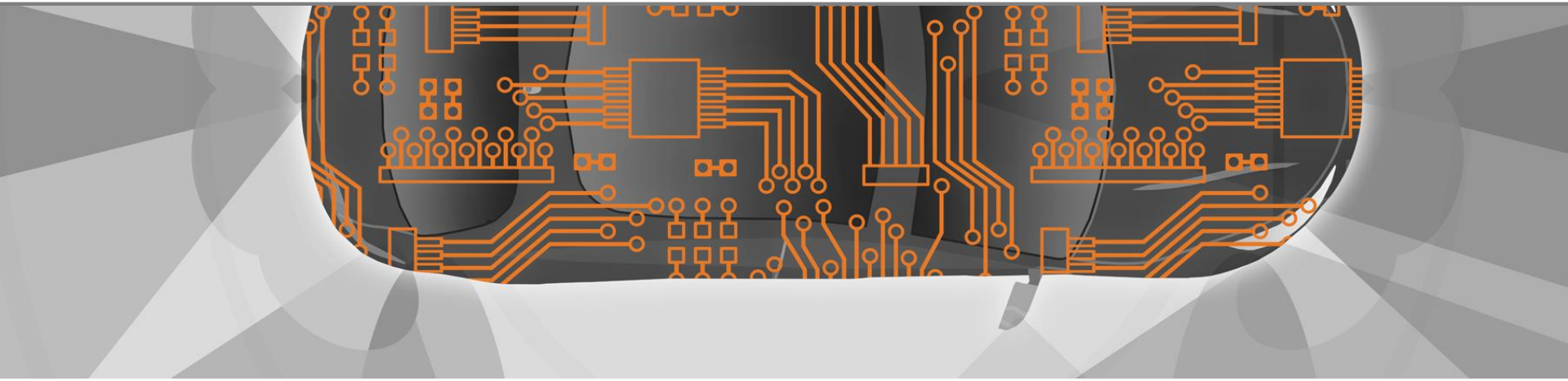


Weiterentwicklung zuverlässiger KI

Integration von Deep Learning mit funktionalen Sicherheitsstandards in autonomen Fahrzeugen

Dr. Rüdiger Nortmann – MCG Management Consult GmbH
Stephen Cobeldick – FSQ Experts – A Brand of Wertefest GmbH

Einführung und Überblick

Herausforderung: KI-Sicherheit im Automobil

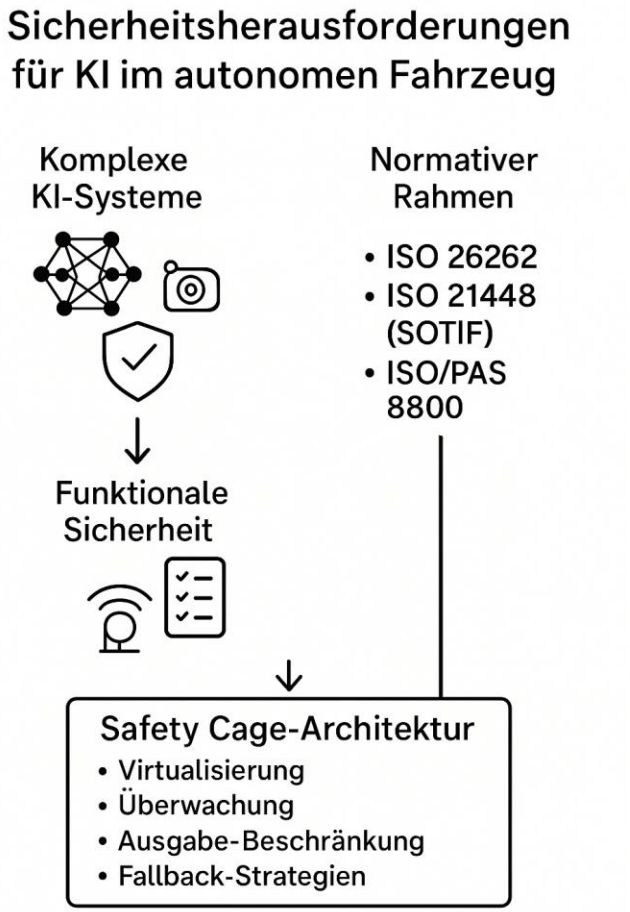
- Komplexität von Wahrnehmung, Planung und Entscheidungsfindung
- Sicherheitsanforderungen als Schlüssel zur Marktreife

Ziele und Agenda

- Was Sie heute mitnehmen werden
- Überblick der Präsentation

Herausforderung: KI-Sicherheit im Automobil

- Komplexität von Wahrnehmung, Planung und Entscheidungsfindung
  - Autonome Fahrzeuge müssen vielfältige Umgebungen erfassen und komplexe, situationsabhängige Entscheidungen in Echtzeit treffen
- Sicherheitsanforderungen als Schlüssel zur Marktreife
  - Nur durch Erfüllung strenger Sicherheits-standards wird eine breite, vertrauenswürdige Markteinführung autonomer Fahrzeuge ermöglicht
- Die Integration von KI in autonome Fahrzeuge erfordert einen normativen Sicherheitsrahmen und spezielle Schutzmechanismen, um Risiken beherrschbar zu machen



Ziele und Agenda

- Was Sie heute mitnehmen werden
  - Sie erhalten praxisnahe Einblicke in Deep Learning, Sicherheitsstandards und deren Zusammenspiel im autonomen Fahren
- Überblick der Präsentationsstruktur
  - Die Agenda führt systematisch von Grundlagen, über Normen und Praxisbeispiele, bis zu Ausblick und Diskussion
- Die Präsentation verbindet Technologie, Normen und Praxiserfahrung für sichere, KI-basierte Fahrzeugfunktionen

Normativer Rahmen für KI-Sicherheit

ISO 26262: Funktionale Sicherheit für Straßenfahrzeuge

- Grundprinzipien und Anwendbarkeit auf KI-Systeme
- Lücken bei KI-spezifischen Anforderungen

ISO 21448: Safety of the Intended Functionality (SOTIF)

- Umgang mit unzureichenden Funktionalitäten ohne Systemfehler
- Anwendbarkeit auf KI-Systeme im Automobil

ISO/PAS 8800: Sicherheit und KI in Straßenfahrzeugen

- Kernkonzepte und Anforderungen
- Integration mit bestehenden Standards

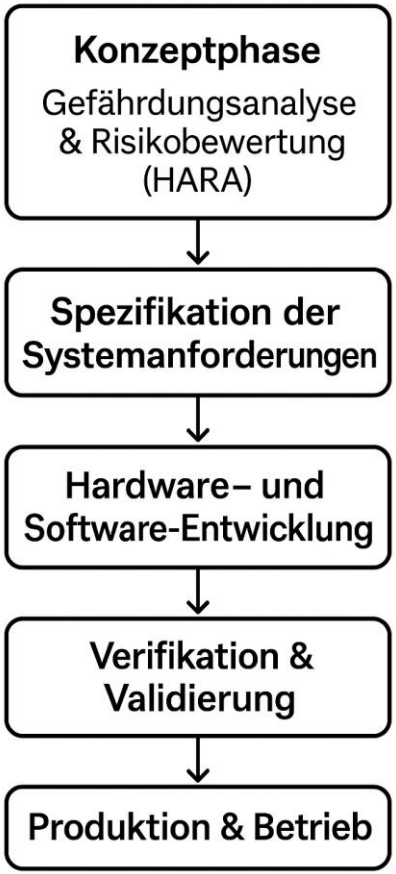
Integration der Standards: Ein ganzheitlicher Ansatz

- Synergie zwischen den Standards
- Sicherheitsnachweisführung bei KI-basierten Systemen
- Neue Maßstäbe für KI-Sicherheit in SAE Level 3-4 Fahrfunktionen

