufzeitverlängerung bis 31.03.2024

## **Abschlussbericht IZEW**

### **Teil II: Darstellung der Arbeiten**

#### **Projektteam:**

Dr. Wulf Loh (Leitung 01/2022 – 03/2025)  
Dr. Simon David Hirsbrunner (Co-Leitung 03/2022 – 12/2023)  
Sol Martinez Demarco (WiMi 06/2023 – 05/2024)  
Dr. Lou Therese Brandner (Co-Leitung 07/2024 – 03/2025)  
Theresa Krampe (WiMi 07/2024 – 12/2024)

**Internationales Zentrum  
für Ethik in den Wissenschaften**  
Universität Tübingen  
Wilhelmstr. 19  
72074 Tübingen

## Inhaltsverzeichnis

<b>Berichtsblatt.....</b>	<b>3</b>
<b>1. Eingehende Darstellung der Arbeiten .....</b>	<b>5</b>
<i>AP 1: Anforderungsanalyse und Spezifikationen .....</i>	<i>5</i>
<i>AP 2: Ethische Anforderungen.....</i>	<i>6</i>
<i>AP 3: Rechtliche Anforderungen.....</i>	<i>10</i>
<i>AP 4: Gesichtserkennung.....</i>	<i>11</i>
<i>AP 5: Textauswertung .....</i>	<i>12</i>
<i>AP 6: Sprechererkennung .....</i>	<i>13</i>
<i>AP 7: Objektdetektion .....</i>	<i>14</i>
<i>AP 8: Methodenspektrum, Prüfkataloge und Standardisierung .....</i>	<i>15</i>
<b>2. Wichtigste Positionen des zahlenmäßigen Nachweises .....</b>	<b>15</b>
<b>3. Notwendigkeit und Angemessenheit der geleisteten Projektarbeiten.....</b>	<b>16</b>
<b>4. Voraussichtliche Nutzen, insbesondere die Verwertbarkeit des Ergebnisses im Sinne des fortgeschriebenen Verwertungsplans .....</b>	<b>17</b>
<b>5. Während der Durchführung des Vorhabens dem Zuwendungsempfänger bekannt gewordenen Fortschritt auf dem Gebiet des Vorhabens bei anderen Stellen.....</b>	<b>18</b>
<b>6. Erfolgte und geplante Veröffentlichungen der Ergebnisse.....</b>	<b>18</b>
<b>7. Im Bericht verwendete Literatur .....</b>	<b>21</b>

## Berichtsblatt

<b>1. ISBN oder ISSN</b>	<b>2. Berichtsart (Schlussbericht oder Veröffentlichung)</b> Schlussbericht	
<b>3. Titel</b> Vertrauenswürdige Künstliche Intelligenz für polizeiliche Anwendungen (VIKING)  Teilvorhaben: Bedingungen für eine Operationalisierung ethischer Anforderungen an vertrauenswürdige KI in polizeilichen Anwendungen		
<b>4. Autor(en) [Name(n), Vorname(n)]</b>  Brandner, Lou Hirsbrunner, Simon Krampe, Theresa Loh, Wulf Martinez Demarco, Sol	<b>5. Abschlussdatum des Vorhabens</b> 31.03.2025	<b>6. Veröffentlichungsdatum</b>
	<b>7. Form der Publikation</b>	
	<b>8. Durchführende Institution(en) (Name, Adresse)</b> Internationales Zentrum für Ethik in den Wissenschaften Universität Tübingen Wilhelmstr. 19 72074 Tübingen	<b>9. Ber. Nr. Durchführende Institution</b>
<b>11. Seitenzahl</b> 22		
<b>12. Fördernde Institution (Name, Adresse)</b> Bundesministerium für Forschung, Technologie und Raumfahrt (BMFTR) Kapelle-Ufer 1 10117 Berlin		<b>13. Literaturangaben</b> 27
	<b>15. Abbildungen</b> 2	
	<b>16. Zusätzliche Angaben</b>	
<b>17. Vorgelegt bei (Titel, Ort, Datum)</b>		

**18. Kurzbeschreibung des Teilvorhabens:**

Die Polizeiarbeit unterliegt ebenso wie andere gesellschaftliche Bereiche der Digitalisierung, die eine Vielzahl neuer Überwachungs- und Aufklärungsmöglichkeiten bietet. Der Einsatz von Softwarelösungen auf der Basis von künstlicher Intelligenz – insbesondere mit Hilfe von Methoden des maschinellen Lernens – in diesem Bereich polizeilicher Arbeit zur Auswertung des anfallenden Datenmaterials ist mit einer Vielzahl von Problemen verbunden, wie z.B. Verzerrungen, Fehlschlüssen und übermäßigem technischen „Solutionismus“ sowie ethischen Fragen. Einerseits besteht ein ethischer Reflexionsbedarf über Definitionen und das Verhältnis von Vertrauen, Nachvollziehbarkeit, Transparenz und Fairness im Zusammenhang mit KI-Technologien. Zum anderen stehen Fragen der Operationalisierbarkeit von ethischen Anforderungen im Hinblick auf den konkreten Anwendungskontext im Vordergrund. In diesem Teilprojekt „Bedingungen für die Operationalisierung ethischer Anforderungen an vertrauenswürdige KI in polizeilichen Anwendungen“ werden eine Risikomatrix und ein Katalog ethischer Anforderungen für den KI-Einsatz in der Polizeiarbeit vorgeschlagen. Dieser operationalisiert die Grundprinzipien vertrauenswürdiger KI im Lichte der neuen EU-Anforderungen (EU AI Act) sowie anderer Operationalisierungsinitiativen. Darüber hinaus wird zwischen verschiedenen Gruppen von Adressaten unterschieden, indem Anforderungen speziell für Anbieter:innen und Anwender:innen definiert werden.

**19. Schlagwörter**

Datenethik, KI-Ethik, Technikethik, polizeiliche Ermittlungsarbeit

**20. Verlag**

**21. Preis**

## 1. Eingehende Darstellung der Arbeiten

Die eingehende Darstellung orientiert sich in der Struktur an den in der Teilvorhabenbeschreibung ausgeführten Arbeitspaketen. Es werden diejenigen Arbeitspakete behandelt, in denen das IZEW mindestens Zuarbeiten erbracht hat. Unter-APs, in denen das IZEW nicht beteiligt war, werden der Übersichtlichkeit halber nicht aufgelistet.

### AP 1: Anforderungsanalyse und Spezifikationen

#### Verwendung der Zuwendung:

Die Zuwendung wurde überwiegend für folgende Positionen verwendet:

- Beschäftigungsentgelte für wissenschaftliche Mitarbeitende
- Studentische Hilfskräfte (Literaturrecherchen, Korrekturlesen)

#### **AP 1.1: Anforderungsanalysen**

#### **AP 1.2: Methodenspezifikation**

#### Ziele:

AP 1 bildete die Grundlage von VIKING. In diesem Arbeitspaket wurden die ethischen Aspekte für einen Anforderungskatalog aus Normen und Standards als Basis der Spezifikation der Demonstratoren für AP4–AP7 identifiziert und mit den Bedarfen der unterschiedlichen Anwender:innengruppen (Polizeibesetzte, KI-Expert:innen, Justizbeschetzte) in Einklang gebracht. Zudem wurden bestehende Kompetenzen, Leerstellen und Verbesserungsbedarfe identifiziert.

#### Ergebnisse:

Es wurde eine Literatursammlung zu digitaler Polizeiarbeit, Critical Algorithm and Data Studies sowie KI- und Datenethikforschung angelegt und analysiert. Die ethischen Anforderungen wurden auf Basis der Literaturrecherche sowie auf Gesprächen mit den technischen, anwendungsbezogenen und rechtlichen Projektpartnern identifiziert und auf den Anwendungskontext zugeschnitten. Dazu wurden in enger Verzahnung mit AP2 und AP3 (EU-)ethische Anforderungen an die zu realisierenden Demonstratoren erfasst und analysiert. Weiterhin wurden Bedarfe der Anwender:innen, Anforderungen an die Gerichtsverwertbarkeit und der aktuelle Stand von Normung, Standardisierung und Zertifizierung erfasst und mit ethischen Anforderungen abgeglichen. Während die rechtlichen Anforderungen hier als Minimalstandard dienen (Datenschutz, Gerichtsverwertbarkeit), gehen die ethischen darüber hinaus, indem sie sowohl inhaltlich als auch die Operationalisierungen betreffend die existierenden Normen weiter in Richtung Vertrauenswürdigkeit spezifizieren und auf diese Weise Best Practices ausdifferenzieren.

