AI ASSURANCE

ABSICHERUNG VON KI BASIERTEN SYSTEMEN

Thorsten Hansler, Dr. Willfried Wienholt

1 Jahr IABG in der Schweiz

Leitstellen-planung – KI in der Leitstelle

spik
SWISS POLICE ICT

APPROVED MEDICAL TRANSPORT www.144.ch

easyCab MEDICAL

iABG

Schweizer Polizei Informatik Kongress 2025, Nutzung von KI in der Leitstelle

01 Anwendungsfall - Einführung von KI in der Leitstelle

Lessons Learned: Was gibt es technisch zu beachten und welche Hürden sind zu überwinden um für die neue Technik eine Akzeptanz bei den Disponenten zu erzielen?

Zentrale Leitstelle Marburg-Biedenkopf

Zuständig für

250.000 Einwohner

Erfasst

90.000 Ereignisse pro Jahr

Zusammenarbeit mit

22 freiwilligen Feuerwehren

162 Ortsteilfeuerwehren

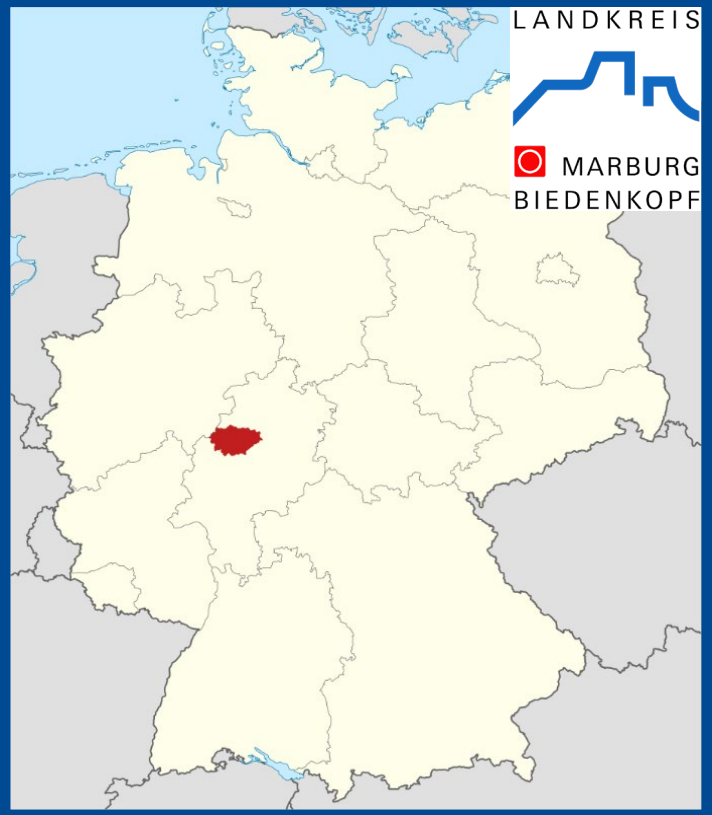
5 Werksfeuerwehren

1 Universitätsklinikum

2 Krankenhäuser der Grundversorgung

Veranlasst

13.700 Gebietsabsicherungen[1] im Jahr



Besteht aus

22 **Städten und Gemeinden**

[1] Verlegefahrten der Rettungswagen, um lokale Engpässe zu vermeiden

Bildquelle: https://commons.wikimedia.org/w/index.php?curid=6232035

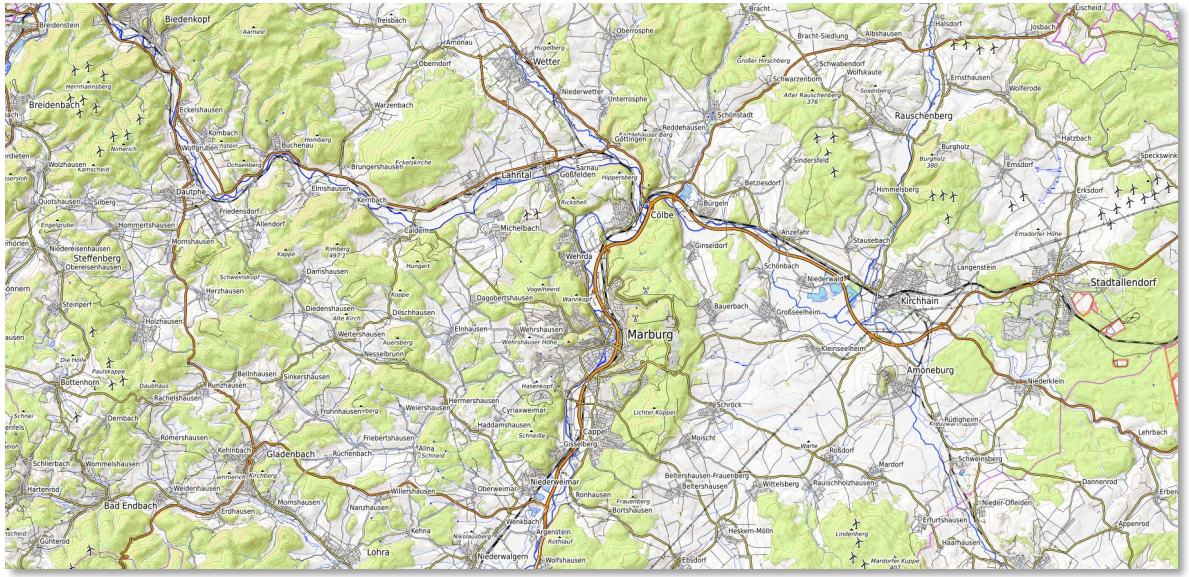
Herausforderungen der Leitstelle Marburg-Biedenkopf

Besonderheiten im Landkreis

- Ländliche Prägung und anspruchsvolle Topografie
- Hohes Hilfsfristniveau in Hessen: 10 min in 90 % der Fälle

Einhaltung der Hilfsfrist über ein risikoabhängiges Gebietsabsicherungskonzept

- Besetzen von nicht besetzten Rettungswachen nach festgelegtem Schema
- 13.700 Einsätze in 2022 zur Gebietsabsicherung
- Auf etwa 50 % der Verlegefahrten erfolgt ein Anschlusseinsatz
- Erheblicher Mehraufwand



Quelle: https://opentopomap.org/#map=12/50.8280/8.7661

Die allgemeine Steigerung der Einsatzzahlen, die höhere Auslastung des Rettungsdienstes und die Personalproblematik machen die Gebietsabsicherung immer schwieriger.

