KI ABSICHERUNG

Safe AI for Automated Driving

Automobilindustrie digital, 02. März 2021

AI Land Meets Safety Land

Dr. Stephan Scholz, Volkswagen AG


Nachweisbar sichere
KI-basierte Funktionsmodule
KI ABSICHERUNG
Safe AI for Automated Driving
Fußgängererkennung

Herausforderung

AI Land

Verheißungsvolle, neue Technologie mit ungeahnten Möglichkeiten

Etablierte Absicherungsprozesse nicht übertragbar

KI ABSICHERUNG
Safe AI for Automated Driving

Safety Land

Sichere, vertrauensvolle Fahrfunktion

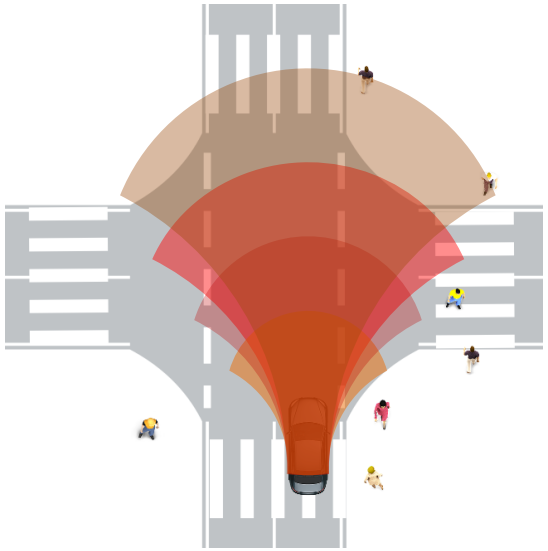
Übergreifender Konsens (Safe AI)
Absicherungsmethodik für sicherheitsrelevante KI-Module