ISO 26262: Funktionale Sicherheit für Straßenfahrzeuge

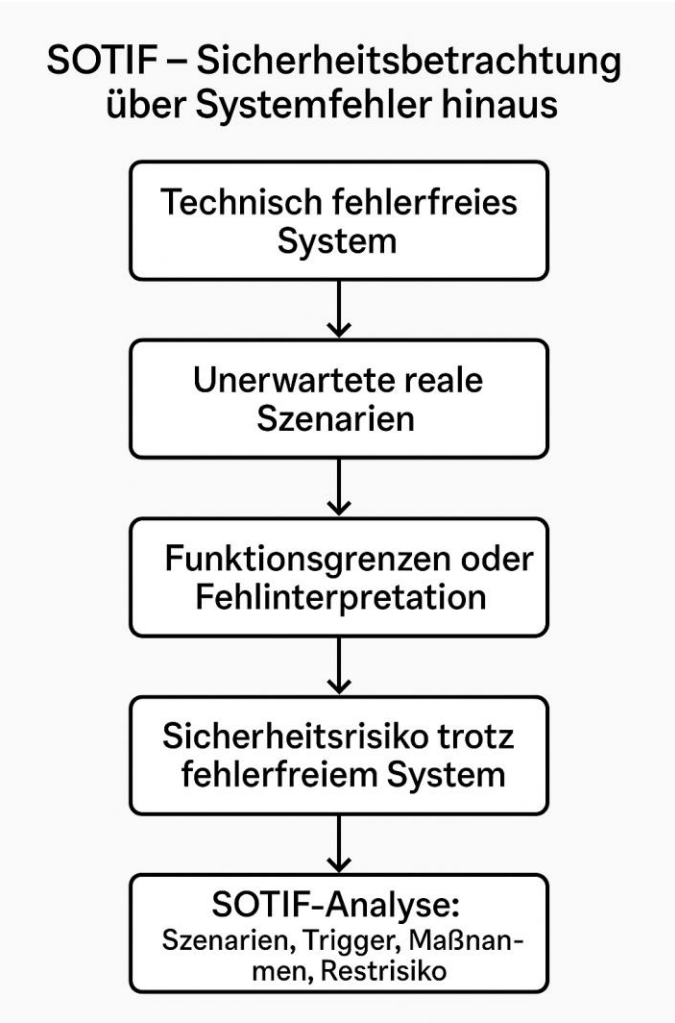
- Grundprinzipien und Anwendbarkeit auf KI-Systeme
  - ISO 26262 definiert Risikomanagement und Sicherheitsziele, die für KI-Systeme adaptiert werden müssen
- Lücken bei KI-spezifischen Anforderungen
  - Fehlende Leitlinien für KI-Dynamik, Datenqualität und unvorhersehbare Szenarien
- ISO 26262 strukturiert den gesamten Entwicklungsprozess und stellt die Erfüllung aller Sicherheitsanforderungen sicher

ISO 26262 - Sicherheitslebenszyklus



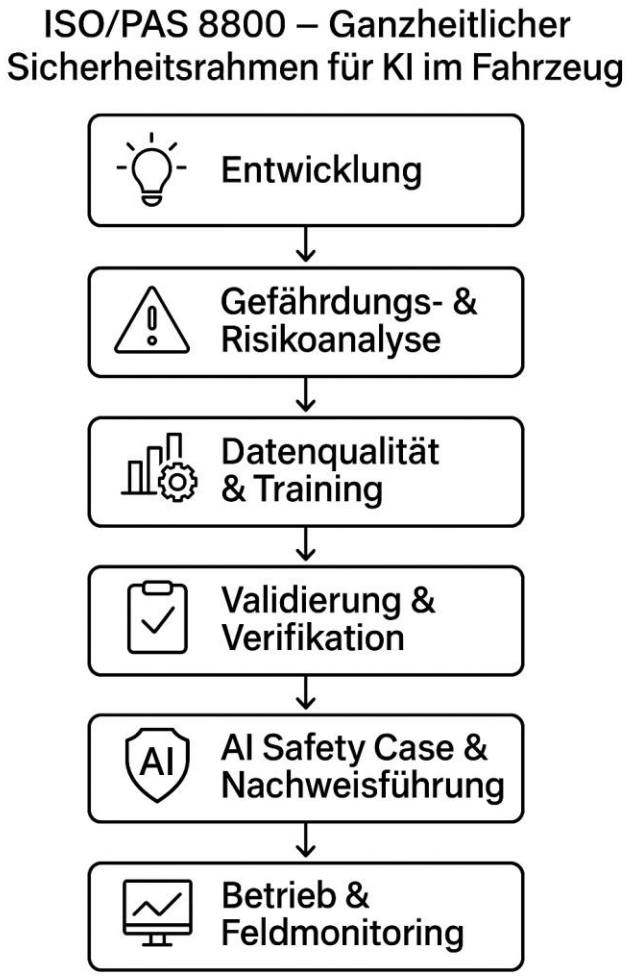
ISO 21448: Safety of the Intended Functionality (SOTIF)

- Umgang mit unzureichenden Funktionalitäten ohne Systemfehler
  - SOTIF adressiert Risiken, die durch funktionale Grenzen trotz fehlerfreiem Systemverhalten entstehen
- Anwendbarkeit auf KI-Systeme im Automobil
  - SOTIF ist hilfreich für KI-basierte Systeme, die komplexe, reale Umgebungen interpretieren müssen
- SOTIF betrachtet Risiken, die durch funktionale Grenzen und nicht durch klassische Fehler entstehen



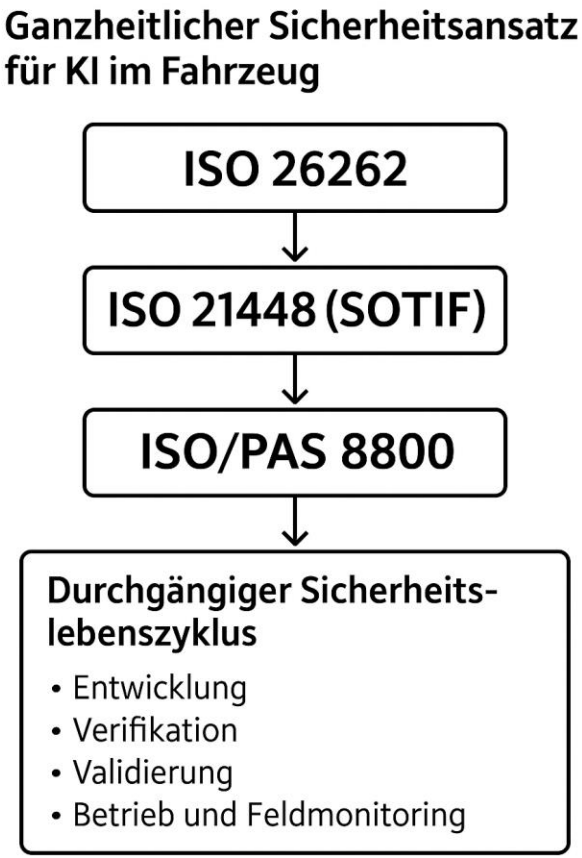
ISO/PAS 8800: Sicherheit und KI in Straßenfahrzeugen

- Kernkonzepte und Anforderungen
  - ISO/PAS 8800 definiert strukturierte Prozesse für Risikoanalyse, Datenqualität, Validierung und Nachweis der KI-Sicherheit
- Integration mit bestehenden Standards
  - ISO/PAS 8800 ergänzt und erweitert ISO 26262, ISO 21448 (SOTIF) und andere relevante Automobilstandards
- ISO/PAS 8800 schafft einen strukturierten, durchgängigen Sicherheitsrahmen für KI-Systeme im Automobil