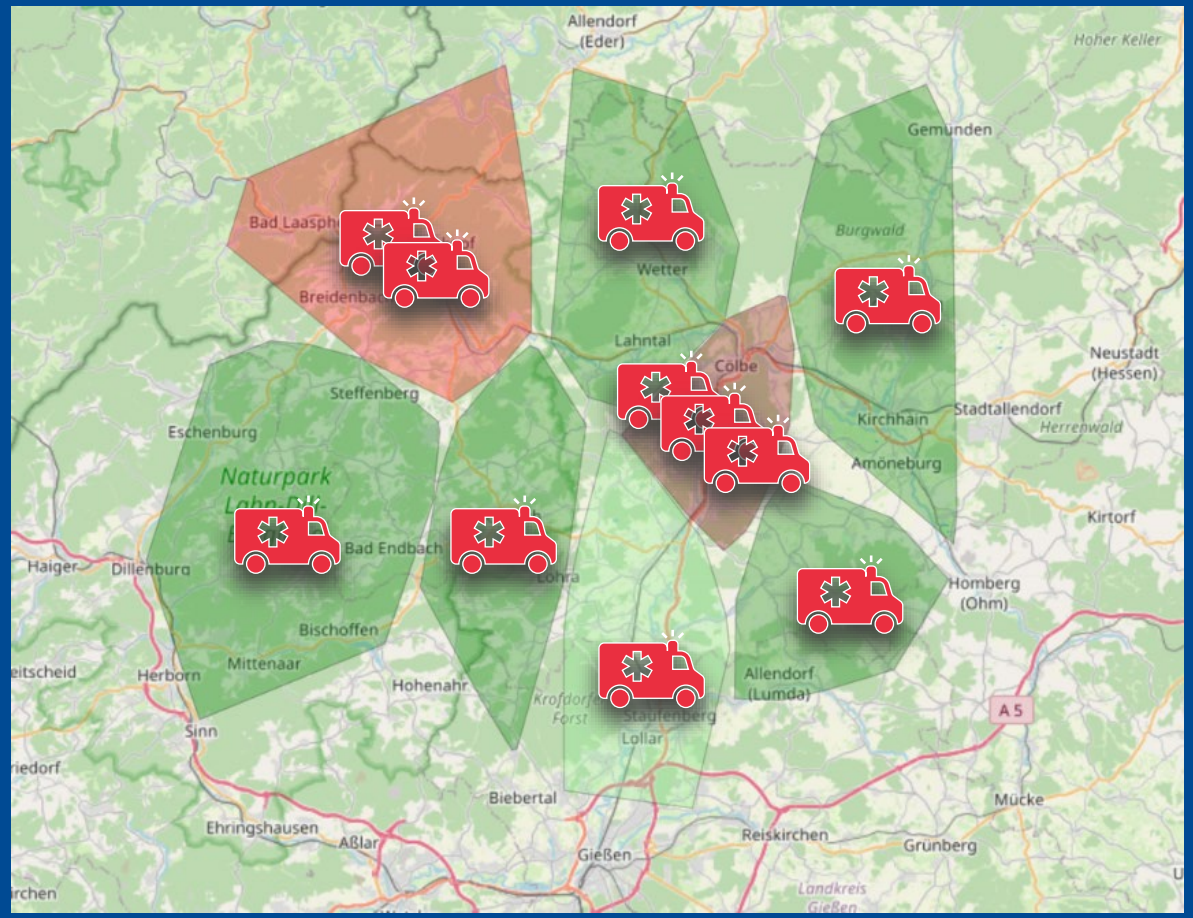
Optimierung der Ressourcenplanung

Einsparung an Ressourcen durch bessere Vorhersage

- des Einsatzaufkommens für die nächsten Stunden und Tage und
- der räumlichen Verteilung der Einsätze

Permanente Berechnung der optimalen Ressourcenallokation, um

- Verlegefahrten zu reduzieren und die
- Anschlusseinsatzquote zu erhöhen



Anwendungsbeispiel für den Landkreis

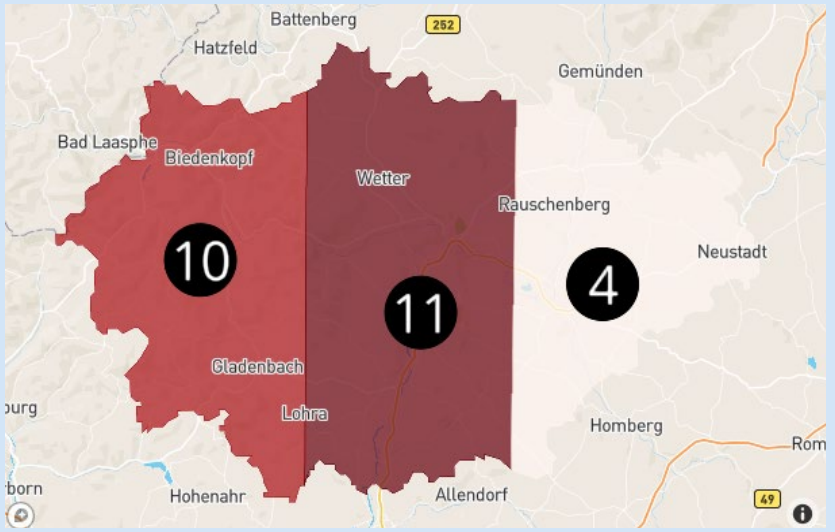
Inputs

- Historische Daten
- Wetterdaten
- Veranstaltungen
- Monat, Wochentag, Uhrzeit
- Demografie
- Geografische Informationen
- ...

KI Modell

Output

Prognose des Einsatzaufkommens für die zukünftigen Rettungseinsätze



LANDKREIS MARBURG BIEDENKOPF

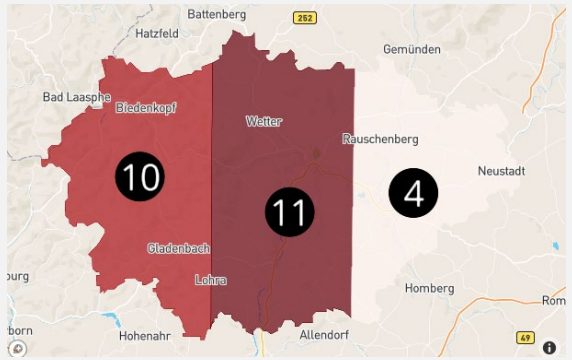
Daten: LK-Marburg

...

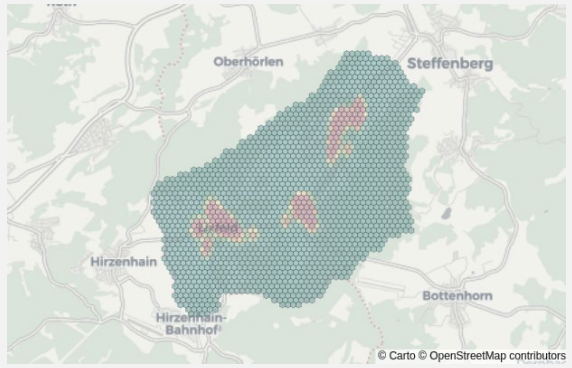
Unterstützung der Leitstelle in Marburg-Biedenkopf