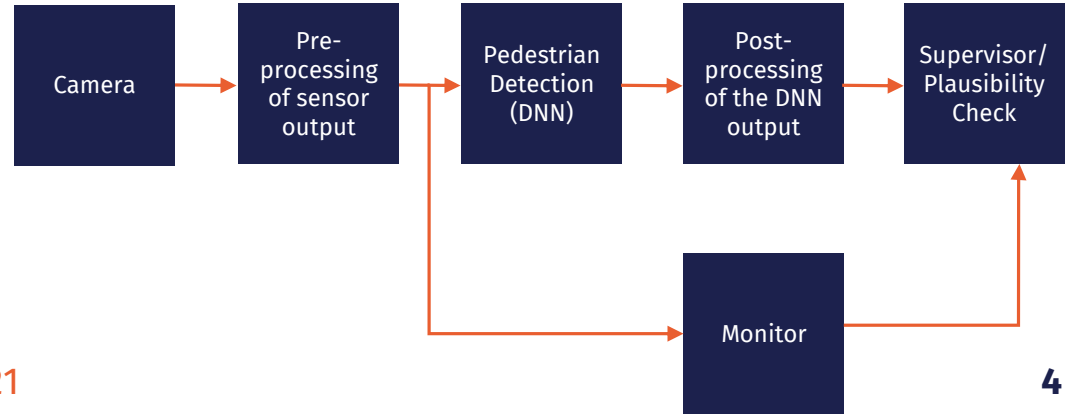
Spezifikation

Funktion

Rahmenbedingung

Systemarchitektur



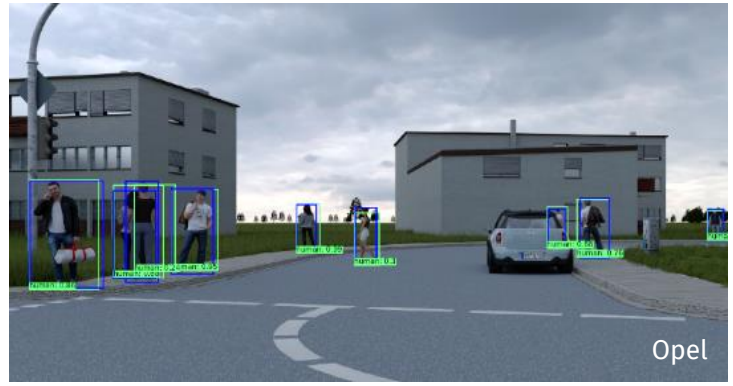


KI Modul: Fußgängererkennung

- Funktion
- Rahmenbedingung
- Systemarchitektur
- (KI-)Modul

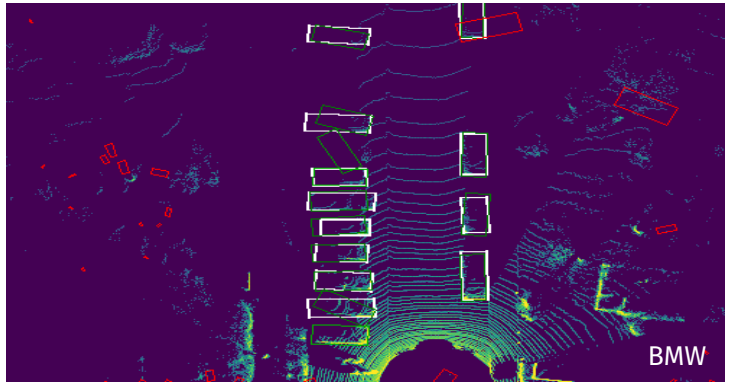
Semantische Segmentierung

2D-Bounding Box Detektion

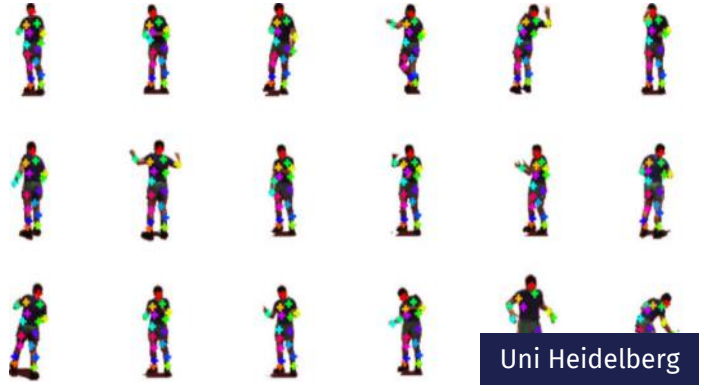


Instance Segmentierung

3D Bounding Box Detektion



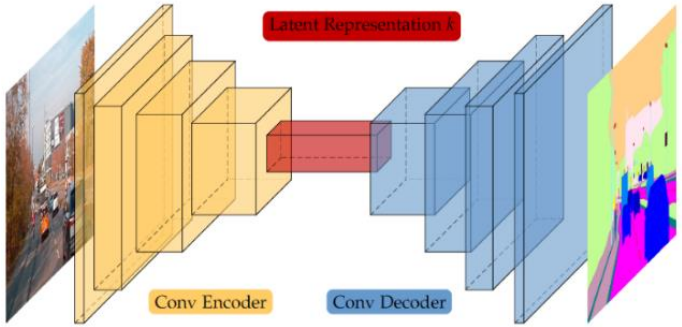
3D - Posenerkennung



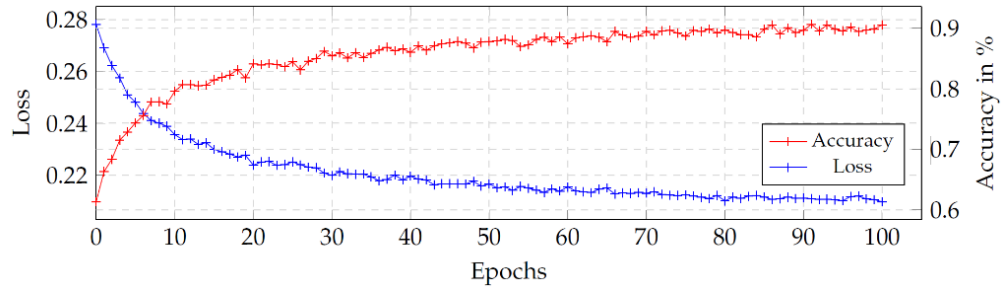
ML-Lifecycle

ML LIFECYCLE

- ML-Daten
- Modeldesign
- Pre-Processing
- Training
- Post-Processing
- Test/ V&V inkl. Daten
- Monitoring
- Maintenance