Vier Themen wurden in einem frühen Stadium der Technologieentwicklung als ethisch wirksam identifiziert: Wertesensitives Design, Schutz vor algorithmischer und menschlicher Diskriminierung, Transparenz und Erklärbarkeit, sowie Schutz der Privatheit. Für diese Thematiken wurde ein Arbeitspapier mit ethischen Anforderungen an die technischen Komponenten (AP4-7) in VIKING erstellt, mit den Partner:innen diskutiert und in das vom Gesamtverbund entwickelte Spezifikationsdokument (GVB AP1) integriert. Weitere ethische Bedenken und Anforderungen (Accountability, menschliche Kontrolle, Aufsicht und Letztentscheidung) wurden für die späteren Phasen der Entwicklung der Demonstratoren identifiziert und zur weiteren Analyse vorgemerkt. Alle diese als ethisch wirksam identifizierten Thematiken flossen in Absprache mit Projektpartner:innen in

die Erstellung des ethischen Anforderungskataloges ein.

## **AP 2: Ethische Anforderungen**

### Verwendung der Zuwendung:

Die Zuwendung wurde überwiegend für folgende Positionen verwendet:

- Beschäftigungsentgelte für wissenschaftliche Mitarbeitende
- Studentische Hilfskräfte (Literaturrecherchen, Assistenz bei Schreibprojekten, administrative Unterstützung)
- Reisegelder für Projekttreffen und Konferenzen
- Benötigte Geräte für Forschungsvorhaben: Apple MacBook Air 13"

Da AP 2 ganz dem ethischen Teilprojekt gewidmet war, werden die Arbeiten in den Unter-APs hier ausdifferenzierter und ausführlicher dargestellt für die anderen APs.

### **AP 2.1: Verfahren zur ethischen Bewertung von KI-Systemen**

#### Ziele:

- (1) Entwicklung von Verfahren und einer Risikomatrix zur ethischen Bewertung von KI-Systemen
- (2) Etablierung einer geeigneten Eingriffsevaluation

#### Ergebnisse:

Zur Operationalisierung der ethischen Bewertung von KI-Systemen und einer Spezifizierung des Kriteriums der Vertrauenswürdigkeit wurde eine Risikomatrix erarbeitet, die die Eingriffstiefe des Systems in die Rechte und Interessen von Betroffenen abschätzen und klassifizieren kann. Hierfür wurde zunächst eine umfassende Literaturrecherche und -analyse zu ethischen Risiko- und Folgenabschätzungen sowie zu Risikomatrizen durchgeführt. Diese diente als Grundlage für die Entwicklung der Risikomatrix, deren grobgranulare Fassung im Meilensteinbericht in Form eines umfangreichen Arbeitspapiers vorgelegt wurde. Unter anderem wurden in diesem Zusammenhang bestehende Ansätze wie die Risiko-Pyramide im zur damaligen Zeit erst entstehenden EU AI Act<sup>1</sup>, die Risiko-Matrix der AI Ethics Impact Group (AIEIG 2020), das Konformitätsassessment Cap.AI (Floridi et al. 2022) und verschiedene andere Modelle aus unterschiedlichen geographischen Regionen verglichen und evaluiert. Auf Basis dieser Untersuchung wurde entschieden, einen dynamischen und iterativen Ansatz der ethischen Bewertung anzuwenden.

In der zweiten Projekthälfte wurde die grobgranulare Fassung in einem iterativen Prozess und in enger Kooperation mit den technischen Partner:innen weiter verfeinert und ergänzt. Anhand der entstandenen Risikomatrix lassen sich Anwendungsszenarien im Einklang mit rechtlichen Anforderungen aus dem AI Act evaluieren. Die Matrix bietet den doppelten Vorteil einer einfachen, intuitiven Visualisierung von Risikobewertungen (siehe Abbildung 1) und verdeutlicht gleichzeitig die Abhängigkeit von soziotechnischen Kontexten. Elemente, die Systemfähigkeiten repräsentieren, werden in der Matrix abhängig von 1) der Intensität des potenziellen Schadens sowie der Wahrscheinlichkeit seines

---

<sup>1</sup> EU-KI-VO, Verordnung (EU) 2024/1689 des Europäischen Parlaments und des Rates vom 13. Juni 2024 zur Festlegung harmonisierter Vorschriften für künstliche Intelligenz und zur Änderung der Verordnungen (EG) Nr. 300/2008, (EU) Nr. 167/2013, (EU) Nr. 168/2013, (EU) 2018/858, (EU) 2018/1139 und (EU) 2019/2144 sowie der Richtlinien 2014/90/EU, (EU) 2016/797 und (EU) 2020/1828 (Verordnung über künstliche Intelligenz) (Text von Bedeutung für den EWR).

Eintretens und 2) der Abhängigkeit der betroffenen Personen platziert, d. h. wie verletzlich diese sind und inwieweit sie die Möglichkeit haben, KI-Entscheidungen zu beeinflussen und/oder zu kritisieren bzw. anzufechten. Die Matrix ermöglicht so eine multidimensionale und anwendungsspezifische Anforderungsanalyse. Zur Risiko-Einordnung der Systemfähigkeiten der einzelnen VIKING-Anwendungen und als „Proof of Concept“ wurde für alle Anwendungsfelder (AP 4-7) eine Risikoeinschätzung anhand der Matrix vorgenommen (siehe Beispiel Gesichtserkennung, Abbildung 2).

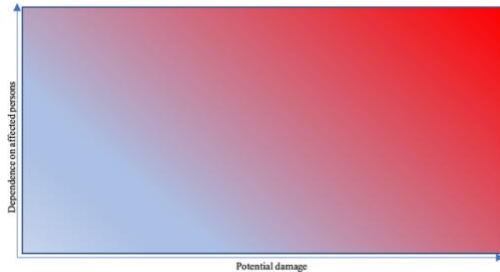


Abbildung 1: Visualisierung der Risikomatrix (Template)

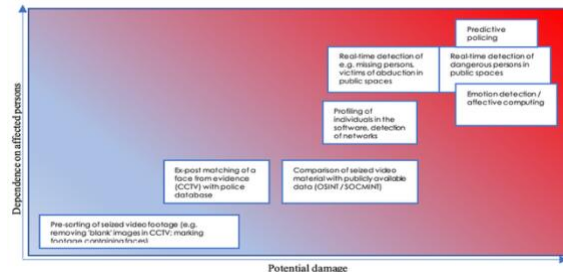


Abbildung 2: Risikovisualisierung für AP 4: Gesichtserkennung

## AP 2.2: Transparenz/Erklärbarkeit

### Ziele:

(3) Analyse und Spezifizierung der ethischen Anforderungen an Transparenz

### Ergebnisse:

In AP 2.2 erfolgt eine Analyse der kontextspezifischen Anforderungen an die Transparenz und Erklärbarkeit von KI-Systemen im polizeilichen Ermittlungskontext. Es wurde eine Literaturrecherche zu verschiedenen Ansätzen der Transparenzschaffung in der KI- und Software-Entwicklung und deren Ethik-Komponenten durchgeführt. Dies beinhaltete unter anderem Transparenzinstrumente wie Data Sheets, Model Cards und System-Folgenabschätzungen. Erkenntnisse aus diesen Recherchen sowie aus Diskussionen mit den Projektpartner:innen wurden in den ethischen Anforderungskatalog und das vom Gesamtverbund entwickelte Spezifikationsdokument integriert (AP 1). Dabei wurde betont, dass Ansätze erklärbarer KI (XAI) Möglichkeiten aufzeigen können, um eine höhere Transparenz in KI-Systemen herzustellen und problematische Konsequenzen der Opazität zu vermindern. Gleichzeitig werden zentrale XAI-Begriffe wie Interpretierbarkeit und Vollständigkeit oft zu vereinfacht gesehen (Rohlfing et al. 2021). Insgesamt werden „Erklärungen“ dabei als einfache Informationsübermittlung verstanden, während es sich realistischere um einen sozialen Prozess der Vermittlung von Wissen handelt.