Inputs

Volumen-Modell



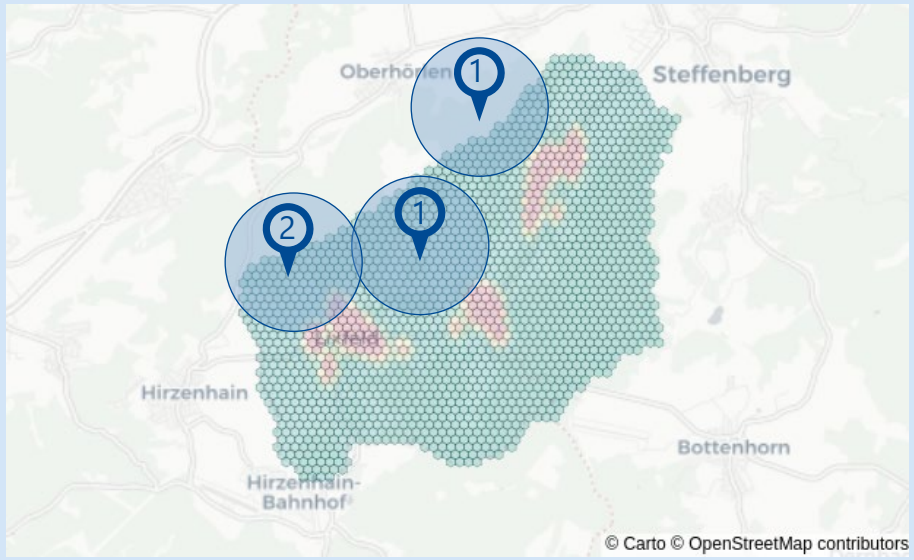
Verteilungs-Modell



Allokations-Modell

Output

Unterstützung bei der Allokation von Ressourcen



LANDKREIS MARBURG BIEDENKOPF

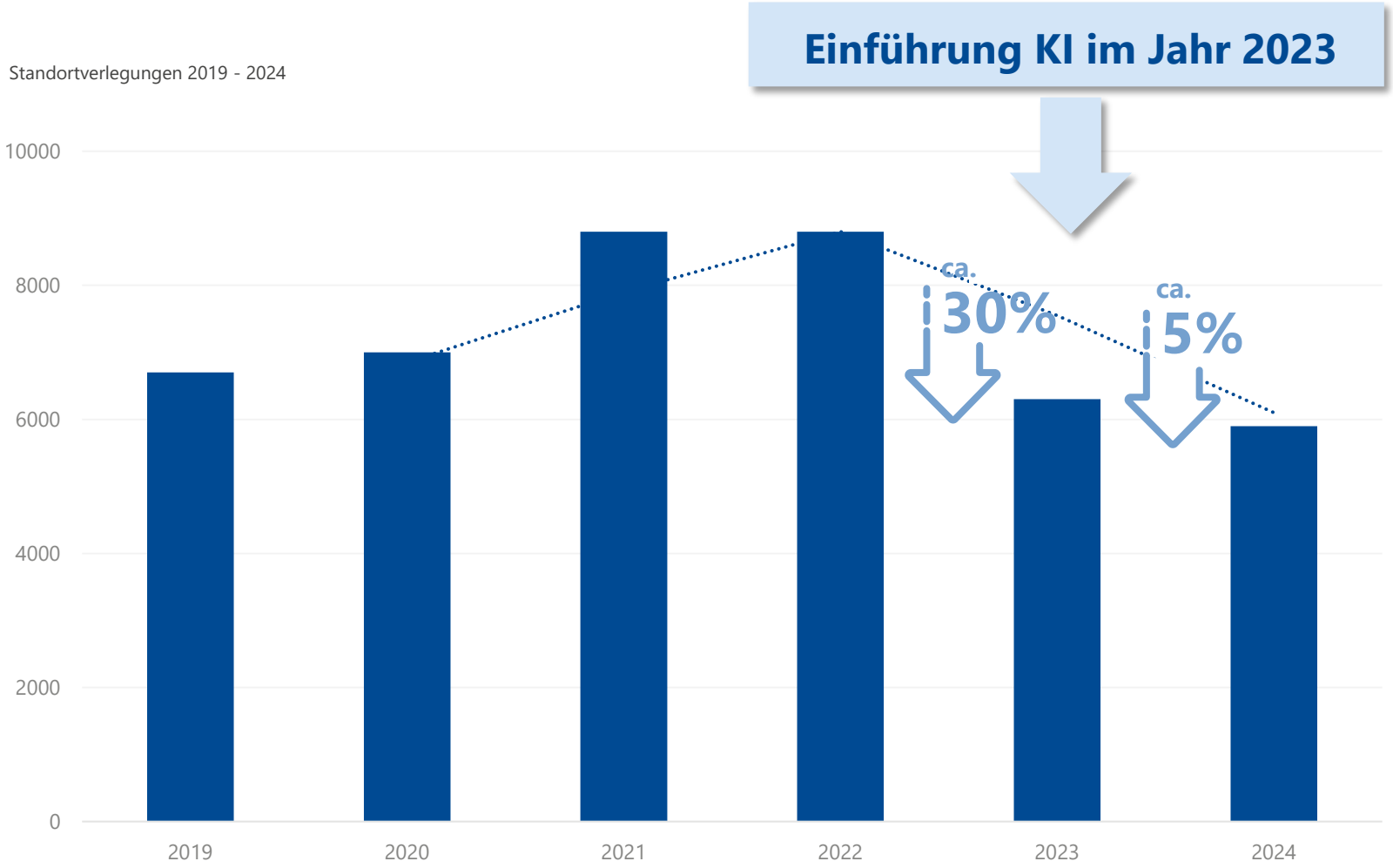
Daten: LK-Marburg

Ergebnisse seit Einführung der KI

Reduzierung der Verlegefahrten

2022 bis 2023: ca. 30%

2023 bis 2024: ca. 5%



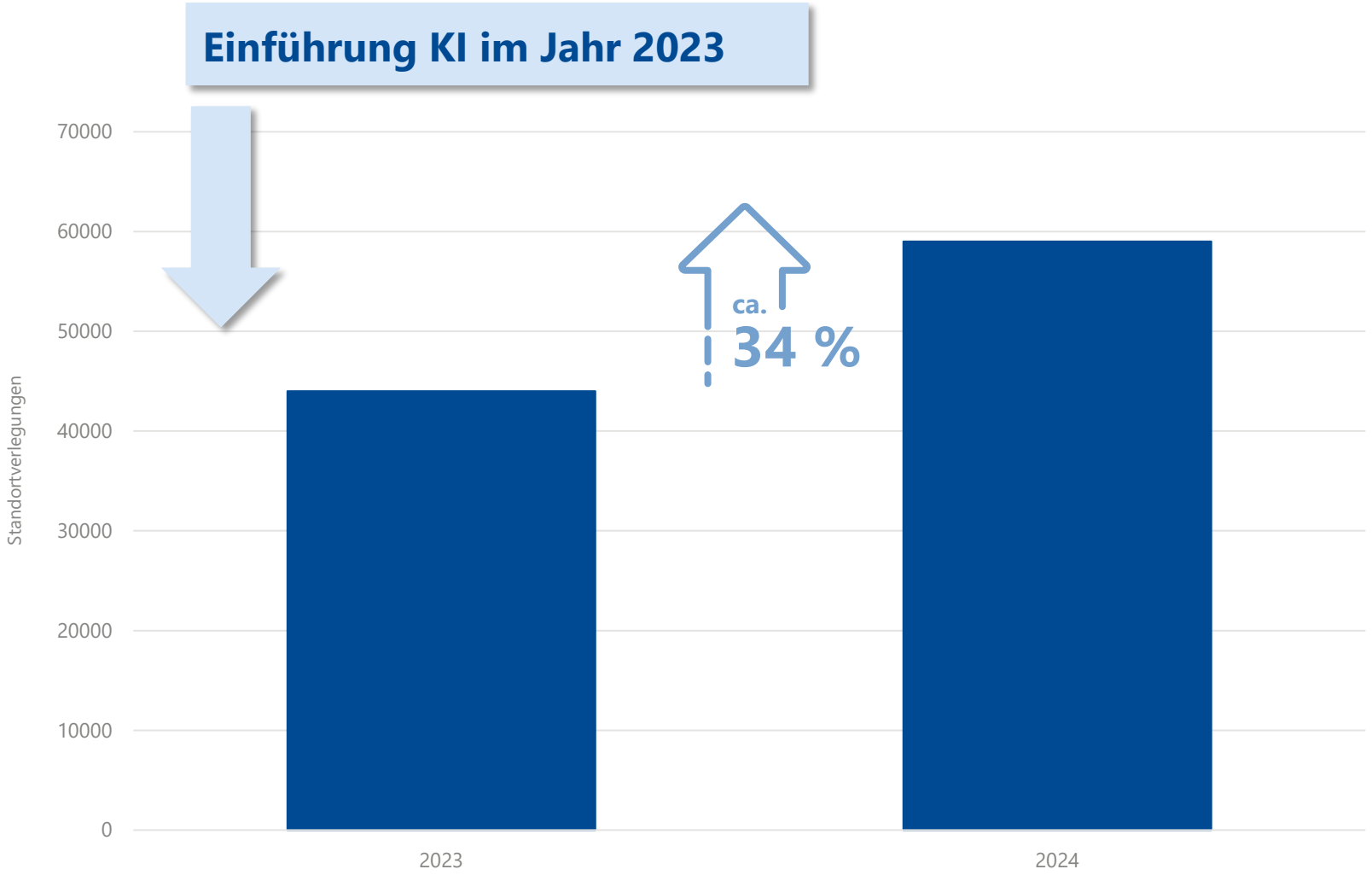
Ergebnisse seit Einführung der KI

Erhöhung der Einsatzzahlen

2023 bis 2024:

ca. **34 %**

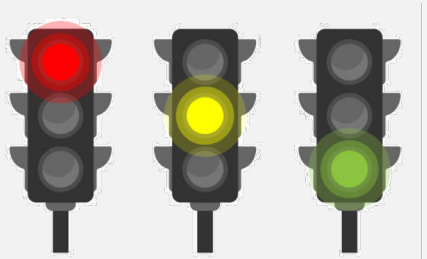
Keine Neubeschaffung von Fahrzeugen



Ergebnis

Faktor Mensch

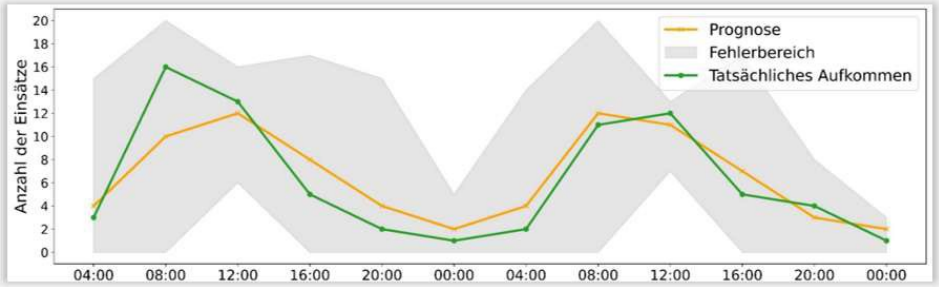
- Schnittstelle Mensch zu KI rein audiovisuell
- Keine automatische Dispo
- Klare Darstellung, ohne Interpretationsmöglichkeiten



- Rechtssichere Grundlage zur Verwendung der Informationen
- Nachvollziehbarkeit