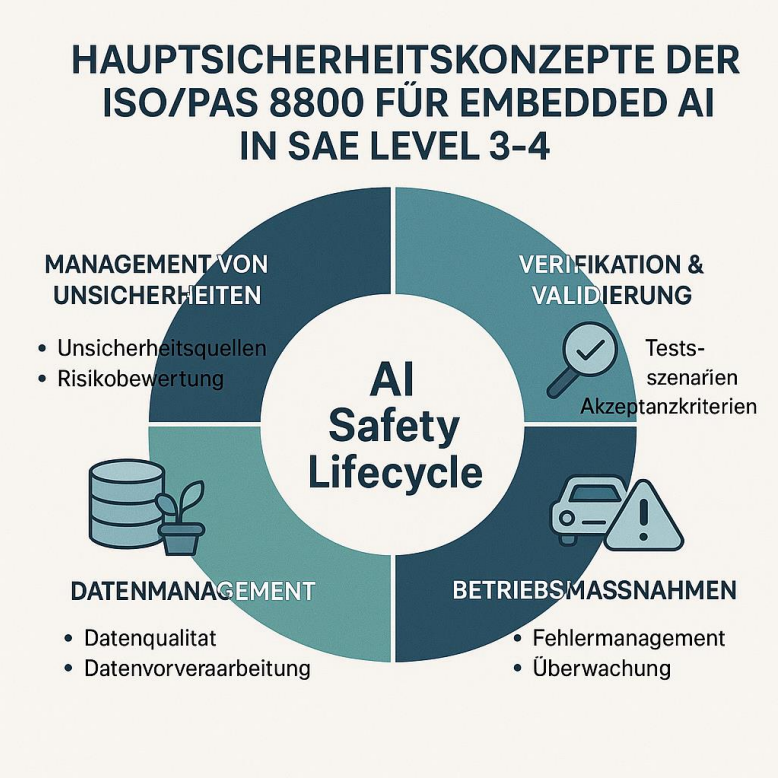


Integration der Standards: Ein ganzheitlicher Ansatz

- Synergie zwischen den Standards
  - Die Normen ergänzen sich, um alle sicherheitsrelevanten Risiken von KI-Systemen umfassend abzudecken
- Sicherheitsnachweis-führung bei KI-basierten Systemen
  - Ein projektspezifischer AI Safety Case konsolidiert Sicherheitsargumente und Nachweise für KI-Komponenten
- Die enge Verzahnung der Normen ermöglicht einen durchgängigen, sicheren Lebenszyklus für KI-Systeme im Fahrzeug



Neue Maßstäbe für KI-Sicherheit SAE Level 3-4 Fahrfunktionen



- **Erweiterte Sicherheitskonzepte für SAE Level 3-4**
- AI Safety Lifecycle für alle Entwicklungsphasen
- Systematisches Management von Unsicherheiten
- Verfeinerte KI-Sicherheitsanforderungen
- Begrenzung von KI-Output-Unzulänglichkeiten
- AI Safety Case als strukturierte Sicherheitsargumentation

- **Erweiterte Verifikation, Validierung & Sicherheitsanalysen**
- Spezifische Testmethoden für KI-Komponenten
- Integrationstests & Bewertung der Systemleistung
- Erweiterte virtuelle und physische Testverfahren
- Dedizierte Sicherheitsanalysen für KI-Systeme
- Sicherheitsanalysen basierend auf Testergebnissen

- **Kontinuierliche Betriebsmaßnahmen & Datenmanagement**
- Kontinuierliche Neubewertung der Sicherheitsargumente
- Technische Sicherheitsmaßnahmen während des Betriebs
- Strukturierte Felddatenerfassung & Feldrisikoauswertung
- Maßn. für KI-Nachtraining, Neuvalidierung und -bereitstellung
- Anforderungen an Datenqualität und Trainingsdaten

Deep Learning für autonomes Fahren

Grundlagen: Deep Learning im Automobilkontext

- Architektur und Funktionsweise neuronaler Netze
- Anwendungsbereiche im autonomen Fahrzeug

Wahrnehmungssysteme und KI

- Sensorfusion und Objekterkennung
- Herausforderungen der Umgebungsinterpretation

Entscheidungsfindung mit Deep Learning

- Von der Wahrnehmung zur Handlung
- Trainingsmethoden und Datenverwaltung

Von der Wahrnehmung zur Entscheidung

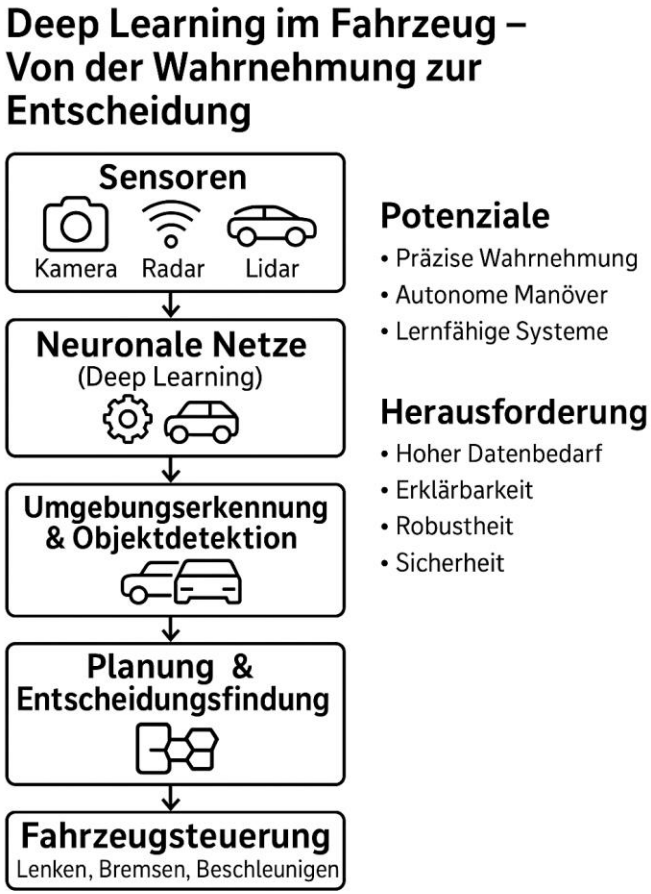
- Wahrnehmung und Objekterkennung
- Entscheidungsfindung und Handlungssteuerung

Grenzen und Ungewissheiten von Deep-Learning-Modellen

- Transparenz und Erklärbarkeit
- Umgang mit unbekannten Szenarien

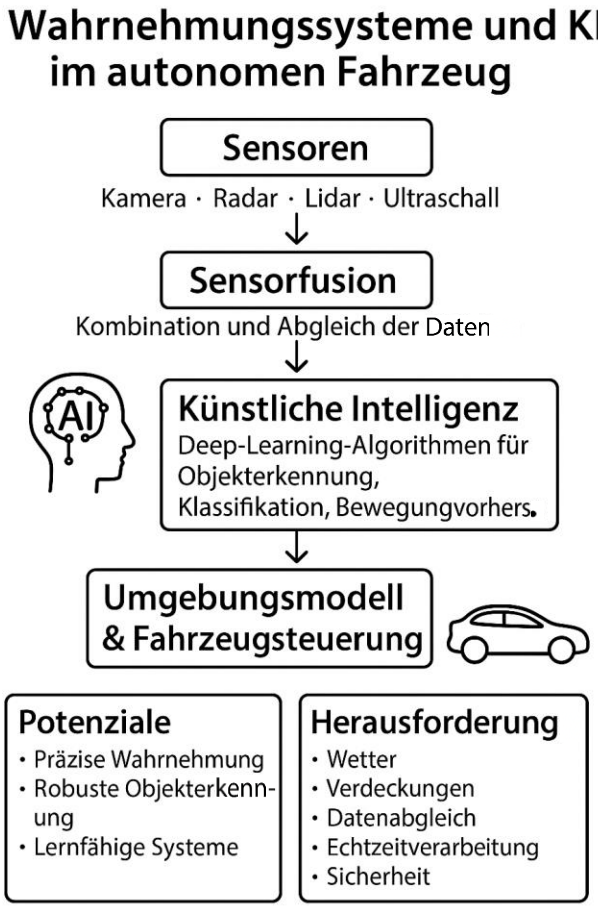
Grundlagen: Deep Learning im Automobilkontext