Aufbauend auf diesen Erkenntnissen richtete das IZEW am Meilenstein-Treffen (M18) gemeinsam mit den Partner:innen der HWR und UKON einen Workshop zu zielgruppen-orientierter Transparenz für das Konsortium aus. Dabei wurden Informationsbedürfnisse und -strategien für unterschiedliche Stakeholder identifiziert zur Optimierung der Handlungsoptionen dieser Akteure (siehe AP 2.3). Es wurde weiterhin eine System Card entwickelt zur Erhöhung der Transparenz und Accountability (siehe AP 2.3) von polizeilicher KI-Software (Table 1). Diese System Card wurde mit den Projektpartner:innen geteilt und bezüglich der VIKING-Anwendungen diskutiert. Während der anwendungsneutralen Verlängerung des Projekts wurden die Items der System Card mit dem Projektpartner UKON in mehreren bilateralen Meetings diskutiert und mit dem von UKON entwickelten VIKING-Transparenztool abgeglichen; wo passend, wurden Aspekte der System Card an das Transparenztool angepasst und in dieses integriert.

**General information**

System name	
Description	
Links to further accessible information	

**Purpose**

Motivation	
Intended use	
Prohibited use	
Unsupported use	

**System features**

Existing features	
Planned features	

**Geographic areas and languages**

Existing target region	
Existing target region	

**Developer(s)**

Main developer	
Others	

**Data**

Documentation of data in use	Data impact assessment (DIA) (representativeness, relevance, accuracy, traceability, completeness, others)	Technical mitigation measures for risks detected through the DIA (missing, data, normalisation, scaling, others)	Legal compliance (GDPR; privacy, consent, protected attributes)
Results			

**Stakeholder analysis**

Stakeholders	Potential benefits	Potential harms
1.		
2.		
3.		
4.		
5.		
6.		
7.		
8.		
9.		

**Fairness**

Conducted fairness evaluation(s) (metric, procedure, benchmark)	
Result	

### Transparency

Mechanisms and documentation to evaluate transparency	Traceability mechanism(s) for data quality, decisions/recommendations made	Explainability: <ul style="list-style-type: none"> <li>• Users receive explanation</li> <li>• Explanation is continuously monitored for users' understanding</li> </ul>	Communication: <p>Users are informed of purpose, criteria and limitations (benefits, technical limitations and potential risks, training material and disclaimers are available)</p>
Results			

### Accountability

Mechanisms to ensure responsibility for development, deployment and/or use of AI system	Auditability: <ul style="list-style-type: none"> <li>• Mechanisms to facility the system's auditability</li> <li>• Possibility of being audited by external/third-party/independent actor</li> </ul>	Risk assessment: <p>Documentation of trade-offs and the reasons behind the decisions made</p>	Report of potential vulnerabilities, bias or other problems (bugs)	Possibility of redress for those negatively affected
Results				

Table 3: System Card / Operation Map

## AP 2.3: Menschliche Letztverantwortung und Generalisierbarkeit

### Ziele:

- (4) Ethische Evaluation von Accountability-Mechanismen
- (5) Operationalisierung von sozio-technischen Anforderungen ethischer Aspekte im Spannungsfeld von Generalisierung und Kontext

### Ergebnisse:

Es wurde eine Literaturrecherche zu verschiedenen Ansätzen der menschlichen Letztverantwortung in der KI- und Software-Entwicklung und deren Ethik-Komponenten durchgeführt (Accountability, menschliche Aufsicht und Contestability). KI-gestützte Technologien bringen aufgrund ihrer Komplexität, ihres dynamischen Lernpotenzials sowie der großen Anzahl der typischerweise beteiligten Stakeholder Bedenken hinsichtlich der Accountability (Rechenschaftspflicht) mit sich. In soziotechnischen Umgebungen bedeutet Accountability nicht notwendigerweise eine kausale Verantwortung von Akteuren im Sinne von Handlungsfähigkeit, sondern typischerweise die Bereitschaft oder Verpflichtung, Verantwortung zu übernehmen (Saurwein 2018, S. 38). Accountability bezieht sich vor allem auf die prospektive Seite der Verantwortung anderen Akteuren gegenüber, also die institutionellen wie individuellen Vorkehrungen, um ethisch unerwünschte Auswirkungen zu vermeiden und so potenziellen Schaden zu minimieren. Accountability-Anforderungen beinhalten sowohl die Spezifizierung der Pflichten als auch die Benennung des verantwortlichen Akteurs. Gerade in autonomen KI-basierten Systemen mit unklaren algorithmischen Selektions- und Entscheidungsstrukturen ist es entscheidend, einer möglichen sogenannten Verantwortungsdiffusion und Verantwortungsfucht entgegenzuwirken. Perspektivisch müssen verantwortliche Personen oder Institutionen benannt werden, die für die Erfüllung der mit dem KI-System verbundenen vordefinierten Transparenz-, Überprüfbarkeits- und Sorgfaltspflichten verantwortlich sind. Als rechenschaftspflichtige Akteure müssen diese Stellen auch für alle Betroffenen identifizierbar und leicht zugänglich sein, z.B. für Fragen, Beschwerden oder Einsprüche (AIEIG 2020), was die Contestability von KI-Systemen erhöht.

Eine weitere wichtige Anforderung des AI Acts an alle Bereitsteller und Anbieter von Hochrisikosystemen ist die Pflicht zur menschlichen Aufsicht. Das bedeutet, dass Hochrisikosysteme so konzipiert und entwickelt werden, dass natürliche Personen ihre Funktionsweise überwachen können. Sie müssen sicherstellen können, dass die Systeme bestimmungsgemäß verwendet werden und dass ihre Auswirkungen während des gesamten KI-Lebenszyklus berücksichtigt werden. KI-gestützte polizeiliche Instrumente werden im Allgemeinen als „Assistenzsysteme“ bezeichnet. Ohne kompetente menschliche Aufsicht kann die Komplexität und wahrgenommene Autorität von KI-gestützten Prozessen und Outputs jedoch dazu führen, dass Biases unerkannt bleiben oder von Anwender:innen von Korrelationen auf Kausalitäten geschlossen wird. Es kann zu einer Autoritätsumkehr kommen zwischen Anwender:innen und Systemen, die die menschliche Entscheidungsfindung eigentlich nur unterstützen sollen (Yeung et al. 2021). Es sollte beim Design also stets darauf geachtet werden, dass dieses Verhältnis zwischen dem maschinellen Assistenten und dem Souverän sich auch in der Realität der Techniknutzung widerspiegelt und Polizeibeamt:innen nicht zu bloßen Assistent:innen maschineller Entscheidungen werden. Das Prinzip der menschlichen Aufsicht ist eng mit Transparenzanforderungen verbunden: Anwender:innen müssen dazu befähigt werden, die Vertrauenswürdigkeit von Modellen, Daten und anderen Elementen kritisch überprüfen zu können. Ebenso hängt das Prinzip der menschlichen Aufsicht eng mit Accountability zusammen. Die Verantwortung für menschliche Aufsicht kann nicht bei Individuen liegen, sondern muss im gesamten Lebenszyklus der Technologie, von Design und Entwicklung bis zum Einsatz, mitgedacht werden.

Aus AP 2.3 sind zwei projektinterne Workshops hervorgegangen: Auf dem Projekttreffen im Januar 2024 fand ein Workshop zur Thematik Accountability statt, an dessen Vorbereitung und Durchführung die Projektpartner:innen der HWR beteiligt waren. Ebenso mit Unterstützung des HWR-Teams fand auf dem Projekttreffen im Dezember 2024 der letzte Ethik-Workshop zur Thematik menschliche Aufsicht statt. Erkenntnisse aus AP 2.3 flossen weiterhin in ein vom IZEW organisiertes Panel namens *“Enacting contestation of Artificial Intelligence (AI) – concepts, approaches and techniques”* ein, das auf der EASST-4S-Konferenz 2024 in Amsterdam stattfand.