- Dokumentation der Prozesse

Faktor Technik

- Evaluation,
  passen die Prognosen zur Realität?



- Gegenüberstellung Prognose/Auslastung/Vorhaltung
- Qualitätssicherung Prognose – Realität
- Grundlage ist eine hohe Datenqualität
- Einsatzdaten, Wetterdaten usw.
- Quantifizierte Unsicherheit (ISO / IEC 42)

Fazit

Der Mensch entscheidet, die KI gibt Empfehlungen

KI ist kein Produkt, sondern ein Prozess, der etabliert werden muss

Vertrauen & Akzeptanz durch zusätzliche Quantifizierung der Unsicherheit

KI muss lernen – Lernen benötigt Zeit!

KI unterstützt den Disponenten

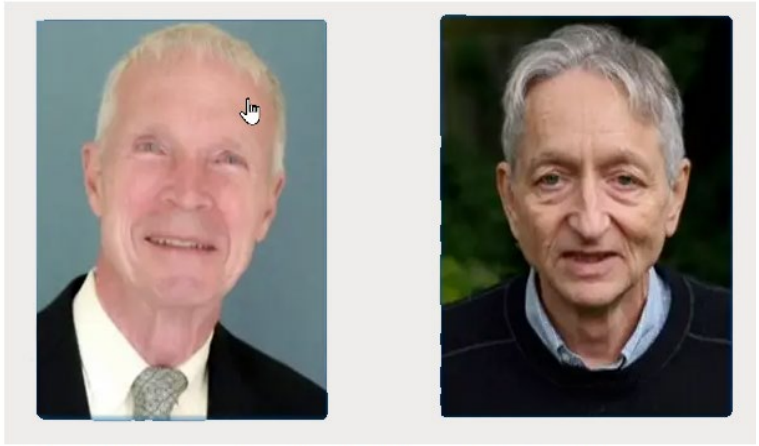
KI erhöht die Qualität in der Leitstelle

Created with help of Midjourney

02 AI Assurance

"A fool with a tool is still a fool"

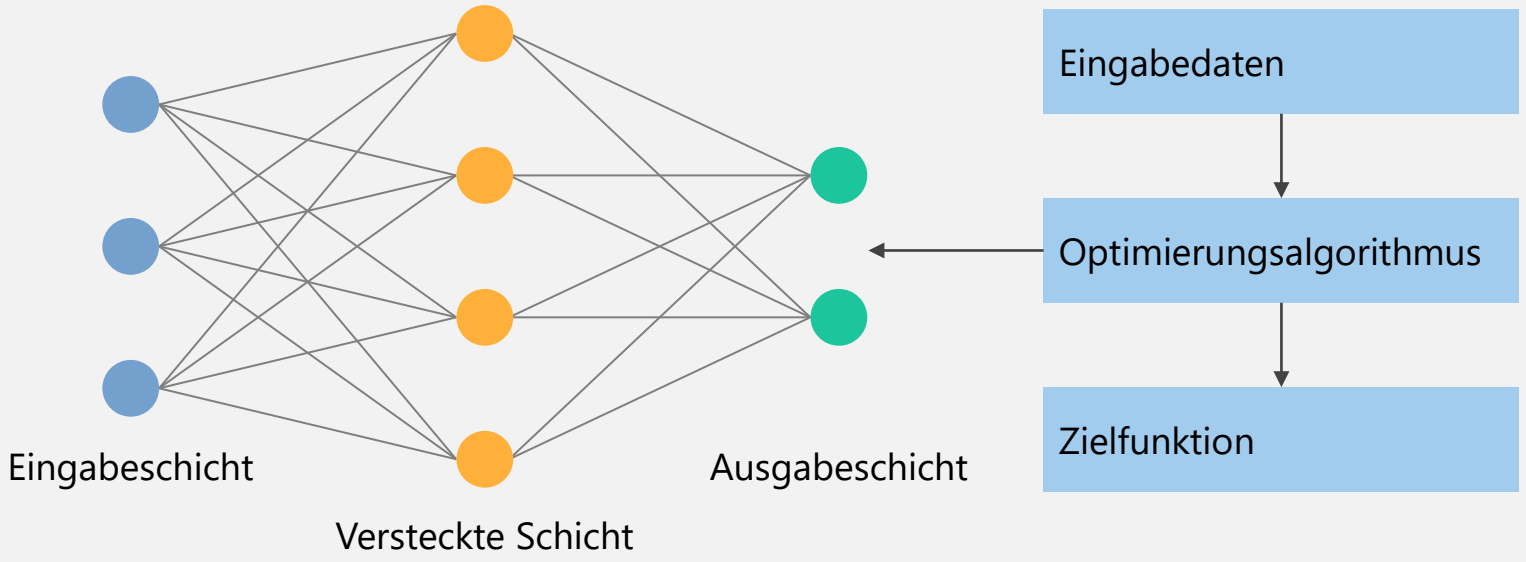
Nobelpreise 2024 in Physik und Chemie für maschinelles Lernen
KI Basics