- Architektur und Funktionsweise neuronaler Netze
  - Neuronale Netze verarbeiten Daten durch Schichten (Input, Hidden, Output) mit gewichteten Verbindungen und Aktivierungsfunktionen.
- Anwendungsbereiche im autonomen Fahrzeug
  - Umgebungserkennung, Entscheidungsfindung, Spurhaltung, Objekterkennung und autonome Fahrmanöver mittels Sensordatenanalyse.
- Deep Learning ermöglicht autonome Fahrfunktionen – Sicherheit und Erklärbarkeit bleiben zentrale Herausforderungen



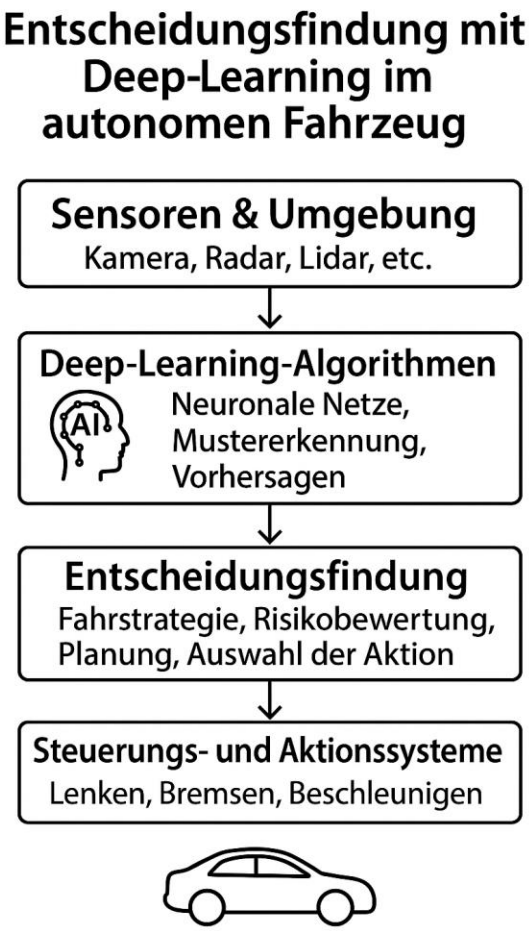
Wahrnehmungssysteme und KI

- Sensorfusion und Objekterkennung
  - Sensorfusion kombiniert Daten verschiedener Sensoren, verbessert Genauigkeit, Ausfallsicherheit und ermöglicht präzise Objekterkennung in Echtzeit
- Herausforderungen der Umgebungs-interpretation
  - Wetter, Verdeckungen, Datenabgleich und Echtzeitverarbeitung erschweren eine zuverlässige, umfassende Umgebungsinterpretation für autonome Fahrzeuge
- Sensorfusion und KI ermöglichen eine präzise, robuste Wahrnehmung und sichere Steuerung autonomer Fahrzeuge



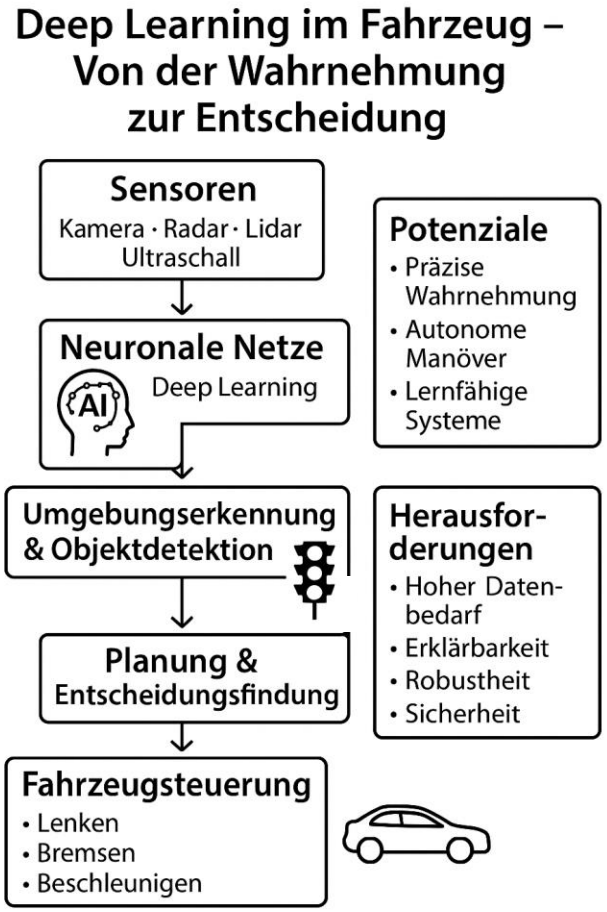
Entscheidungsfindung mit Deep Learning

- Von der Wahrnehmung zur Handlung
  - Sensordaten werden analysiert, interpretiert und in konkrete, sichere Fahrzeugaktionen umgesetzt
- Trainingsmethoden und Datenverwaltung
  - Große, vielfältige Datensätze und gezielte Trainingsmethoden sind essenziell für zuverlässige KI-Modelle
- Deep Learning ermöglicht autonome, situationsabhängige Entscheidungen und die sichere Umsetzung in Fahrzeugaktionen



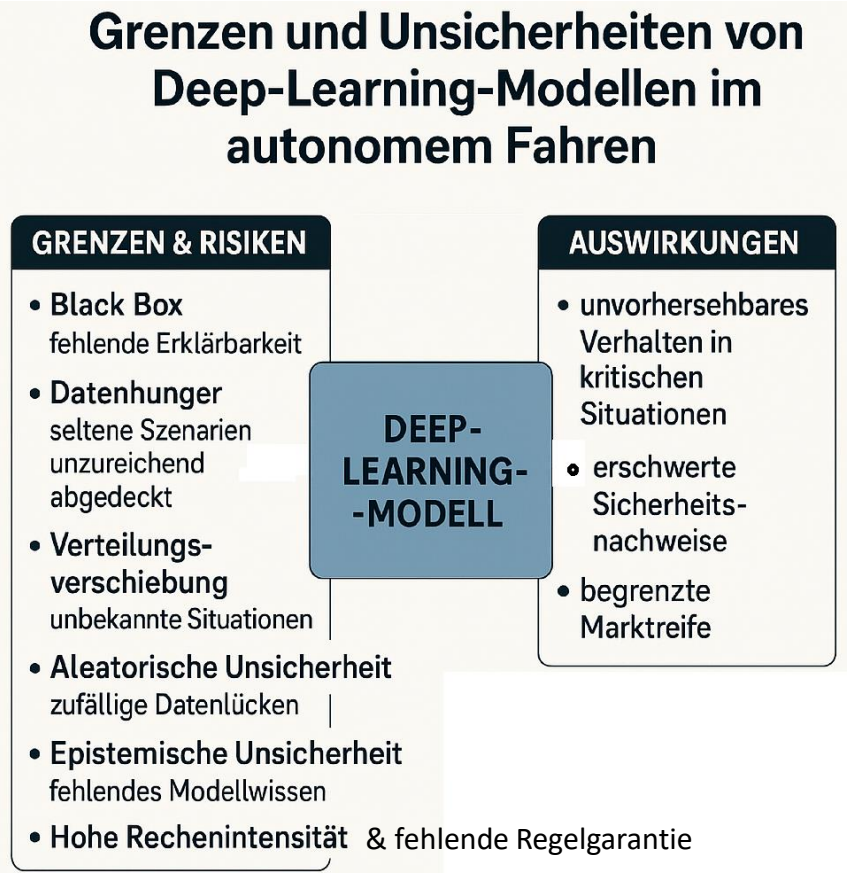
Von der Wahrnehmung zur Entscheidung

- Wahrnehmung und Objekterkennung:
  - Deep Learning erkennt zuverlässig Verkehrsobjekte, Fahrbahnmarkierungen und klassifiziert komplexe Umgebungen in Echtzeit.
- Entscheidungsfindung und Handlungssteuerung:
  - KI ermöglicht autonome Fahrmanöver durch Analyse, Planung und Umsetzung sicherer Fahrzeugaktionen
- Deep Learning ermöglicht autonome Fahrfunktionen – Sicherheit und Erklärbarkeit bleiben zentrale Herausforderungen