Aufbauend auf den Literaturrecherchen, Erkenntnissen aus den projektinternen Workshops sowie auf Analysen des mittlerweile finalisierten AI Acts wurden während der Verlängerung von VIKING Anforderungen zu Accountability und menschlicher Aufsicht in den ethischen Anforderungskatalog integriert. Weiterhin wurden, wie bereits in den Ausführungen zu AP 2.2 genannt, die Aspekte bezüglich Accountability und menschlicher Aufsicht aus System Card in das interaktive UKON-Transparenztool eingearbeitet.

### **AP 3: Rechtliche Anforderungen**

#### Verwendung der Zuwendung:

Die Zuwendung wurde überwiegend für folgende Positionen verwendet:

- Beschäftigungsentgelte für wissenschaftliche Mitarbeitende
- Studentische Hilfskräfte (Literaturrecherche)

#### **AP 3.1: Verfahren zur rechtlichen Bewertung von KI-Systemen**

#### **AP 3.2: KI-Anwendungen und Grundrechte**

#### **AP 3.4: Polizeiliche KI-Anwendungen und Datenschutz**

### Ziele:

Abgleich ethischer Anforderungen mit rechtlichen Minimalstandards

### Ergebnisse:

Mit dem Rechtspartner HWR wurde von Anfang an eine enge Kollaboration aufgebaut, um potenzielle Synergien zu identifizieren und um die technischen Partner:innen zu begleiten und in ethischer und rechtlicher Hinsicht zu beraten. Es fanden eine Reihe von Treffen zwischen IZEW, HWR und den technischen Partner:innen für jedes AP statt, um Anforderungen zu konkretisieren und mögliche ethische und/oder rechtliche Konflikte zu identifizieren. Für eine intern kongruente Analyse normativer Anforderungen und Handlungsempfehlungen im Einklang mit den entsprechenden Anforderungen aus dem AI Act wurden die ethischen Aspekte (AP 2) mit den rechtlichen Kriterien für die Bewertung und Standardisierung (DIN) polizeilicher KI-Anwendungen abgeglichen. Existierende Fairness-Maßnahmen wurden kritisch in Hinblick auf das Ziel der Vermeidung von Diskriminierung überprüft.

In Zusammenarbeit mit der HWR wurde auf der Rechtssoziologie-Konferenz “Zugang zum Recht – zugängliche Rechte” in Innsbruck (September 2023) ein Panel namens “Künstliche Intelligenz und staatliche Institutionen der (Un)Sicherheit” durchgeführt, in dem diverse Aspekte aus den Arbeiten zu VIKING aufgegriffen und diskutiert wurden. Gemeinsame Forschungsergebnisse der VIKING-Wissenschaftler:innen vom IZEW und von der HWR zum Thema „*Fairness, Erklärbarkeit und Transparenz bei KI-Anwendungen im Sicherheitsbereich - ein unmögliches Unterfangen?*“ wurden in der Zeitschrift „vorgänge, Zeitschrift für Bürgerrechte und Gesellschaftspolitik“ veröffentlicht.

In der zweiten Projekthälfte wurde die enge Zusammenarbeit mit der HWR fortgesetzt. Neben den Schwerpunkten Fairness, Erklärbarkeit und Transparenz wurde die Kooperation auf die Themen Verantwortung, Accountability und menschliche Aufsicht ausgeweitet, um mögliche ethische und/oder rechtliche Konflikte mit AP-Entwicklungen zu identifizieren (siehe auch AP 2.3). Auf dem vom IZEW organisierten Panel “*Enacting contestation of Artificial Intelligence (AI) – concepts, approaches and techniques*” auf der EASST-4S-Konferenz in Amsterdam (siehe AP 2.3) wurde der Vortrag “*Facets of (in-)contestability: the case of AI-powered police intelligence applications*” von Simon Hirsbrunner mit den HWR-Partnern Milan Tahraoui und Steven Kleemann präsentiert.

## **AP 4: Gesichtserkennung**

### Verwendung der Zuwendung:

Die Zuwendung wurde überwiegend für folgende Positionen verwendet:

- Beschäftigungsentgelte für wissenschaftliche Mitarbeitende

### **AP 4.1: Bereitstellung geclusterter Daten**

### **AP 4.2: De-Biasing im Trainingsprozess**

### **AP 4.5: Demonstrator**

### Ziele:

Das Ziel des AP4 war die Entzerrung (De-Biasing) biometrischer Gesichtserkennung bzgl. Alter, Geschlecht und Ethnizität in Abstimmung mit der Ethik-Matrix aus AP 2.

### Ergebnisse:

Das IZEW nahm regelmäßig an den Arbeitstreffen von AP 4 teil, um eine genauere Vorstellung von

den technischen Funktionsweisen zu erhalten und zielgerichtet ethische Einschätzungen liefern zu können. Schwerpunkte lagen hier in der Problematisierung der verwendeten Trainingsdaten (Gesichterfotos aus dem Internet) und der Geschlechterbinarität von Gesichtserkennungs-Software sowie die Berücksichtigung holistischer Konzepte zu Diskriminierung, Fairness und Diversität. Die Dichotomie (männlich/weiblich) der Klasse Geschlecht wurde problematisiert, da hier von einem reduktionistischen Geschlechterbild ausgegangen wird, welches nicht immer der geschlechterbezogenen Selbstwahrnehmung betroffener Menschen entspricht (Hamidi et al. 2018).

Ein besonderes Augenmerk lag weiterhin auf der Frage, wie nicht nur mit *False Positives*, sondern auch mit *False Negatives* umgegangen werden kann. False Positives sind für den Polizeibereich sehr relevant, da eine Verdächtigung und Strafverfolgung, selbst wenn sie sich im Nachhinein als fälschlich erweist, nachhaltig destruktive Folgen für die Lebenschancen einer Person und ihr direktes Umfeld nach sich ziehen kann. Doch auch falsch-negative Ergebnisse müssen möglichst verhindert werden; beispielsweise kann bei der Suche nach einem vermissten Kind eine falsch-negative Ergebnis dazu führen, dass das vermisste Kind fälschlicherweise nicht identifiziert wird und die Suche durch den Fehler des Systems verlängert wird oder sogar erfolglos bleibt. Überdies können problematische Biases durch übermäßige Kontrollen bestimmter Personengruppen entstehen. Es wurde insbesondere eine umfassende Technologiebewertung mit dem Fokus auf die Vermeidung von algorithmischer Diskriminierung und die Herstellung von KI-Fairness durchgeführt. Ein daraus resultierender wissenschaftlicher Artikel, „*Algorithmische Fairness in der polizeilichen Ermittlungsarbeit: Ethische Analyse von Verfahren des maschinellen Lernens zur Gesichtserkennung*“ wurde mit den AP4-Partnern diskutiert und im März 2023 in der Zeitschrift „TATuP: Zeitschrift für Technikfolgenabschätzung in Theorie und Praxis“ veröffentlicht (Brandner, L. T., & Hirsbrunner, S. D. (2023).

Die identifizierten ethischen Aspekte und Risiken für die Gesichtserkennung flossen in den ethischen Anforderungskatalog, die feingranulare Version der Matrix sowie in die vom Konsortium erarbeitete DIN SPEC (AP 8) ein.

## **AP 5: Textauswertung**

### Verwendung der Zuwendung:

Die Zuwendung wurde überwiegend für folgende Positionen verwendet:

- Beschäftigungsentgelte für wissenschaftliche Mitarbeitende

### **AP 5.1: Erstellung von Benchmark-Datensätzen und Modell-Training**

### **AP 5.2: Entwicklung von De-Biasing-Methoden für Sprachmodelle**

#### Ziele:

Ziel von AP 5 war die Erforschung vertrauenswürdiger KI-Sprachmodelle zur Textklassifikation für den Einsatz bei Sicherheitsbehörden. Ziel des ethischen Teilvorhabens in diesem AP war die Begleitung der Ergebnisse der technischen Arbeiten und der Abgleich dieser mit ethischen Mindestanforderungen und Best Practices.