John J. Hopfield (l.), Geoffrey E. Hinton (Bild: Mary Waltham / University of Toronto)

David Baker, Demis Hassabis und John M. Jumper (v.l.n.r.) (Bild: Nobel Media / Niklas Elmehed)

Vorwärtsgekoppeltes Neuronales Netz



Stark vereinfacht: die Kombination aus **Architektur** (Feedforward Neural Network), **Modell** (Konkrete Ausprägung der Architektur), **Optimierungs-algorithmus** (Backpropagation), **Zielfunktion** (Quadratischer Fehler) und **Daten** als **Metapher** für den **Lernvorgang in biologischen neuronalen Netzen**.

Typische Architekturen der KI

(begleitet von einer Vielzahl Bibliotheken mit Zielfunktionen und Optimierungsverfahren)

		Beispiele
Neuronale Netze	**1957**	**Perceptron**, das erste künstliche neuronale Netz
	1982	Einführung von **Rekurrenten Neuronalen Netzen**
	1997	Einführung von **LSTM** (Long Short-Term Memory) Netzen
	2006	Durchbruch im **Deep Learning** beim Training tiefer neuronaler Netze (viele Schichten / Parameter)
	2024	**Kolmogorov-Arnold Netze** (Basierend auf Arbeiten von A.N. Kolmogorov in 1956)
Objekterkennung	**1980**	Neocognitron, ein hierarchisches neuronales Netz für visuelle Mustererkennung
	1989	CNN LeNet, die erste erfolgreiche Anwendung von **Convolutional Neural Networks**
	2012	AlexNet gewinnt den ImageNet-Wettbewerb: bedeutender Fortschritt bei tiefen CNNs
	2015	ResNet führt tiefes residuales Lernen für Bilderkennung ein
Sprache-zu-Text	**1998**	Erste Software zur kontinuierlichen Spracherkennung mit großem Vokabular
	2014	**Encoder-Decoder-Architektur** für maschinelle Übersetzung
Natürliche Sprachverarbeitung / LLM	**2013**	Mit **Word2Vec** gelingt effiziente Erfassung von Semantik
	2017	**Transformer-Modell**, basierend auf Encoder-Decoder-Architektur
	2018	**BERT** (Bidirectional Encoder Representations from Transformers)
	2020	GPT-3, ein Sprachmodell mit 175 Milliarden Parametern
	2022	ChatGPT, eine Konversations-KI basierend auf GPT-3.5
	2023	GPT-4 → zunehmender „Hunger an Ressourcen“
	2025	Januar: DeepSeek V3 (671 Milliarden Parameter)
Vertrauen & Akzeptanz	**2024**	**EU AI Act**

Wettstreit LLM:
Open Source versus proprietäre Lösungen

https://huggingface.co/

Typische Anwendungsfälle für KI am Beispiel Militär

Aufgaben fallen bei vielen Systemen an, Lösungen wollen verschiedene KI Technologien nutzen

Missionsplanung und Durchführung	Zieldetektion, -erkennung und -identifikation (DRI)	Situationsbewusstsein (Situational Awareness)	Navigation und Steuerung	Cyber Security und Resilienz
• Missionsplanung mit Hilfe von Reinforcement Learning • Routenoptimierung	• Bildklassifizierung • (Multi-)Objekterkennung	• Prädiktive Analytik • Autonome Überwachung • Sprache-zu-Text-Anwendungen	• Optimierung Manned/ Unmanned Teaming • Autonome Routenplanung	• Jamming-Detektion • Erkennung von Hackingversuchen
KI Technologien				
• **Hybrid (KI plus Graphentheorie)**	• **Convolutional Neural Network** • **Transformer**	• **Recurrent Neural Network** • **Transformer** • **Hybrid (KI plus Graphentheorie)**	• **Recurrent Neural Network** • **Reinforcement Learning** • **Hybrid (KI plus Graphentheorie)**	• **Convolutional Neural Network** • **Reinforcement Learning**

Problemlösung

Früher

Forschung

So wenig Daten und Rechenkapazität … Was soll ich tun?

Definiere Rahmenbedingungen, entwickle **ein** Modell und beweise (mathematisch), dass das Modell die Lösung des Problems ist. Nutze Daten, um die Parameter des Modells zu schätzen.

- ⊕ Mathematischer Beweis
- ⊖ Mangelnde Passung des Modells an die Realität

Source: Midjourney "/imagine a mathematician writing a lengthy formula on the chalk board"

Heute

Empirie

So viele Daten und Rechenkapazität … Was soll ich tun?

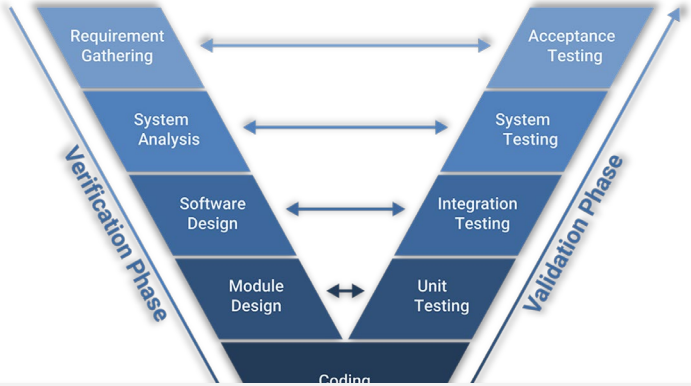
Wähle eine Architektur und Zielfunktion und finde durch Nutzung von Daten und Variation von Architektur- und Modell-Parametern eine Lösung, die das Problem bestmöglich löst.