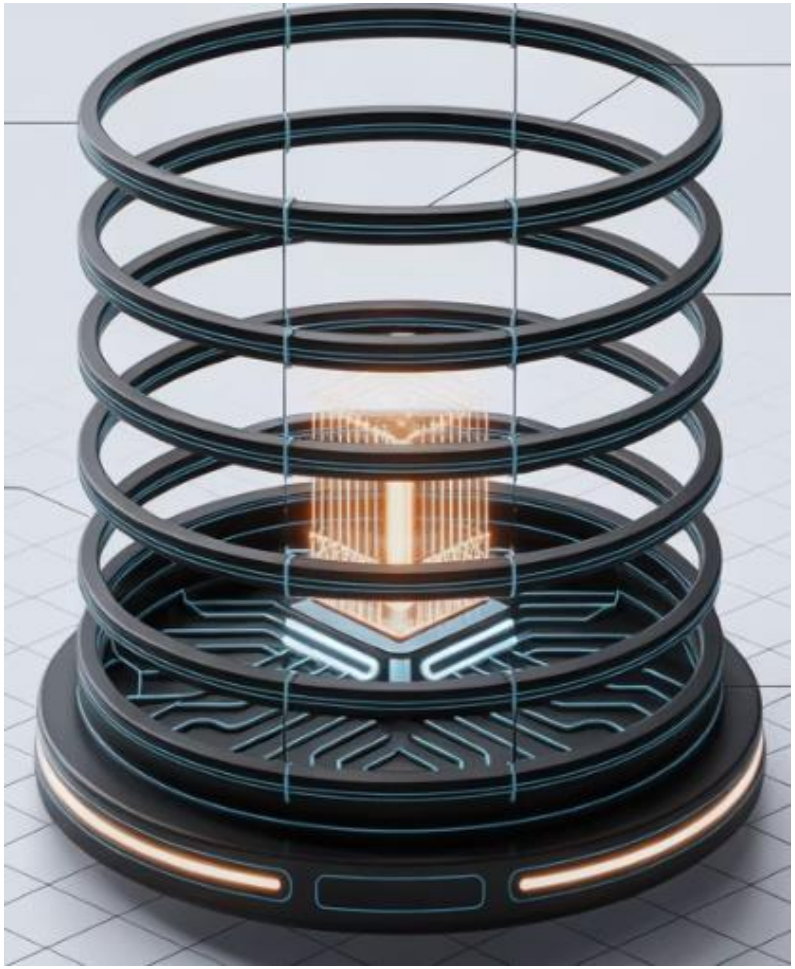


Grenzen und Ungewissheiten von Deep-Learning-Modellen

- Transparenz und Erklärbarkeit
  - Erklärbare KI erhöht Vertrauen, Nutzerakzeptanz und ist für Sicherheitsnachweise im Fahrzeug essenziell
- Umgang mit unbekannten Szenarien
  - KI-Systeme müssen unbekannte Situationen wahrnehmen, Unsicherheiten bewerten und sicher darauf reagieren
- Unsicherheiten, Datenlücken und fehlende Erklärbarkeit begrenzen die Sicherheit und Marktreife von Deep Learning im Fahrzeug



Safety Cage Architecture



Konzept der Safety Cage Architecture

- Definition und Komponenten
- Schutz vor unerwarteten KI-Verhaltensweisen

Virtualisierung und Monitoring

- Abgrenzung durch Virtualisierung
- Überwachung des KI-Systemverhaltens und Betriebszustände
- Erkennung von Abweichungen und Anomalien

Einschränkung der KI-Outputs

- Begrenzung auf zulässige Wertebereiche
- Beispiel: Sicherheitskorridore für Fahrzeugsteuerung

Redundanz-Konzepte

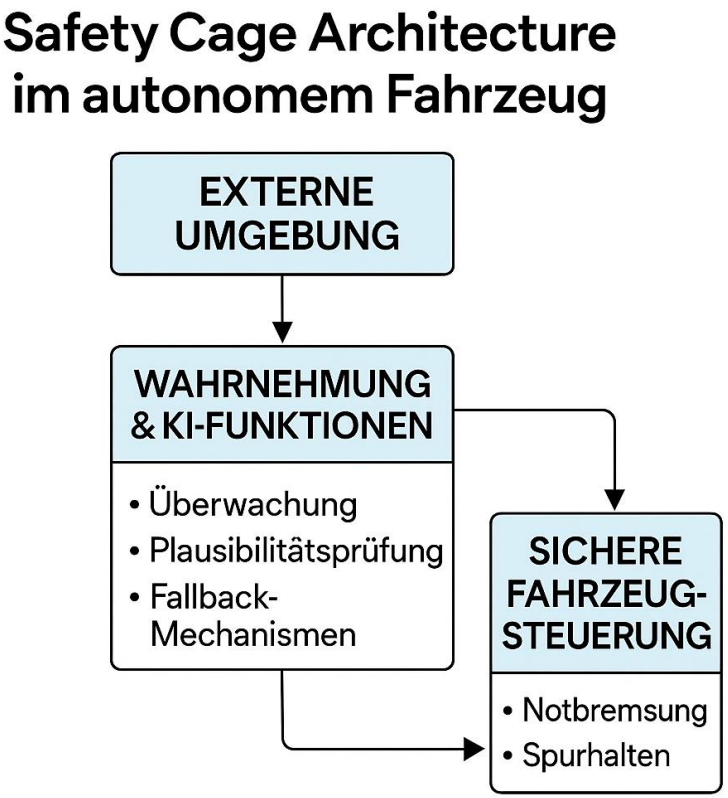
- Mehrfache Absicherung
- Sichere Degradation

Fallback-Strategien

- Übergänge in sichere Zustände
- Rückfallebenen und Fahrer-Übernahme

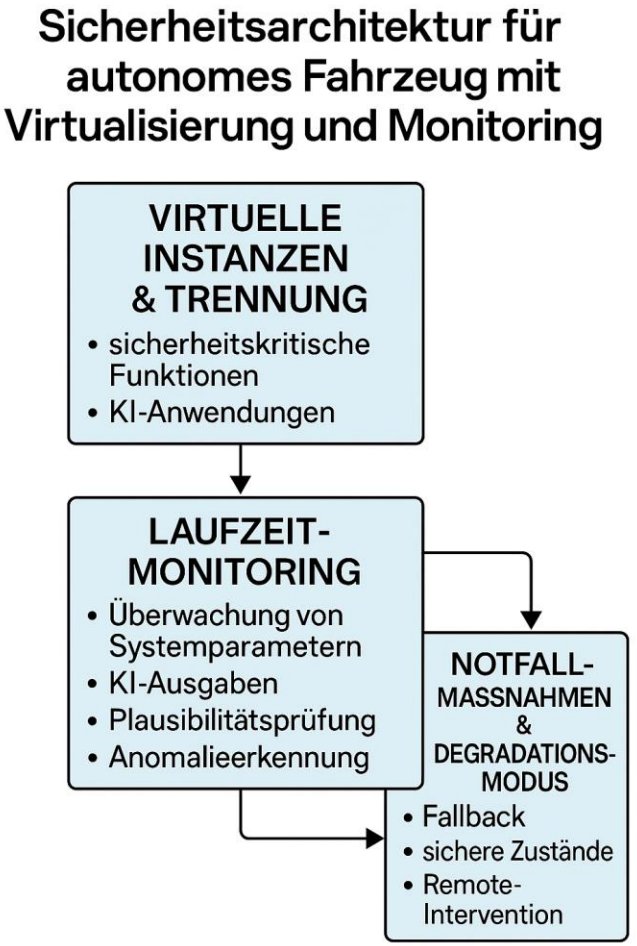
Konzept der Safety Cage Architecture

- Definition und Komponenten
  - Die Safety Cage Architecture umfasst Virtualisierung, Monitoring, Output-Restriktionen und Fallback-Strategien für KI-Systeme
- Schutz vor unerwarteten KI-Verhaltensweisen
  - Sie überwacht KI-Ausgaben und aktiviert bei Risiken sichere Zustände oder Notfallmaßnahmen
- Die Safety Cage Architecture überwacht KI-Entscheidungen und stellt bei Unsicherheiten den sicheren Fahrzeugbetrieb sicher