Volkswagen AG



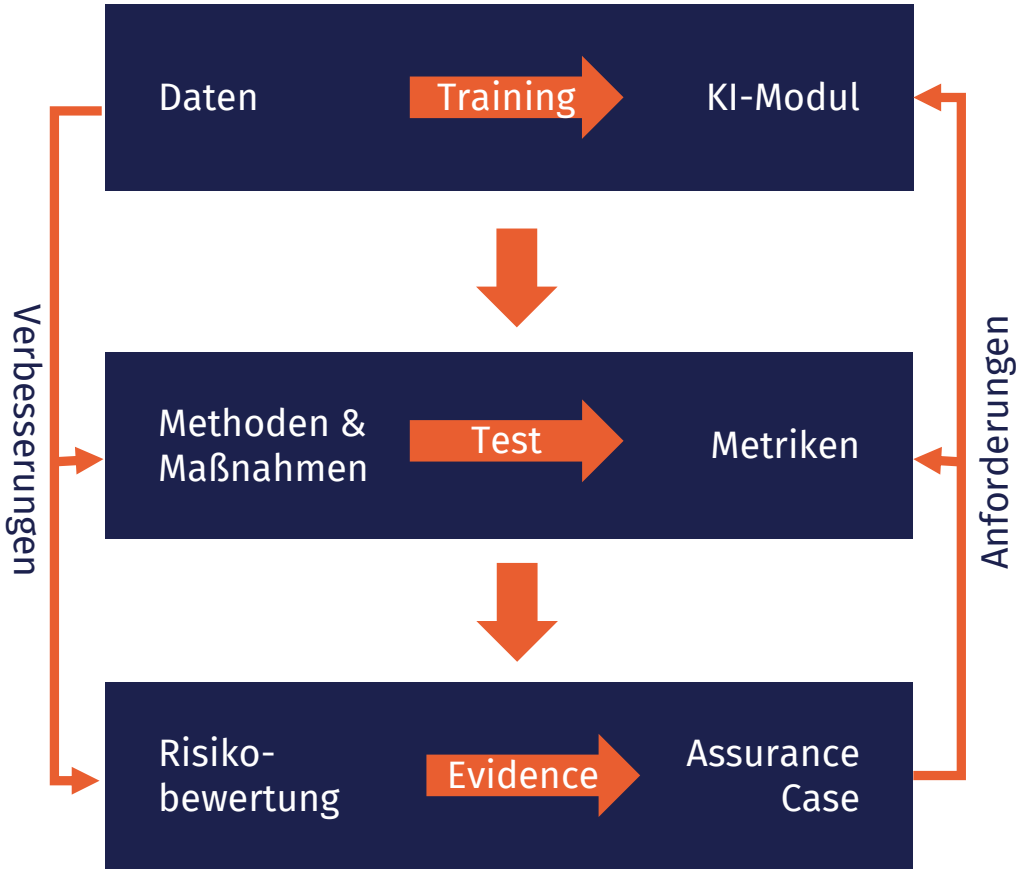
Volkswagen AG

ML-Lifecycle-Absicherungsdaten

ML LIFECYCLE

- ML-Daten
- Modelldesign
- Pre-Processing
- Training
- Post-Processing
- **Test/ V&V inkl. Daten**
- Monitoring
- Maintenance

Stringente Sicherheitsargumentation



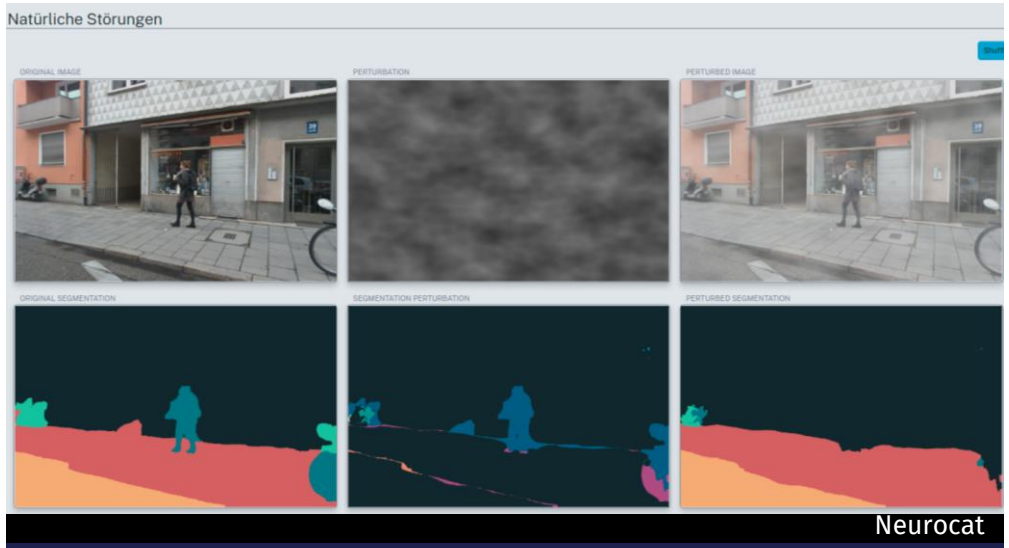
DNN-spezifische „Safety Concerns” erkennen, messen & entgegenwirken

Safety Concern:

- False positive / negative: Fußgängererkennung ist fehlerhaft bzw. nicht robust genug

Methode:

- Unsicherheitsbewertung: Stochastische Auswertung einer Vielzahl von Modellveränderungen (Monte Carlo Dropout)