- ⊕ Meist gute Passung des Modells an die Realität
- ⊖ Fehlender "Beweis"

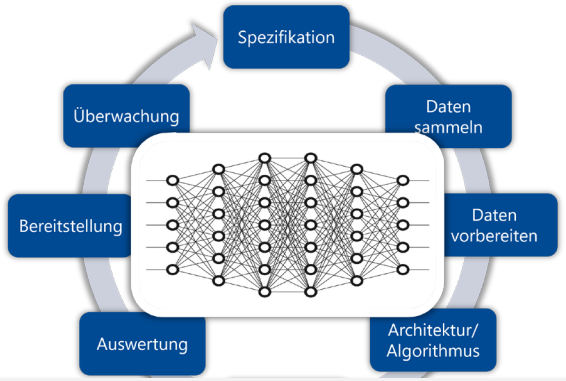
Source: Midjourney "/imagine an ai specialist using data to find a solution"

V-Modell versus KI-Zyklus

Ist KI *automatisch* die bessere Lösung?



"Typische" **Ingenieurlösungen**, z.B.
Regelungstechnik: Kalman Filter
Zeitreihen: Autoregressive Modelle
Encoder / Decoder: Hauptkomponentenanalyse (PCA)



Viel **Empirie** am Werk (leider):
Architektur, Optimierungsverfahren, Zielfunktion, Daten?

Oft fehlt Bezug zur einer typischen Referenzlösung
Leistungsversprechen meist unklar

EU AI Act:
Vertrauen & Akzeptanz in KI

Für die **Bewertung**, **Prüfung** und **Beurteilung** von KI-Systemen mit Blick auf Sicherheit und Vertrauenswürdigkeit sind **(neue) Methoden und Standards** erforderlich!

EU Artificial Intelligence Act

AI Assurance

AI Assurance bezeichnet den Ansatz zur Sicherstellung, dass KI-Systeme …

… sicher

Ziele der AI Assurance

- Stärkung des **Vertrauens** in KI-Systeme
- Förderung der **Akzeptanz** bei den Anwendern und Stakeholdern
- Minimierung von **Risiken** und potenziellen Schäden

… zuverlässig

Schlüsselkomponenten

- **Transparenz:** Nachvollziehbarkeit von Entscheidungen
- **Robustheit:** Widerstandsfähigkeit gegen Angriffe und Störungen
- **Ethik:** Vermeidung von Diskriminierung und Verzerrungen
- **Compliance:** Einhaltung gesetzlicher Vorgaben und Standards

… fair &

Anwendungsszenarien im Militär/Sicherheitsbehörden

- Einsatz von KI zur Unterstützung von **Entscheidungsprozessen** in Echtzeit
- Entwicklung sicherer, **autonomer Systeme** für Aufklärung und Schutz
- Schutz militärischer Netzwerke und Systeme vor Cyberangriffen durch KI-basierte **Anomalie-Erkennung und automatisierte Abwehr-mechanismen**
- Einsatz von KI zur **Früherkennung** und **Prävention** von physischen und digitalen Angriffen auf kritische Infrastrukturen
- Simulationsbasierte Szenarien zur **Reduzierung von Risiken** bei der Entwicklung neuer Technologien.

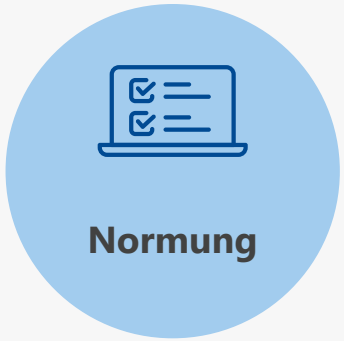
… rechtskonform sind

Herausforderungen

- **Schutz** sensibler Daten und Abwehr von Cyberbedrohungen
- **Nachvollziehbarkeit** von KI-Entscheidungen
- Einhaltung **gesetzlicher und normativer Vorgaben**
- **Robustheit** und Integration in bestehende Systeme
- Sicherstellung von qualitativ hochwertigen und vollständigen Daten
- **Anpassung** an sich ändernde Bedrohungen

safeAI kurz und bündig – Umsetzung AI Assurance

Europäisches KI-Gesetz



Einhaltung der Vorschriften

Beitrag

ISO/IEC JTC 1/SC 42 Künstliche Intelligenz

ISO JTC 1 IEC
INFORMATION TECHNOLOGY STANDARDS

Europäische Ebene

EUROCAE

cen

Nationale Ebene

- Leitung des DIN / DKE Gremiums zu KI Vertrauenswürdigkeit
- Leitung der DIN SPEC 92005 Uncertainty Quantification in Machine Learning

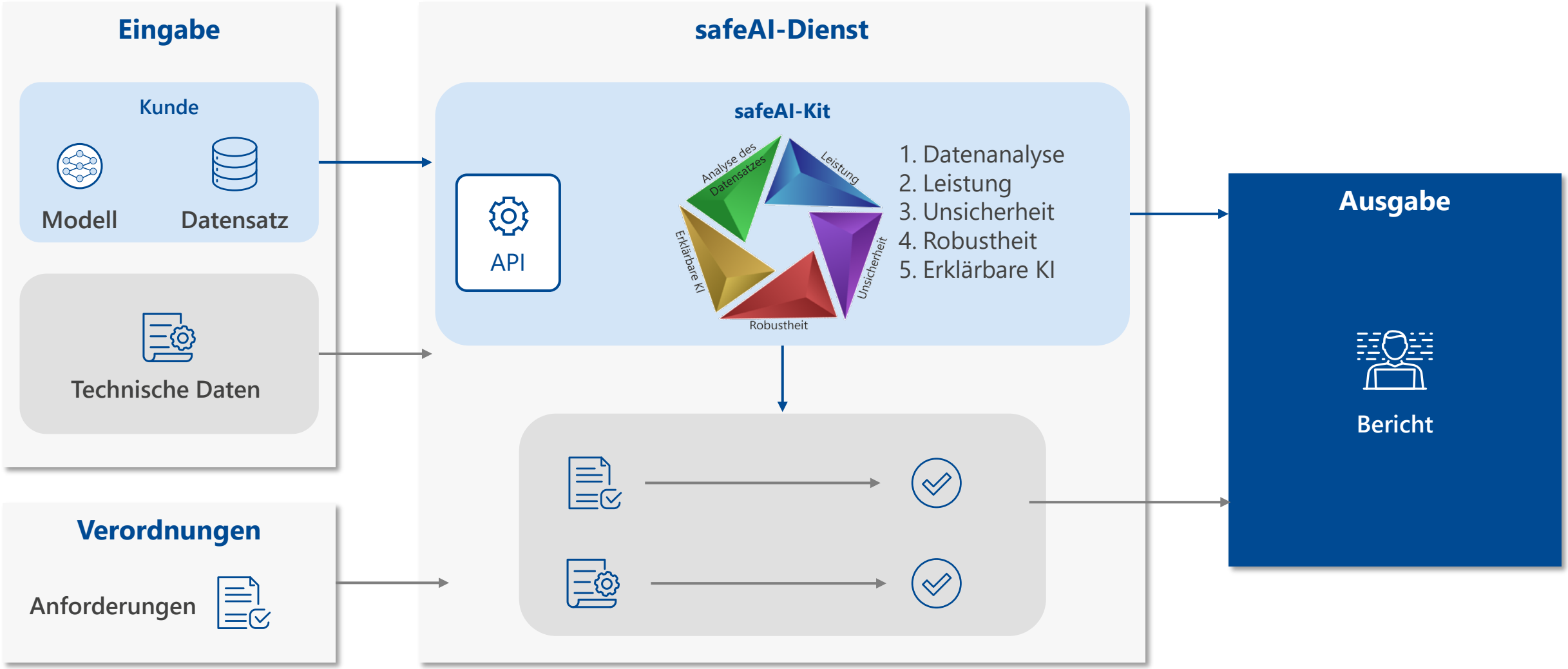
KI-Bewertungsmethoden

Wissenschaftliche Arbeiten

KI-System

Bewertung der Konformität

safeAI-Dienst



Zusammenfassend: ein fundiertes und fachlich-übergreifendes Verständnis ist hilfreich, um das Beste *aus* und *mit* KI zu machen