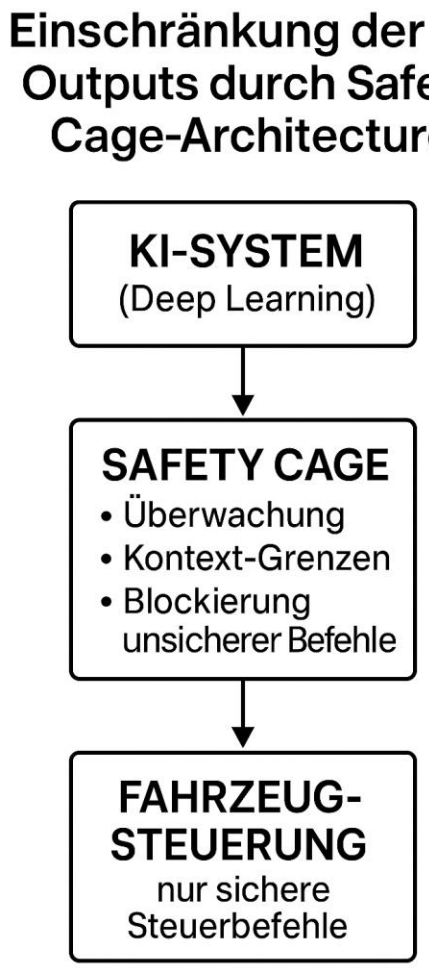
Virtualisierung und Monitoring

- Abgrenzung durch Virtualisierung
  - Virtualisierung trennt sicherheitskritische Funktionen von komplexen, KI-basierten Anwendungsfunktionen durch virtuelle Instanzen
- Überwachung des KI-Systemverhaltens und Betriebszustände
  - Das Monitoring prüft kontinuierlich das Verhalten und die Betriebszustände des KI-Systems auf Auffälligkeiten
- Erkennung von Abweichungen und Anomalien
  - Abweichungen und Anomalien werden erkannt, um rechtzeitig sichere Maßnahmen oder Fallbacks einzuleiten
- Virtualisierung und Monitoring verhindern Fehlerausbreitung und ermöglichen sicheres Eingreifen bei kritischen Situationen im autonomen Fahrzeug



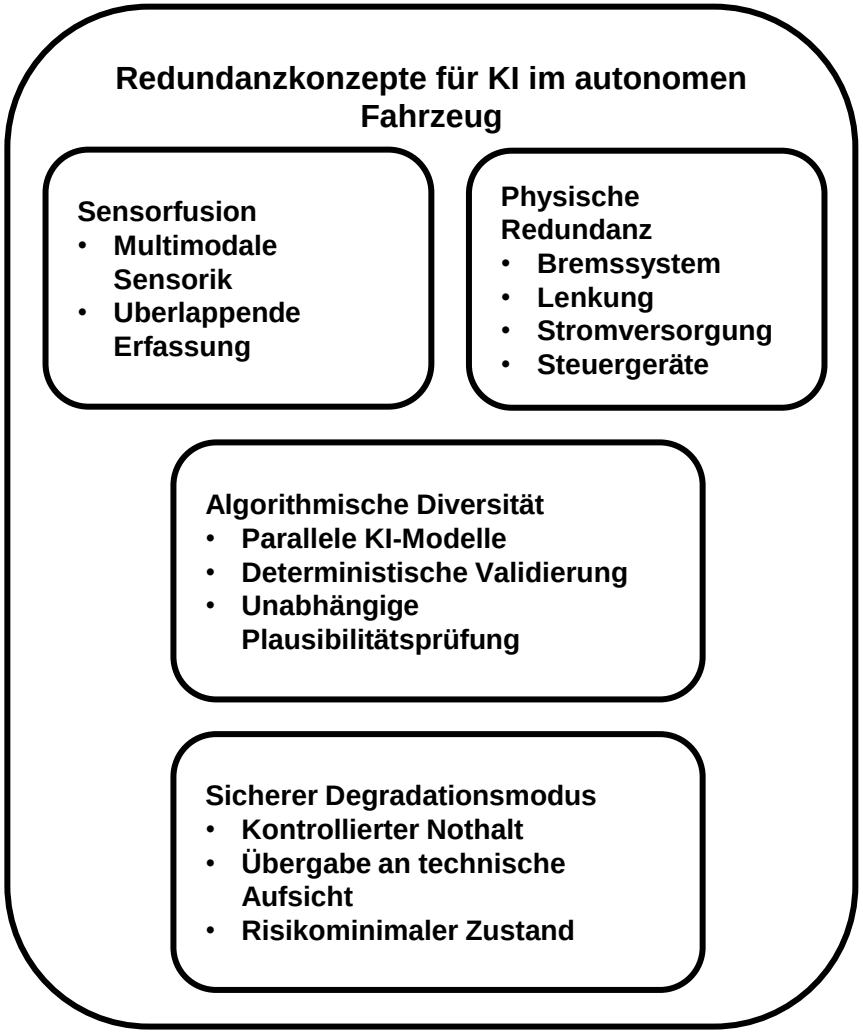
Einschränkung der KI-Outputs

- Begrenzung auf zulässige Wertebereiche
  - KI-Ausgaben werden auf vorher definierte, sichere Wertebereiche begrenzt und kontinuierlich überwacht
- Beispiel: Sicherheitskorridore für Fahrzeugsteuerung
  - Sicherheitskorridore stellen sicher, dass Lenk- und Bremsbefehle nur innerhalb sicherer Grenzen erfolgen
- Die Safety Cage Architecture begrenzt KI-Ausgaben auf sichere Werte und verhindert gefährliche Fahrzeugreaktionen in Echtzeit



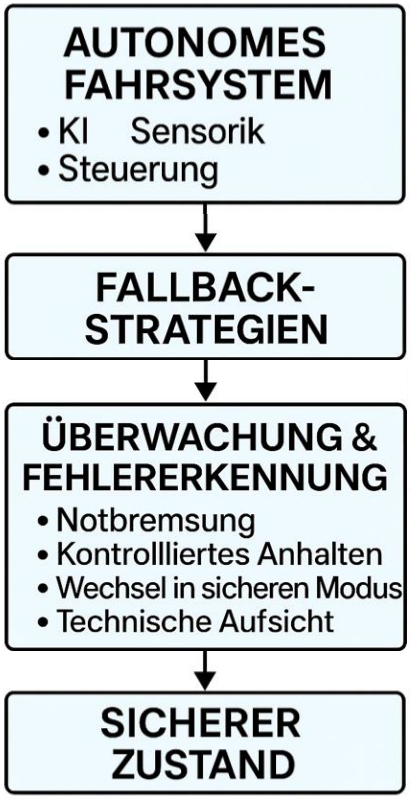
Redundanz-Konzepte

- Mehrfache Absicherung
  - Parallele KI-Modelle (z.B. CNN, RNN), deterministische Validierung und multimodale Sensorik verhindern Einzelpunktfehler
- Sichere Degradation
  - Kontrollierte Notfallstrategien und risikominimale Zustände gewährleisten Sicherheit auch bei Teilausfällen
- Mehrschichtige Redundanz bei KI-Systemen im Fahrzeug gewährleistet Sicherheit durch parallele Komponenten, diverse Algorithmen und sichere Degradationsmodi