#### Ergebnisse:

Das IZEW nahm gemeinsam mit der HWR an mehreren Online-Treffen der AP5-Partner:innen teil und brachte ethische Erwägungen für die Technikentwicklung ein. Während sich die gewählten Objektklassen zur Named-Entity-Recognition als ethisch unproblematisch erwiesen, waren

Trainingsdaten eine wichtige Problematik. Ethisch umstritten ist z.B. Ursprung der Trainingsdaten aus vortrainierten Sprachmodellen (bspw. alle verfügbaren Daten aus dem Netz) aufgrund von Privatheits- und Fairness-Bedenken. Es können z.B. Leistungsunterschiede bei verschiedenen Sprachen und diskriminierende Verzerrungen vorliegen. Den Partnern wurden daher Daten aus ethisch und datenschutzrechtlich unbedenklichen Quellen nahegelegt. Die verwendeten Trainingsdaten sollten die erwarteten Sprachvarianten der Zielpopulation darstellen und die Datensatzmerkmale sollten in späteren Phasen erneut darauf überprüft werden, wie gut sie mit den zuvor spezifizierten erwarteten Anwendungsfällen übereinstimmen.

Ein weiteres Augenmerk der IZEW-Forschung lag auf dem Problem der Kontextgebundenheit von Sprache in Bezug auf ihre Semantik und Sinnhaftigkeit oder anderen Faktoren wie Ironie und Sarkasmus, welche eine automatische Analyse erschweren und zu in ethischer Weise folgenreichen Fehlschlüssen führen können. Menschliche Aufsicht spielt hier deshalb eine zentrale Rolle, um Ergebnisse ggf. von geschulten Aufsichtspersonen kritisch gegenprüfen zu lassen. Ein wichtiger Punkt für die Implementierung menschlicher Aufsicht in AP 5 war die Eignung der Benutzerführung und der *Explanations*/Erklärungen im Demonstrator für Nicht-Expert:innen. Es wurden Methoden wie Interviews, Tests mit einem Mockup oder andere Methoden der Mensch-Computer-Interaktion bzw. des Human-Centered Machine Learnings empfohlen, um die Eignung der Benutzerführung auch im Rahmen einer formalen Evaluation zu überprüfen.

Die identifizierten ethischen Aspekte und Risiken für die Textauswertung sind in den ethischen Anforderungskatalog, die feingranulare Version der Matrix sowie in die vom Konsortium erarbeitete DIN SPEC (AP 8) eingeflossen.

## **AP 6: Sprechererkennung**

### Verwendung der Zuwendung:

Die Zuwendung wurde überwiegend für folgende Positionen verwendet:

- Beschäftigungsentgelte für wissenschaftliche Mitarbeitende

### **AP 6.1: Analyse eines bestehenden Sprachkorpus**

### **AP 6.2: Aufbau eines Sprachkorpus**

### **AP 6.3: Analyse von Netzwerkstrukturen**

### Ziele:

Das Ziel von AP 6 war die Formulierung ethischer Anforderungen für die Entwicklung erklärbarer und vertrauenswürdiger neuronaler Netze zur automatischen Sprechererkennung bzw. zur Bewertung bestehender Black-Box-Systeme. Begleitend erfolgte die Anpassung der ethischen Matrix an den Kontext der Sprechererkennung.

### Ergebnisse:

Das IZEW nahm gemeinsam mit der HWR an mehreren Online-Treffen der AP6-Partner:innen teil und brachte ethische Erwägungen für die Technikentwicklung ein. Auch bei der Sprechererkennung lag ein Fokus auf der Vermeidung von Verzerrungen, die zur Diskriminierung vulnerabler Gruppen und Minderheiten beitragen und potenziell bestehende Vorurteile verstärken können. So sind die bereits für die Textauswertung genannten möglichen Leistungsunterschiede bei verschiedenen Sprachen (z.B. Englisch gegenüber Deutsch) ebenso für die Sprechererkennung relevant, hinzu kommen

Leistungsunterschiede bei Dialekten, Akzenten, sowie bei Stimmen und Sprachmustern (z.B. männlich gegenüber weiblich). Mit Blick auf die Robustheit des Systems war weiterhin die nötige Akkuratheit unter suboptimalen Umständen (z.B. schlechte Soundqualität, Störgeräusche, Akzente, Sprachstörungen) ein Schwerpunkt, um die Vermeidung ethisch unverantwortlicher Effekte sicherzustellen. Auch die bereits im Abschnitt zur Gesichtserkennung angesprochene Dichotomie von weiblich und männlich wurde hinsichtlich der Sprechererkennung problematisiert, da in den entsprechenden Verfahren die Selbstwahrnehmung von Individuen gegebenenfalls nicht berücksichtigt wird.

Die identifizierten ethischen Aspekte und Risiken für die Sprechererkennung sind in den ethischen Anforderungskatalog, die feingranulare Version der Matrix sowie in die vom Konsortium erarbeitete DIN SPEC (AP 8) eingeflossen.

## **AP 7: Objektdetektion**

### Verwendung der Zuwendung:

Die Zuwendung wurde überwiegend für folgende Positionen verwendet:

- Beschäftigungsentgelte für wissenschaftliche Mitarbeitende

### **AP 7.1: Erstellung von Test- und Trainingsdatensätzen**

### **AP 7.5: Evaluation und Bewertung der Ergebnisse**

#### Ziele:

Das Ziel von AP7 war die ethische Begleitung und Bewertung des zu entwickelnden KI-Verfahrens zur Detektion von Objekten in Bildmaterial. Ein besonderes Augenmerk bei der ethischen Begleitung galt überdies der Spezifikation ethischer Anforderung für die geplante Entwicklung der Methode, welche Fehler erklären kann und dadurch das gezielte Nachtrainieren des Detektors (Weiterlernen) ermöglichen sollte.

#### Ergebnisse:

Das IZEW nahm an mehreren Online-Treffen der AP7-Partner:innen teil und brachte ethische Erwägungen für die Technikentwicklung ein, z.B. darüber wie Perspektiven und Bedarfe der Endnutzer:innen und Erklärungen/erklärbare Ergebnisse in die Webschnittstelle des Demonstrators integriert werden konnte. Ein besonderer Schwerpunkt lag auf der Selektion der Anwendungsfelder. Im Gegensatz zu anderen technischen APs waren in AP 7 Anwendungsfälle ohne Bezug zu biometrischen Daten vorhanden und praxisrelevant. Aus Gründen des ethischen und (datenschutz-)rechtlichen Risikomanagements wurde betont, dass möglichst keine Anwendungsfälle gewählt werden sollten, in welchen biometrische Daten eine tragende Rolle spielten. Potenzielle Szenarien wie die Objekterkennung in Videoaufnahmen mit Personenbezug wurden problematisiert, da für sie die Verarbeitung sensibler biometrischer Daten im Vordergrund stand und wenig technische Zuverlässigkeit zu erwarten war. Daraufhin wurden andere Anwendungsfelder (Schmuckstücke, Waffen) gewählt, die aus ethischer Sicht viel unproblematischer zu bewerten waren.

Die identifizierten ethischen Aspekte und Risiken für die Objekterkennung sind in den ethischen Anforderungskatalog, die feingranulare Version der Matrix sowie in die vom Konsortium erarbeitete DIN SPEC (AP 8) eingeflossen.

## AP 8: Methodenspektrum, Prüfkataloge und Standardisierung

### Verwendung der Zuwendung:

Die Zuwendung wurde überwiegend für folgende Positionen verwendet:

- Beschäftigungsentgelte für wissenschaftliche Mitarbeitende
- Studentische Hilfskräfte (Korrekturlesen)

### **AP 8.1: Übertragbarkeit zwischen Anwendungen**

### **AP 8.2: Gesamtheitliches Framework**

### **AP 8.3: Standardisierung**

### **AP 8.4: Evaluierung der Demonstratoren**

### Ziele:

AP 8 befasste sich mit Studien zur Generalisierbarkeit der in AP 4-7 erarbeiteten ethischen Anforderungen für unterschiedliche technische Verfahren auf andere Anwendungen und der ethischen Prüfung eines standardisierten Rahmens für Entwickler:innen und polizeiliche Anwender:innen.