- **Architekturen / Modelle / Zielfunktionen**
  Lineare Algebra … Funktionalanalysis, partielle DGLn
- **Daten**
  "Feature Engineering", lineare multivariate Stochastik, Zeitreihenanalyse, Umgang mit große Datenmengen
- **Optimierungsverfahren**
  Informatik, angewandte Mathematik, Konvergenz von Algorithmen (numerische Mathematik)
- **Entwicklung**
  SW-Entwicklung und -Prozess, Kubernetes, GitHub, KI-Bibliotheken, Nutzung von Open Source SW
- **Konformitätsbewertung**
  Funktionale Sicherheit, Referenzmodell und quantifizierte Unsicherheit sollten Minimal-anforderungen sein

- Einbettung / Kombination mit weiteren disruptiven Technologien: **Cloud Technologien**
- **Ressourceneffizienz**: eine Lösung sollte so einfach wie möglich sein (!)

- Moderne **Lernplattformen** bieten beeindruckende Lerner(l/g)ebnisse:
  - https://www.coursera.org/
  - https://www.deeplearning.ai/

IABG LEISTUNGEN

Wie sicher ist künstliche Intelligenz?

Wir entwickeln **Lösungen** für die **Evaluierung** und **Prüfung** von KI-Systemen basierend auf neuesten Erkenntnissen aus **KI-Forschung**, -**Standardisierung** und **Regulierung** sowie auf unserer langjährigen Erfahrung bei **Test-**, **Analyse-** und **Zertifizierungsprozessen**.

Unser **Leistungsversprechen**:

- **Beratung** bei der Erfüllung von KI Standardisierung, Normierung und Regulierung inkl. der relevanten Prozesse
- **Evaluierung und Prüfung** der Leistung und Sicherheit von KI-Systemen und Datensätzen. Beispiel: KI in der Leitstelle.

Vielen Dank

Ihre Ansprechpartner

Industrieanlagen-Betriebsgesellschaft mbH

Einsteinstr. 20
85521 Ottobrunn

Tel. +49 89 6608-0
Fax +49 89 6608-2220

info@iabg.de
www.iabg.de

IABG, NL Bonn

Königswintererstr. 552b
53227 Bonn

Thorsten Hansler
Programm Manager
KI in der Leitstelle

hansler@iabg.de
Telefon: +49 228 91767-23

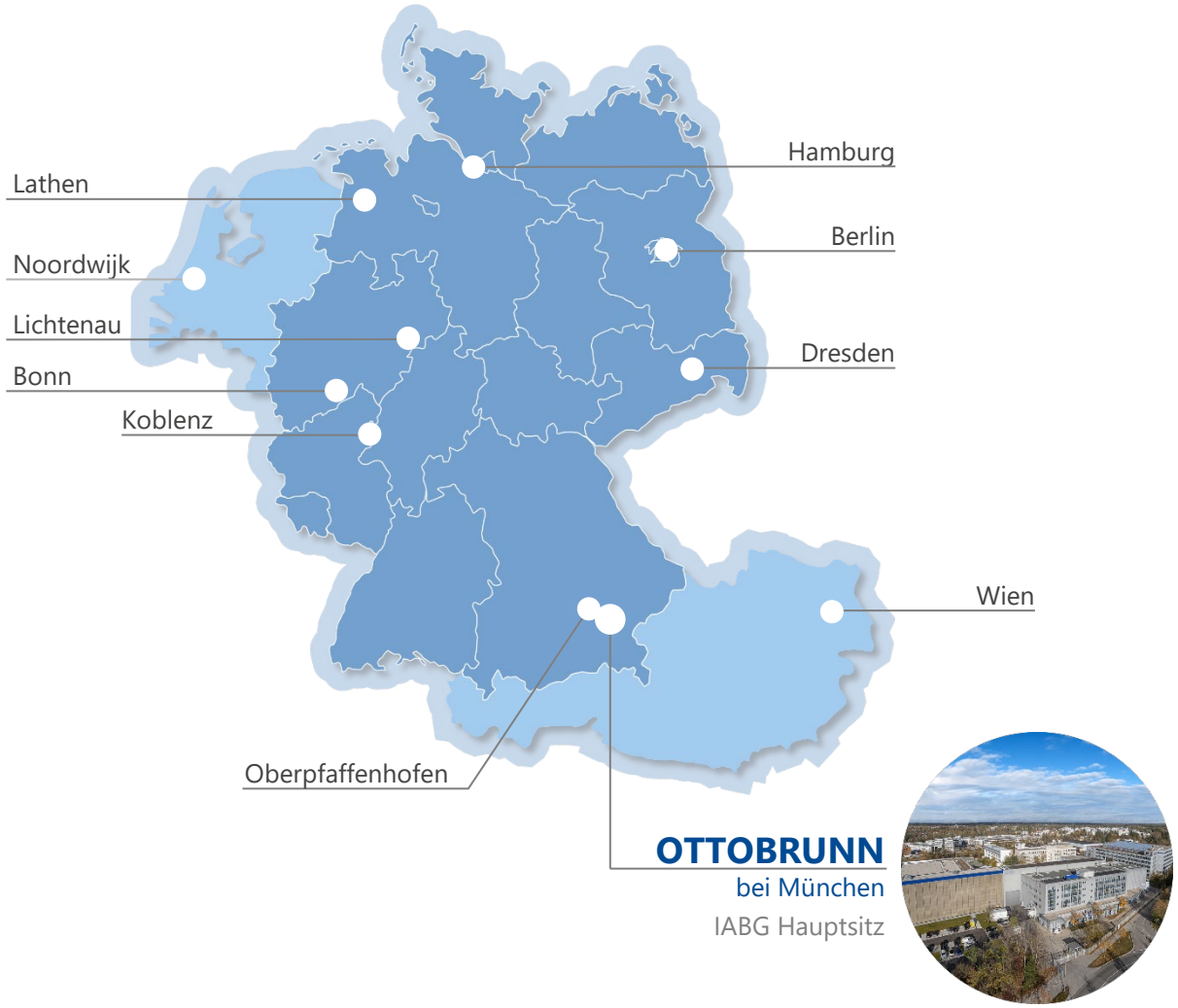
Dr. Willfried Wienholt
Ressortleiter InfoKom
Sichere Cloud & KI

wienholt@iabg.de
Telefon: +49 151 74431308

00 Unternehmen IABG

Technologieunternehmen im Herzen Europas.

Daten und Fakten auf einen Blick.