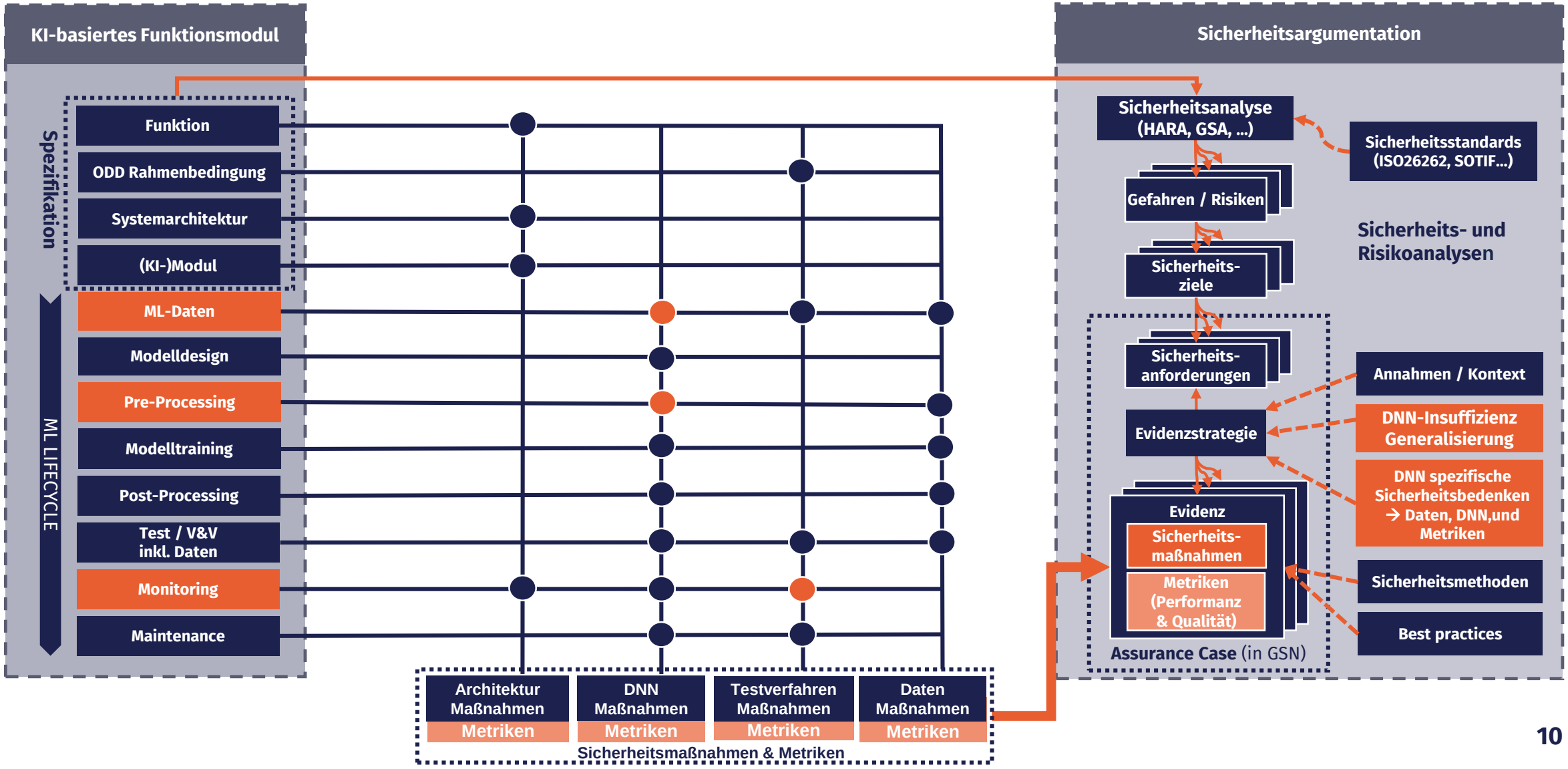
Safety Concern:

- Adversiale Störungen: führt zum Übersehen (false negative)

Methode:

- Systematische Analyse von adversialen Störungen
- Methoden zur Bewertung der adversialen Resilienz
- Abwehrmechanismen im Training oder zur Laufzeit

KI spezifische evidenzbasierte Sicherheitsargumentation



KI ABSICHERUNG

Safe AI for Automated Driving

Projektkoordinator: Dr. Stephan Scholz | Volkswagen AG

Stellv. Konsortialleiter und wissenschaftlicher Koordinator: PD. Dr. Michael Mock | Fraunhofer IAIS

E-Mail: ki-absicherung-konsortialfuehrung@eict.de

KI Absicherung ist ein Projekt der KI Familie
und wurde aus der VDA Leitinitiative autonomes
und vernetztes Fahren heraus entwickelt.

KI FAMILIE

Gefördert durch:

Bundesministerium für Wirtschaft und Energie

aufgrund eines Beschlusses des Deutschen Bundestages

www.ki-absicherung.vdali.de @KI_Familie KI Familie