### Ergebnisse:

Die Erkenntnisse und Ergebnisse aus AP 2 wurden im Hinblick auf ihre Relevanz für die Normung und Standardisierung in AP 8.3 mit dem DIN diskutiert und abgestimmt. Das IZEW identifizierte für das DIN aus ethischer Sicht relevante Normen und Standards. PI Wulf Loh arbeitete im DIN / DKE Normungsausschuss NA 043-01-42 GA DIN/DKE Künstliche Intelligenz mit. Weiterhin beteiligte sich das IZEW maßgeblich an der Entstehung der DIN SPEC mit dem Titel „*Vertrauenswürdige KI-Methoden in polizeilichen Anwendungen*“. Ziel des Standarddokumentes war die Spezifikation rechtlicher und ethischer Überlegungen sowie technischer Empfehlungen für den vertrauenswürdigen Einsatz von KI-Technologien in Strafverfolgungsanwendungen. Lou Brandner übernahm die stellvertretende Leitung dieses Vorhabens (Leitung: Maximilian Fischer/UKON). IZEW-Mitglieder Wulf Loh und Simon Hirsbrunner wirkten an der Ausarbeitung mit. Ethische Erkenntnisse aus den Arbeiten in allen VIKING-APs, dem ethischen Anforderungskatalog sowie aus der Risikomatrix sind in die Spezifikationen in der DIN SPEC eingeflossen. Die Koordination und Erarbeitung der DIN SPEC wurde während der Verlängerung von VIKING erfolgreich abgeschlossen und das Dokument wurde im Mai 2025 veröffentlicht.

## 2. Wichtigste Positionen des zahlenmäßigen Nachweises

### **Position 0812 (Beschäftigte E12-E15): 218.019,37 €**

Aus Position 0812 wurden die Beschäftigungsentgelte für die vier wissenschaftlichen Angestellten, in wechselnder Besetzung und Prozentzahl, gezahlt.

### **Position 0822 (sonstige Beschäftigungsentgelte): 20.695,92 €**

Aus Position 0822 wurden die Beschäftigungsentgelte für sieben wissenschaftlichen Hilfskräfte, in wechselnder Besetzung und Stundenzahl, gezahlt, welche die Arbeiten innerhalb des Projekts unterstützten.

**Position 0850: 1.117,47 €**

Aus Position 0850 wurde ein für im Projekt stattgefundene Recherchen und Analysen notwendiger Laptop bezahlt:

- Apple MacBook Air 13" (UB#42000286/2022), CANCOM GmbH

**Position 0846: 14.172,38 €**

Aus Position 0846 wurden Reisen zu Projekttreffen sowie Reisen und Anmeldegebühren für Konferenzen und Fachtagungen im Inland und Ausland bezahlt.

**Tabelle 1:** Projekttreffen

Mitarbeitende	Anlass	Ort	Datum
Wulf Loh	Projekttreffen	Konstanz	27.-28.6.2022
Lou Brandner, Simon Hirsbrunner	Projekttreffen	München	18.-19.1.2023
Simon Hirsbrunner, Sol Martinez Demarco	Meilensteintreffen	München	19.-20.6.2023
Sol Martinez Demarco	Projekttreffen	Berlin	16.-18.1.2024
Lou Brandner	Projekttreffen	Konstanz	12.-13.6.2024
Lou Brandner, Theresa Krampe	Projekttreffen	Oldenburg	9.-11.12.2024

**Tabelle 2:** Konferenzen und Fachtagungen Inland

Mitarbeitende	Anlass	Ort	Datum
Simon Hirsbrunner	Ringvorlesung "Data Literacy", KIT	Karlsruhe	6.-8.3.2023
Wulf Loh	Deutscher Kongress für Philosophie	Münster	22.-27.9.2024

**Tabelle 3:** Konferenzen und Fachtagungen Ausland

Mitarbeitende	Anlass	Ort	Datum
Sol Martinez Demarco	CSCW Summer School	Como, Italien	20.-26.8.2023
Simon Hirsbrunner	CSCW-Konferenz	Minneapolis, USA	13.-19.10.2023
Sol Martinez Demarco	STS-Konferenz	Graz, Österreich	5.-9.5.2024
Lou Brandner	SSN-Konferenz	Ljubljana, Slowenien	27.5.-1.6.2024
Wulf Loh	EASST-Konferenz	Amsterdam, Niederlande	13.-19.7.2024
Wulf Loh	World Congress of Philosophy	Rom, Italien	1.-9.8.2024
Lou Brandner	ECREA-Konferenz	Ljubljana, Slowenien	24.-28.9.2024

### 3. Notwendigkeit und Angemessenheit der geleisteten Projektarbeiten

Die gegenwärtige rasante Entwicklung und Verbreitung von KI-Technologien, die auch den Bereich der zivilen Sicherheit, Strafverfolgung und -prävention betrifft, macht die ethische Evaluation von gesellschaftlichen Folgen derartiger Systeme unerlässlich. Ethische Forschungsarbeiten bergen immer das Risiko, dass sie in praktischen Kontexten auf Problematiken verweisen, die ihrerseits jedoch auf faktische Akzeptanz treffen. Somit können teils inkompatible Ansätze normativer Bewertungen entstehen, die eine Metaebene der Reflexion nötig machen, welche schließlich zur Schaffung von ethischer Akzeptabilität beitragen.

Erst die Förderung des Forschungsvorhabens ermöglichte eine solche Reflexion auf der Metaebene, die gesellschaftliche Akzeptanz nicht als hinreichendes Kriterium hinnimmt, sondern diese mit ethischer Akzeptabilität kontrastiert. So konnten nicht nur gewünschte, sondern ebenso wünschenswerte Ergebnisse erarbeitet werden. Die unterschiedlichen theoretischen Arbeiten innerhalb des Projekts machten zudem die Arbeit in einem breit aufgestellten Forschungsverbund nötig. Das Projekt VIKING im Allgemeinen und die Arbeiten des Teilvorhabens im Besonderen waren also auf die Förderung angewiesen. Es bestand durch den inter- sowie transdisziplinären Charakter des Projekts und wegen des hohen Praxisbezugs keine Alternative zu dieser Förderung durch andere Institutionen.

#### **4. Voraussichtliche Nutzen, insbesondere die Verwertbarkeit des Ergebnisses im Sinne des fortgeschriebenen Verwertungsplans**

Aufgrund der Alleinstellungsmerkmale des Projektfokus sind die wissenschaftlichen Ergebnisse des Projekts innovativ, passen sich jedoch nahtlos in die bestehende Forschung ein und profitieren gleichzeitig von dieser. Hier kann das IZEW mit den Projektergebnissen den wissenschaftlichen Austausch und Fortschritt auf dem Gebiet der KI-Ethik, der Operationalisierung von KI-ethischen Prinzipien und Werten, sowie der Technologienutzung in der zivilen Sicherheit allgemein unterstützen und voranbringen. Durch die veröffentlichten Artikel – und weitere, folgende Publikationen, die geplant oder bereits unter Begutachtung sind, siehe Kapitel 6 „Erfolgte und geplante Veröffentlichungen der Ergebnisse“ – sind die Erkenntnisse für die Fachwelt aufgearbeitet und stehen zu weiterführenden Forschungen zur Verfügung.

Auch weiterhin werden IZEW-Mitarbeitende VIKING-Ergebnisse in Vorträgen, wissenschaftlichen Artikeln und Workshops zu ethischen Aspekten der polizeilichen KI-Nutzung verwerten. Die Ergebnisse des Teilvorhabens haben Eingang in die Unternehmen und Forschungsinstitute der Projektpartner:innen gefunden und das erlangte Verständnis über Fragen der Daten- und KI-Ethik kann über VIKING hinaus zu einer institutionellen Auseinandersetzung mit ethischen Aspekten beitragen. Die Projektergebnisse tragen insgesamt dazu bei, das IZEW als Standort für Technikethik – und insbesondere KI-Ethik – zu stärken.