Industrieanlagen-Betriebsgesellschaft mbH

Name	Industrieanlagen-Betriebsgesellschaft mit beschränkter Haftung
Gründung	1961
Hauptsitz	Ottobrunn bei München
Aufsichtsratsvorsitzender	
Geschäftsführung	Prof. Dr. Rudolf F. Schwarz Thomas Köhler
Anzahl Mitarbeiter	Ca. 1.200
Umsatz	Ca. 203 Mio. € (2023)
Standorte	11

Wir unterstützen die Schlüsselbranchen Europas

Automotive

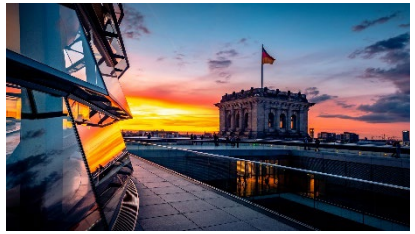
Entwicklung und Bau von Prüfsysteme und Testzentren; Methodenentwicklung & Virtuelle Qualifikation; Absicherung neuer Technologien für autonomes Fahren sowie der Elektromobilität

Bahn und Schiene

Qualifizierung von Wagenkästen und Drehgestellen. Unterstützung der Digitalisierung, z. B. SmartTram / Automated Train Operation

Energietechnik

Absicherung von Anlagen und Verfahren zur Gewinnung erneuerbarer Energien, z. B. Batterien, Hochvoltspeichern, H2-Tanks, Brennstoffzellen, Windenergieanlagen



Öffentlicher Sektor

Sichere Vernetzung von ITK, Leitstellen & Lagezentren sowie Unterstützung bei der Vergabe. Geoinformation für Krisenmanagement, Katastrophenschutz, Energiewende & Klimaanpassung

Luftfahrt

Qualifikation von Bauteilen, Komponenten und komplexen Gesamtsystemen; Lebensdauervorhersagen; Digital Twins & virtuelle Qualifikation

Raumfahrt

Umwelt- & Qualifikationstests von Satelliten und Trägersystemen im Raumfahrtzentrum sowie im Competence Centre Optics. Serviceprovider in den EU-Programmen Galileo und Copernicus

Verteidigung

Unabhängige Beratung, Prüfdienstleistungen und Simulation für Streitkräfte in allen operationellen Dimensionen und Fähigkeitsdomänen

Innovation

Innovationszentrum als Inkubator für das Leistungsportfolio mit Fokus auf Digitalisierung, Data Science & sichere Anwendung von KI sowie vernetzte, elektrische, autonome Systeme

Digitalisierung in kritischen Infrastrukturen sicher gestalten

Mitgliedschaften

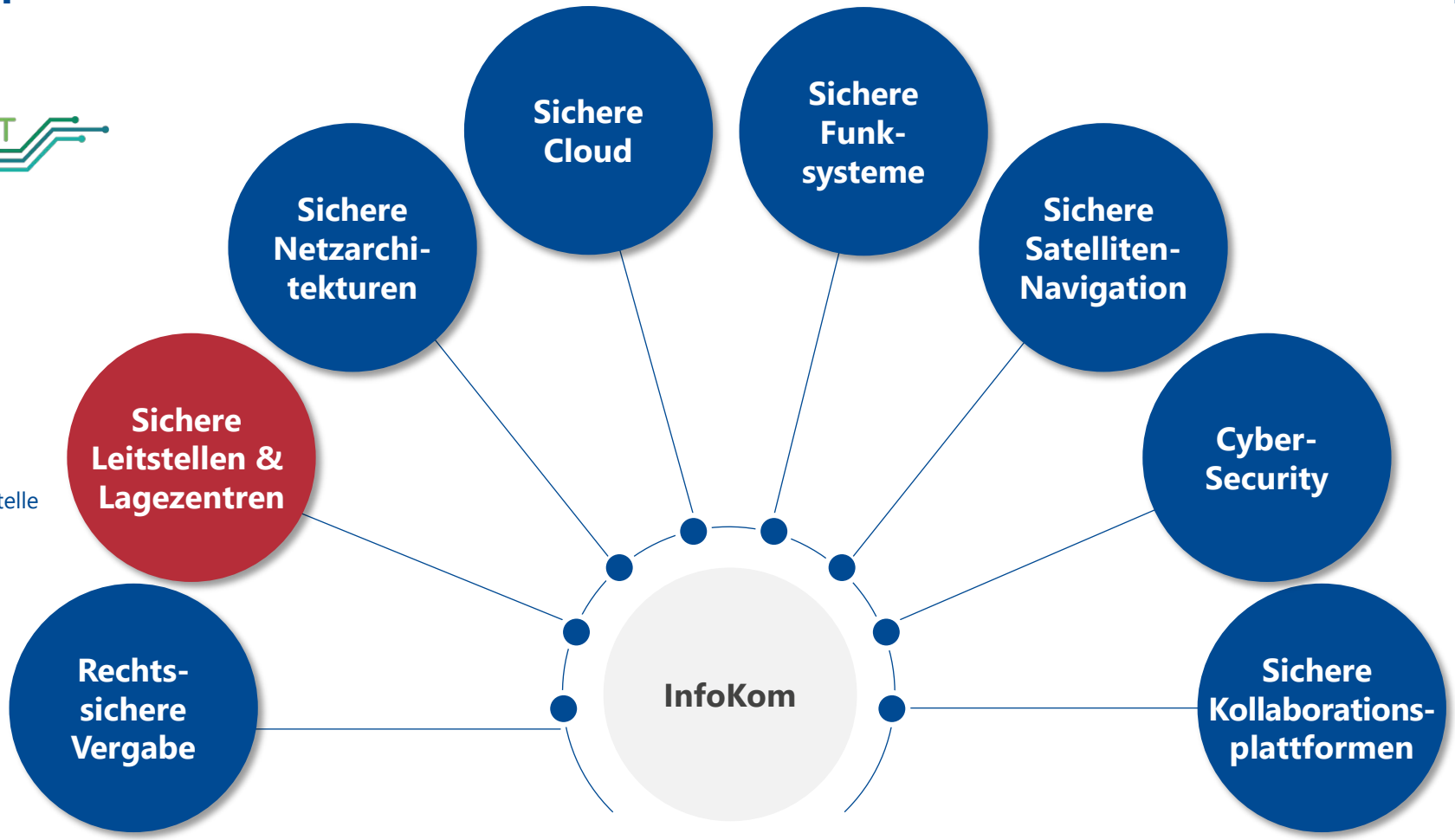
SWISS POLICE ICT

Fachverband Leitstellen e.V.

PMeV NETZWERK SICHERE KOMMUNIKATION
Expertenforum Universelle Leitstellenschnittstelle

Zukunftsforum Öffentliche Sicherheit

BSI-zertifizierter IT-Dienstleister



Langjährige Erfahrungen im Bereich Leitstellen & Lagezentren

Beispielhafte Leistungen und Kunden

- **Anforderung & Planung**
- **Modell-gestützte IT-Systemarchitekturen**
- **Vernetzungskonzepte**
- **Cyber-Security**
- **Informationssicherheit ISO 27001, BSI-Grundschutz**
- **Vergabeunterstützung**
- **IT-Qualitätssicherung**
- **Großprojektmanagement & Controlling**