Fallback-Strategien

- Übergänge in sichere Zustände
  - Das Fahrzeug führt bei Störungen automatisiert ein Minimum-Risk-Manöver oder sicheres Anhalten durch
- Rückfallebenen und Fahrer-Übernahme
  - Unabhängige Rückfallebenen sichern Fahrzeugsteuerung; Fahrerübernahme erfolgt nur bei klarer, rechtzeitiger Aufforderung
- Fallback-Strategien gewährleisten, dass autonome Fahrzeuge bei Fehlern oder Unsicherheiten sicher reagieren und Risiken minimieren



Praktische Umsetzung und Erfahrungen

Safety Case für KI-Systeme nach ISO/PAS 8800

- Strukturierte Sicherheitsargumentation
- Integration von Nachweisen und Evidenzen

Verifizierungs- und Validierungsmethoden

- Simulation und Testing
- Formale Methoden und deren Grenzen

Studien: Deep Learning im Automobil

- Projekterfahrungen und praktische Umsetzung
- Lessons Learned aus der Industrieanwendung

Automobile KI-Projekte

- Ganzheitliches Sicherheitskonzept von NVIDIA
- Sicherheitsmechanismen in der NVIDIA-Architektur

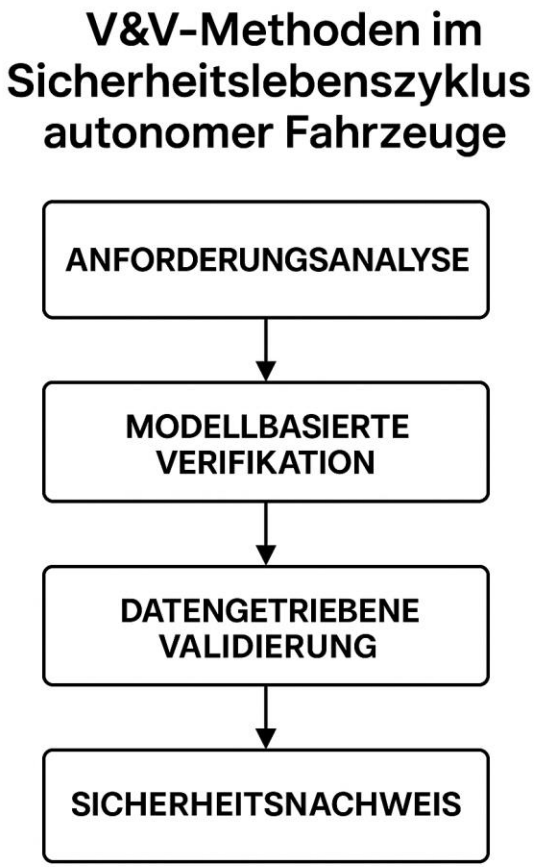
Safety Case für KI-Systeme nach ISO/PAS 8800

- Strukturierte Sicherheitsargumentation
  - Der AI Safety Case nach ISO/PAS 8800 bündelt Sicherheitsargumente systematisch in einem dokumentierten Gesamtnachweis.
- Integration von Nachweisen und Evidenzen
  - Simulationen, Tests und Felddaten werden konsolidiert, um die Sicherheit des KI-Systems nachzuweisen
- Der AI Safety Case verbindet Risikoanalyse, Sicherheitsmaßnahmen und Nachweisführung zu einem strukturierten Gesamtkonzept für KI-Systeme

- Datenqualität: Nachweis vollständiger, repräsentativer Trainingsdaten und Vermeidung von Bias.
- Runtime-Monitoring: Echtzeitüberwachung der KI-Ausgaben zur Erkennung von Anomalien.
- Lebenszyklusmanagement: Kontinuierliches Update des Safety Case durch Felddaten und Retraining.
- Transparenz: Dokumentation aller Unsicherheiten und Limitationen des KI-Modells

Verifizierungs- und Validierungsmethoden

- Simulation und Testing
  - Virtuelle Umgebungen und reale Tests validieren Sicherheit und Funktionalität von KI-Systemen
- Formale Methoden und deren Grenzen
  - Formale Verifikation sichert Korrektheit, stößt aber bei KI-Komplexität und Skalierbarkeit an Grenzen
- Verifikation und Validierung sichern KI-Systeme über den gesamten Entwicklungslebenszyklus hinweg ab – von der Simulation bis zur Zertifizierung
- KI-spezifische Risiken: Adversarial Attacks, Data Bias und Overfitting werden durch gezielte Teststrategien adressiert



Studien: Deep Learning im Automobil

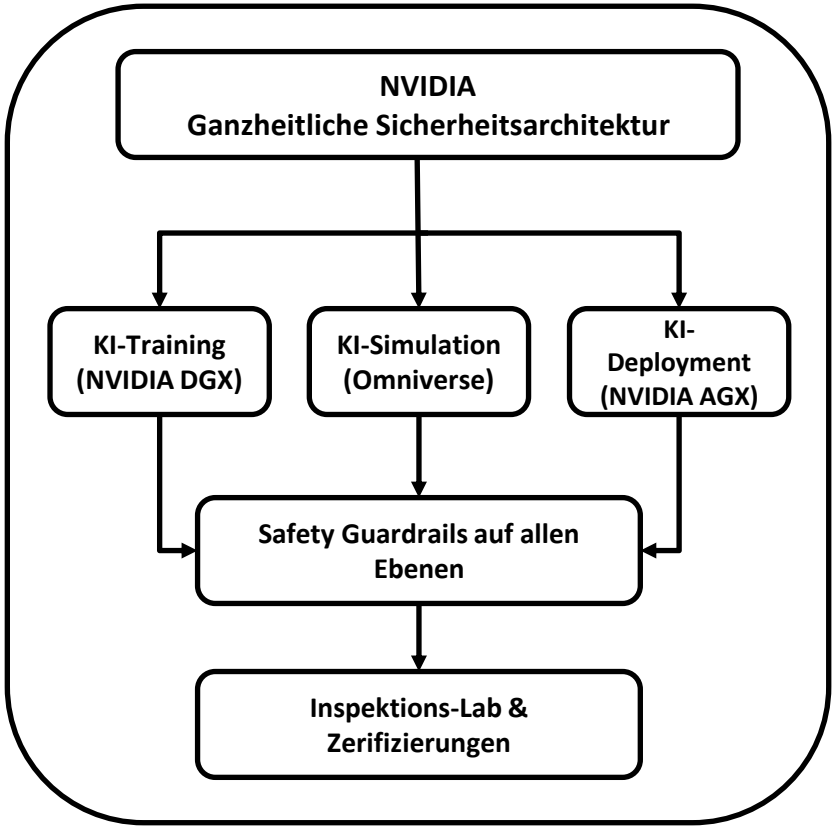
- Projekterfahrungen und praktische Umsetzung
  - → In Projekten zeigte sich: Deep Learning bewältigt komplexe Fahrsituationen, benötigt aber umfangreiche Tests.
  - → Die praktische Umsetzung erfordert enge Zusammenarbeit, leistungsfähige Hardware und kontinuierliche Sicherheitsüberwachung
- Lessons Learned aus der Industrieanwendung
  - → Die Skalierbarkeit von Tests und die Erklärbarkeit der KI bleiben zentrale Herausforderungen in Projekten.
  - → Kontinuierliche Verbesserung und enge Zusammenarbeit sind entscheidend für sichere, robuste Deep-Learning-Anwendungen im Fahrzeug
- Deep Learning ermöglicht autonome Fahrfunktionen – doch Sicherheits-architekturen wie die Safety Cage sind unverzichtbar, um Risiken zu begrenzen

Automobile KI-Projekte (Auswahl)