Die Erkenntnisse aus VIKING sollen zudem in weitere Projekte zur Entwicklung und Nutzung von KI-Technologien im Kontext von Polizei und Sicherheit einfließen. Konkret geschieht dies bereits in der KI-Allianz BW (Ministerium für Wirtschaft, Arbeit und Tourismus, Baden-Württemberg), in der KI-Aktivitäten aus Wirtschaft und Wissenschaft verknüpft werden. Die Projektergebnisse können auch in anderen Operationalisierungsinitiativen, an denen das IZEW beteiligt ist, verwendet werden und geben nicht zuletzt Hinweise und Handreichungen für die Arbeit des IZEW in KI-Innovationsgremien (Plattform Lernende Systeme, Cyber Valley). Das IZEW nutzt die Ergebnisse zu Vernetzungsaktivitäten im Bereich KI und ziviler Sicherheit, um hier in interdisziplinärer Zusammenarbeit die ethischen Anforderungen an KI-Systeme weiterzuentwickeln.

## 5. Während der Durchführung des Vorhabens dem Zuwendungsempfänger bekannt gewordenen Fortschritt auf dem Gebiet des Vorhabens bei anderen Stellen

Das Themenfeld Künstliche Intelligenz hat gegenwärtig Konjunktur und eine Vielzahl an Projekten zu KI-Themen, auch im Sicherheitsbereich, werden gefördert. So förderte das BMBF/BMFTR im Projektzeitraum von VIKING auch die Projekte KISTRA (Einsatz von KI zur Früherkennung von Straftaten), PEGASUS (Polizeiliche Gewinnung und Analyse heterogener Massendaten zur Bekämpfung organisierter Kriminalitätsstrukturen) und FAKE-ID (Videoanalyse mit Hilfe künstlicher Intelligenz zur Detektion von falschen und manipulierten Identitäten). Da das IZEW auch an PEGASUS beteiligt war, konnte bis zum Ende von PEGASUS (2023) ein konstanter Austausch zwischen den beiden Vorhaben stattfinden. Die entstandenen Synergien waren dem Fortschritt von VIKING in hohem Maße zuträglich. Das Projekt KISTRA untersuchte auf KI basierende Textanalytik-Modelle zur Bewertung von Hasskriminalität. Hierbei lag der Forschungsfokus auf der Generalisierbarkeit der Modellergebnisse, nicht auf deren Erklärbarkeit. Die untersuchte Angreifbarkeit von Bildmodellen in KISTRA wird durch die Forschung zur Angreifbarkeit von Textmodellen in VIKING ergänzt. Insofern baute das Vorhaben VIKING auch auf Erkenntnissen aus KISTRA auf, bearbeitete aber neue Fragestellungen und Forschungsschwerpunkte. Da zudem VIKING-Mitarbeitende aus dem Team der HWR auch in FAKE-ID arbeiteten, konnten hier ebenso Synergien geschaffen werden und Erkenntnisse aus FAKE-ID waren dem Fortschritt von VIKING zuträglich.

Auf EU-Ebene wurden z.B. die Projekte AIDA (Artificial Intelligence and Advanced Data Analytics for Law Enforcement Agencies) und ALIGNER (Artificial Intelligence Roadmap for Policing and Law Enforcement) gefördert. Die Ergebnisse anderer Projekte wurden regelmäßig hinsichtlich eines möglichen Beitrags für die Projektarbeiten in VIKING überprüft. Jedoch zeigte sich, dass die Schwerpunktlegung auf verschiedene Erkennungs- und Analysetechnologien in der polizeilichen Ermittlungsarbeit sowie auf die konkrete Operationalisierung von Vertrauenswürdigkeit in diesem Kontext ein Alleinstellungsmerkmal des VIKING-Verbunds darstellte.

## 6. Erfolgte und geplante Veröffentlichungen der Ergebnisse

### Publikationen:

- Lou Brandner und Simon Hirsbrunner (2023): Algorithmische Fairness in der polizeilichen Ermittlungsarbeit: Ethische Analyse von Verfahren des maschinellen Lernens zur Gesichtserkennung. TATuP - Zeitschrift für Technikfolgenabschätzung in Theorie und Praxis 32(1), 24-29.
- Steven Kleemann, Simon Hirsbrunner und Hartmut Aden (2023): Fairness, Erklärbarkeit und Transparenz bei KI-Anwendungen im Sicherheitsbereich – ein unmögliches Unterfangen?, vorgänge. Zeitschrift für Bürgerrechte und Gesellschaftspolitik Nr. 242 [62(2)], S. 29-47.
- DIN SPEC 91517 (VIKING-Konsortium, Mitarbeit Lou Brandner, Simon Hirsbrunner, Wulf Loh) (2025): Anforderungen an vertrauenswürdige KI-Methoden in polizeilichen Anwendungen. DIN Deutsches Institut für Normung. <https://dx.doi.org/10.31030/3612025>

### Im Erscheinen:

- Anna Louban, Lou Brandner und Sabrina Schönrock (2025): KI in der Polizeiarbeit: Menschliche Aufsicht als ethische und institutionelle Herausforderung. In: Schönrock, S. & Geißler, S. (Hrsg.): Breitscheidplatz-Symposium 2024: Zukunftssicherheit: Die Rolle von KI im Kampf gegen den Terrorismus.
- Sol Martinez Demarco, Milan Tahraoui und Steven Kleemann (2025): Isolated in the service of society? The siloed logic and the implementation of the Artificial Intelligence Act in the law enforcement context: legal and ethical analysis of the applicability of accountability and responsibility to high-risk AI systems, in: Marie Eneman, Diana Miranda, William Webster und Jan Canbäck (Hrsg.): AI and Surveillance in Policing and Law and Order: Opportunities, Threats, Perspectives and Cases, Routledge.

### Eingereicht:

- Simon Hirsbrunner, Steven Kleemann und Milan Tahraoui (2025): AI contestation as a practice: from a system-centered, lifecycle-focused and developer-oriented perspective of contestability towards normative contextualization, critical practices and organizational cultures. *Frontiers in Communication*
- Lou Brandner, Theresa Krampe und Simon Hirsbrunner (2025): Policing AI: Towards an ethical framework for AI risk assessment in criminal investigation contexts. CEPE2025

### Konferenzen und Workshops auf denen Fragestellungen und Ergebnisse von VIKING von Mitarbeitenden des IZEW präsentiert wurden:

- **2022:**
  - Wulf Loh: *Trustworthy, and fair? Ethical challenges for AI systems*. Esprit Futur, Episode 2: La fabrique du futur, Universität Aix-Marseille (online)
  - Simon Hirsbrunner: *KI-Ethik und Datenkompetenzen im Kontext digitaler Polizeiarbeit*, Eröffnung der Ringvorlesung ‚Data Literacy‘ am Karlsruher Institute für Technologie (KIT)
- **2023:**
  - Simon Hirsbrunner: Panel *Platform (In)Justice*, Workshops *Supporting User Engagement in Testing, Auditing, and Contesting AI* und *Understanding and Mitigating Cognitive Biases in Human-AI Collaboration*, CSCW Konferenz, Minneapolis, USA
  - Simon Hirsbrunner: Seminar zu *Responsible Data Science* am Hasso-Plattner-Institut in Potsdam, bei dem der Fall KI-basierter Ermittlungsmethoden als Fallbeispiel thematisiert wurde. Im Rahmen des Seminars fand auch ein Gastvortrag des HWR-Forschers Milan Tahraoui zum KI-Akt und Konsequenzen für die Polizeiarbeit statt.
  - Wulf Loh: *Operationalisierung, Risikomatrix, und Kritikalitätsstufen*, Ringvorlesung „Ethik der künstlichen Intelligenz“, HLRS Uni Stuttgart
  - Wulf Loh: *Zur Nachhaltigkeit künstlicher Intelligenzen*, Forschungskolloquium Institut für Philosophie, Uni Stuttgart
  - Wulf Loh: *Ethische Überlegungen zum Einsatz von KI-Systemen*, Vorstandssitzung Alphabet, BMW Group, (online)
  - Simon Hirsbrunner, Lou Brandner (mit dem HWR-VIKING-Team): Organisation des Panels *Künstliche Intelligenz und staatliche Institutionen der (Un)Sicherheit*, 5. Kongress der deutschsprachigen Rechtssoziologie-Vereinigungen, Innsbruck,

Österreich.