NVIDIA	Mobileye	Horizon /Baidu	Qualcomm	Eigene
• Toyota • Hyundai Kia • General Motors • Mercedes-Benz • Volvo • Kia • Tata Motors JLR • Lucid Motors • Rivian • Nio • BYD	• Toyota • Stellantis • Ford • Honda • Nissan • BMW • Renault • Geely • Subaru • Mahindra • Volkswagen • Holon	• SAIC • BYD • Changan • Geely • Dongfeng • FAW • GAC • Great Wall Motor	• BMW • Volkswagen • General Motors • Ford • SAIC	• Waymo • Cruise • Zoox • Pony.ai • WeRide • AutoX • Baidu Apollo
Plattformsicherheit - SoC - DriveOS - DL Inference - DL Accelerator - Virtualization Zertifizierungen (ASIL D) SFF Safety-Force-Field Redundanz & Fallback Cybersecurity Iterative Verbesserung - L4 fähig	Plattformsicherheit - Responsibility-Sensitive Safety auf EyeQ-SoC Safety Envelope - True Redundany - L2+/L3 fähig - L4 ODD fähig	Plattformsicherheit Horizon - sichere Hardware-/KI-Integration mit Upgradability Baidu - Rgelebasiert - RSS (Responsibility Sensitive Safety - Simulation - 5G Teleoperation - L2+/L3 fähig	Plattformsicherheit - SoC - AI Accelerator - Safety Virtualization AI Safety Guard Rails ASIL-D-Architektur Mixed Criticality Safety Island Zert. SW-Frameworks - L2+/L3 fähig	Robotaxis - Mehrschichtiger Sicherheitsansatz - Full-Stack-Entwicklung - massive Simulation - Flottenbetrieb im ODD - L4 fähig

Ganzheitliches Sicherheitskonzept von NVIDIA für Automotive AI

- NVIDIA Halos ist ein umfassendes, mehrschichtiges Sicherheitssystem für autonome Fahrzeuge, das Hardware, Software, KI-Modelle, Tools und Services von der Cloud bis ins Fahrzeug integriert.
- Sicherheitsleitplanken werden in allen Entwicklungsphasen gesetzt:
  - Design, Deployment und Validierung.
- Die Plattform basiert auf drei zentralen Säulen
  - KI-Training: NVIDIA DGX für sichere, bias-freie Modellentwicklung.
  - Simulation: NVIDIA Omniverse und Cosmos für realitätsnahe, skalierbare virtuelle Tests.
  - Deployment: NVIDIA AGX für sichere Echtzeitverarbeitung und Entscheidungsfindung im Fahrzeug.
- Algorithmische Sicherheit
  - KI-Modelle werden mit diversen, realen und synthetischen Daten trainiert, kontinuierlich überwacht und auf Unsicherheiten geprüft. Ein unabhängiger Stack läuft parallel als Fallback und für Redundanz.
- Inspektions-Lab
  - Das weltweit erste von ANAB akkreditierte AI Systems Inspection Lab prüft die Integration und Sicherheit von Partnerlösungen



Umsetzung & Sicherheitsmechanismen in der NVIDIA-Architektur

- Plattform-Sicherheit
  - Safety-assessed SoCs (z. B. DRIVE AGX Hyperion) mit hunderten integrierten Safety-Mechanismen und dem sicherheitszertifizierten Betriebssystem DriveOS.
- Redundanz & Fallback
  - Parallele, unabhängige Software-Stacks und Hardwarepfade sichern kritische Fahrfunktionen ab. Kontinuierliches Monitoring erkennt Anomalien und Unsicherheiten in Echtzeit.
- Cybersecurity
  - Zero-Trust-Architektur, Secure Boot, Hardware-Root-of-Trust und kontinuierliche Überwachung schützen vor Angriffen und Manipulationen.
- Zertifizierungen
  - Erfüllt höchste internationale Standards (ISO 26262 ASIL D, ISO/SAE 21434, UNECE), unabhängige Prüfungen durch TÜV SÜD und TÜV Rheinland.
- Iterative Verbesserung
  - Over-the-air-Updates, Datenflywheel und automatisierte Sicherheitsbewertungen sorgen für ständige Weiterentwicklung und Anpassung an neue Bedrohungen.

NVIDIA DRIVE AGX Platform

- **Safety-zertifiziertes SoC & DriveOS**
- **Parallele Software-Stacks & Hardwarepfade**
- **Echtzeit-Monitoring & Anomalieerkennung**
- **Zero-Trust Cybersecurity**
- **Zertifizierungen & unabhängige Prüfungen**
- **OTA-Updates & kontinuierliche Verbesserung**

Ausblick und Zusammenfassung

Zukünftige Herausforderungen und Forschungsfelder

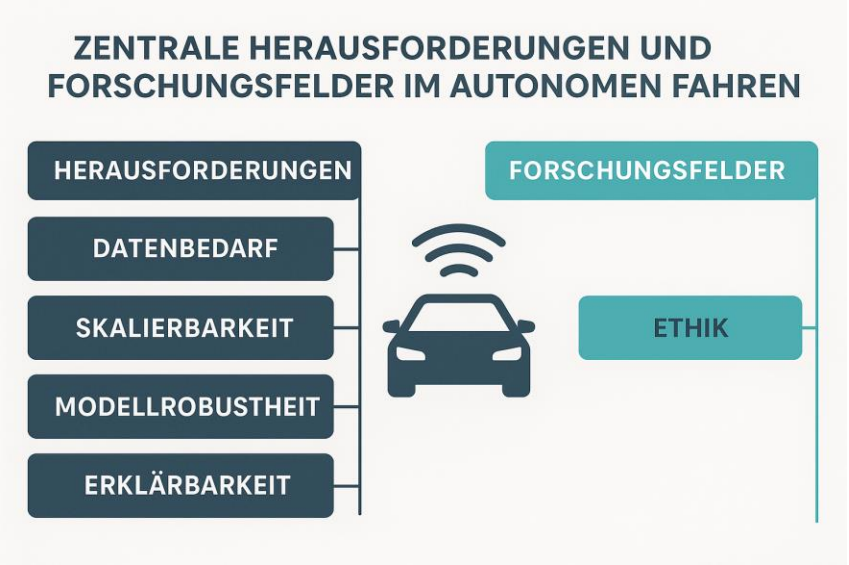
- Skalierbarkeit von Testmethoden
- Verbesserte Erklärbarkeit
- Qualifizierte Werkzeuge

Zusammenfassung und Schlüsselbotschaften

- Deep Learning sicher anwenden
- Normative Frameworks optimal nutzen
- Balance zwischen Innovation und Sicherheit finden

Zukünftige Herausforderungen und Forschungsfelder

- Skalierbarkeit von Testmethoden
  - Synthetische Daten, Simulationen und automatisierte Validierung ermöglichen skalierbare Tests für seltene und komplexe Szenarien
- Verbesserte Erklärbarkeit und qualifizierte Werkzeuge
  - Explainable AI, hybride Ansätze und zertifizierte Tools erhöhen Nachvollziehbarkeit, Vertrauen und Sicherheit von KI-Systemen
- Datenqualität, Skalierbarkeit, Erklärbarkeit und Ethik sind zentrale Forschungsfelder für die sichere Zukunft autonomer Fahrzeuge



Zusammenfassung und Schlüsselbotschaften

- Deep Learning sicher anwenden
  - Nur mit Safety-Architektur und Risikoanalyse ist Deep Learning im Fahrzeug verantwortbar
- Normative Frameworks optimal nutzen
  - ISO 26262, ISO 21448 und ISO/PAS 8800 bieten klare Leitlinien für sichere KI-Anwendungen
- Balance zwischen Innovation und Sicherheit finden
  - Innovative KI-Lösungen müssen stets mit funktionaler Sicherheit und Transparenz kombiniert werden

Weiterentwicklung zuverlässiger KI

VIELEN DANK FÜR IHRE AUFMERKSAMKEIT

FRAGEN?

Contact

FSQ Experts
A Brand of Wertefest GmbH
Balanstraße 14
81669 München
+49 89 588087571
info@fsq-experts.com
www.fsq-experts.com

MCG
Management Consult GmbH
Waldstraße 25
30890 Barsinghausen
+49 5105 585919
mcg@mcgnet.de
www.mcgnet.de