- **2024:**
  - Sol Martinez Demarco: *Accountability as a relational ethical value*, STS Konferenz, Graz, Österreich
  - Lou Brandner: *Borderline decisions: Is the use of automatic deception detection at EU borders justified?* SSN (Surveillance Studies Network) Konferenz, Ljubljana, Slowenien
  - Wulf Loh: *Empirical Ethics and Critical Theory* und *Contestability and Practices of Critique*, EASST, Amsterdam, Niederlande
  - Sol Martinez Demarco: *Reflecting on being accountable for high-stake decisions*, EASST, Amsterdam, Niederlande
  - Simon Hirsbrunner (mit Milan Tahraoui und Steven Kleemann, HWR): *Facets of (in-)contestability: the case of AI-powered police intelligence applications*, EASST, Amsterdam, Niederlande
  - Wulf Loh: *Transparenz und Fairness von KI-Systemen*, World Congress of Philosophy, Rom, Italien
  - Lou Brandner: Panel *Mobilising methods for post-digital cities*, ECREA-Konferenz, Ljubljana, Slowenien
  - Wulf Loh: *Generative KI und epistemische Ungerechtigkeit*, Deutscher Kongress für Philosophie, Münster
  - Sol Martinez Demarco (mit Steven Kleemann und Milan Tahraoui, HWR Berlin), Vortrag auf dem International Workshop on AI and Surveillance in Policing and Law and Order: Opportunities, Threats, Perspectives and Cases, Universität Göteborg, Schweden.
  - Sol Martinez Demarco (mit Steven Kleemann und Milan Tahraoui, HWR): *The Implementation of the EU AI Act in the Law Enforcement Context: Legal and Ethical Considerations in Terms of Responsibility and Accountability Principles for High-Risk AI Systems*, Eu-SPRI Annual Conference, Twente, Niederlande.
- **2025:**
  - Lou Brandner, Theresa Krampe: *Policing AI: Towards an ethical framework for AI risk assessment in criminal investigation*, CEPE 2025, Rom, Italien (Vortrag angenommen für September 2025)

## 7. Im Bericht verwendete Literatur

1. AI Ethics Impact Group (AIEIG) (2020): From Principles to Practice – An interdisciplinary framework to operationalise AI ethics, VDE/Bertelsmann. <https://www.ai-ethics-impact.org/>
2. Andrejevic, M. (2018): Data Collection without Limits: Automated Policing and the Politics of Framelessness. In: A. Završnik (Hrsg.): Big Data, Crime and Social Control. London: Routledge: 93–107.
3. Angwin, J., J. Larson, S. Mattu & L. Kirchner (2016): Machine Bias. There's software used across the country to predict future criminals. And it's biased against blacks. ProPublica, 23.05.2016. <https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing>
4. Brakel, R. van & de Hert, P. (2011): Policing, surveillance and law in a pre-crime society. Understanding the consequences of technology-based strategies. Journal of Police Studies 20(20): 163–192. [https://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=3316781](https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3316781)
5. Brandner, L. & Hirsbrunner, S. (2023): Algorithmische Fairness in der polizeilichen Ermittlungsarbeit: Ethische Analyse von Verfahren des maschinellen Lernens zur Gesichtserkennung. TATuP - Zeitschrift für Technikfolgenabschätzung in Theorie und Praxis 32(1): 24-29.
6. Brandner, L., T. Krampe & S. Hirsbrunner (forthcoming): Policing AI: Towards an ethical framework for AI risk assessment in criminal investigation contexts. CEPE2025 (eingereicht)
7. DIN & DKE (2020): German Standardisation Roadmap on Artificial Intelligence.
8. DIN & DKE (2022): German Standardisation Roadmap on Artificial Intelligence (2nd edition). <https://www.din.de/en/innovation-and-research/artificial-intelligence/ai-roadmap>
9. DIN SPEC 91517 (2025): Anforderungen an vertrauenswürdige KI-Methoden in polizeilichen Anwendungen. DIN Deutsches Institut für Normung. <https://dx.doi.org/10.31030/3612025>
10. Floridi, L., Holweg, M., Taddeo, M., Amaya Silva, J., Mökander, J. & Wen, Y. (2022): capAI - A Procedure for Conducting Conformity Assessment of AI Systems in Line with the EU Artificial Intelligence Act (SSRN Scholarly Paper No. 4064091). <https://doi.org/10.2139/ssrn.4064091>
11. Hagendorff, T. (2020): The Ethics of AI Ethics. An Evaluation of Guidelines. Minds and Machines 30: 99–120.
12. Hamidi, F., Scheuerman, M. K. & Branham, S. M. 2018. Gender recognition or gender reductionism? The social implications of embedded gender recognition systems. Proceedings of the 2018 CHI conference on human factors in computing systems: 1–13.
13. Helm, P. & T. Hagendorff (2021): Beyond the Prediction Paradigm. Chances and Risks of AI in the fight against Organized Crime. In: L. Floridi, J. Ward & C. Rudin (Hrsg.): Black Box Artificial Intelligence and the Rule of Law. 84 Law and Contemporary Problems: 1-17.
14. Hirsbrunner, S., S. Kleemann & M. Tahraoui (forthcoming): AI contestation as a practice: from a system-centered, lifecycle-focused and developer-oriented perspective of contestability towards normative contextualization, critical practices and organizational cultures. Frontiers in Communication (eingereicht)
15. Institute of Electrical and Electronics Engineers (n/d). IEEE 7000 Standards. <https://ieeexplore.ieee.org/browse/standards/get-program/page/series?id=93>
16. Kleemann, S., S. Hirsbrunner & H. Aden (2023): Fairness, Erklärbarkeit und Transparenz bei KI-Anwendungen im Sicherheitsbereich – ein unmögliches Unterfangen? vorgänge. Zeitschrift für Bürgerrechte und Gesellschaftspolitik 62(2): 29-47.
17. Louban, A. & Lou Brandner (forthcoming): KI in der Polizeiarbeit: Menschliche Aufsicht als ethische und institutionelle Herausforderung. In: Schönrock, S. & Geißler, S. (Hrsg.):

Breitscheidplatz-Symposium 2024: Zukunftssicherheit: Die Rolle von KI im Kampf gegen den Terrorismus

18. Martinez Demarco, S., M. Tahraoui & S. Kleemann (2025): Isolated in the service of society? The siloed logic and the implementation of the Artificial Intelligence Act in the law enforcement context: legal and ethical analysis of the applicability of accountability and responsibility to high-risk AI systems, in: Marie Eneman, Diana Miranda, William Webster und Jan Canbäck (Hrsg.): AI and Surveillance in Policing and Law and Order: Opportunities, Threats, Perspectives and Cases, Routledge.
19. Mittelstadt, B. (2019): AI Ethics – Too Principled to Fail? In: SSRN Journal: 1–15.
20. Ribeiro, M. T., S. Singh & C. Guestrin (2016): Introduction to Local Interpretable Model - Agnostic Explanations (LIME). An Introduction. O'Reilly, 12.08.2016.  
<https://www.oreilly.com/learning/introduction-to-local-interpretable-model-agnostic-explanations-lime>
21. Rohlfing, K. J., P. Cimiano, I. Scharlau, T. Matzner, H. M. Buhl, H. Buschmeier, E. Esposito et al. (2021): Explanation as a Social Practice: Toward a Conceptual Framework for the Social Design of AI Systems. IEEE Transactions on Cognitive and Developmental Systems 13 (3): 717–28.  
<https://doi.org/10.1109/TCDS.2020.3044366>
22. Saurwein, F. (2019): Automatisierung, Algorithmen, Accountability: Eine Governance Perspektive. Maschinenethik: Normative Grenzen autonomer Systeme: 35-56.
23. Selbst, A. (2017): Disparate impact in big data policing. Georgia Law Review 52: 109–195.  
<http://dx.doi.org/10.2139/ssrn.2819182>
24. Shapiro, A. (2017): Reform Predictive Policing. Nature 541, S. 458-459.
25. Yeung, D., Khan, I., Kalra, N. & Osoba, O. A. (2021): Identifying Systemic Bias in the Acquisition of Machine Learning Decision Aids for Law Enforcement Applications. RAND Corporation.  
<http://www.jstor.org/stable/resrep29576>
26. Zednik, Carlos (2019): Solving the Black Box Problem. A Normative Framework for Explainable Artificial Intelligence. Philosophy and Technology 34: 265-288.