

Zukunft der Robotik aus Sicht der Wissenschaft

SVLFG-Fachkolloquium:

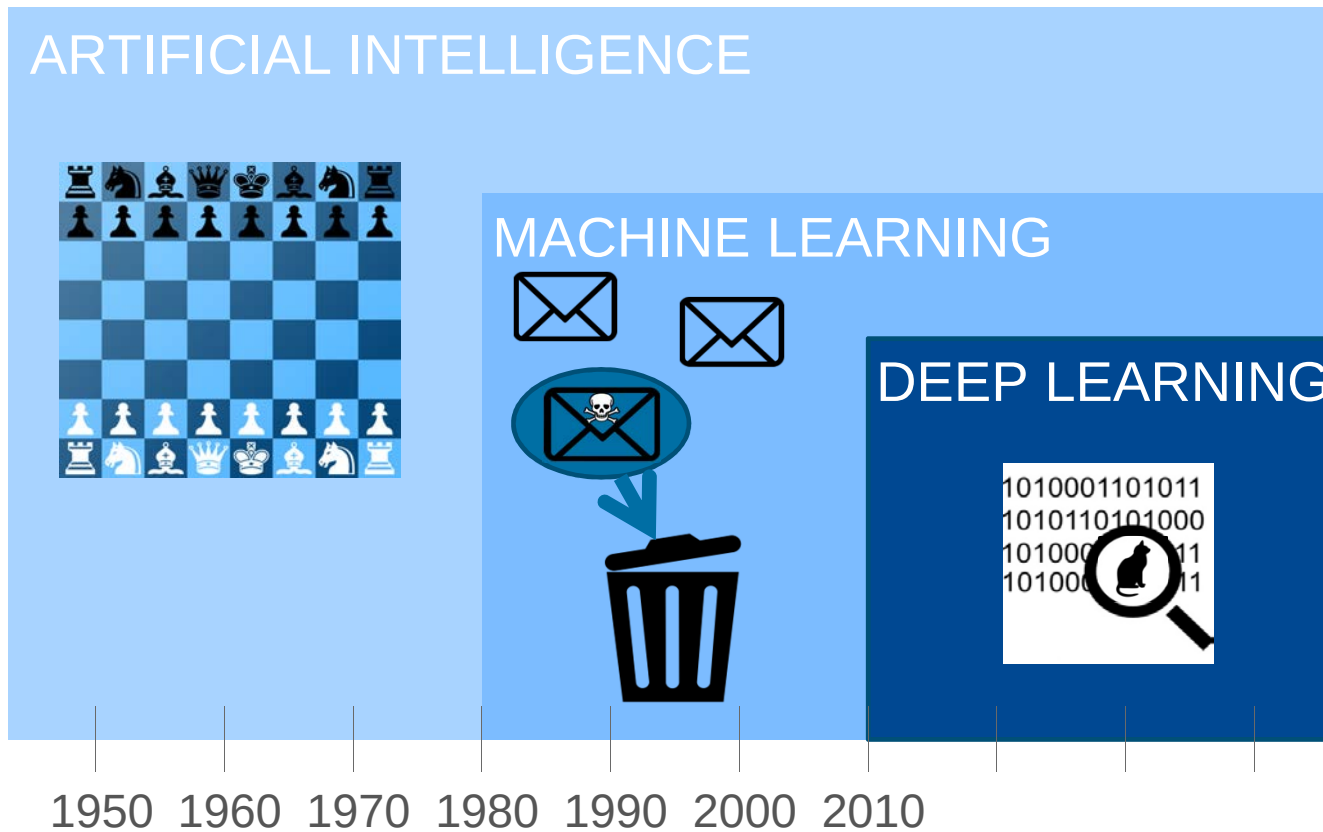
Robotik in der Landwirtschaft – Sicher in die Zukunft

18.07.2023

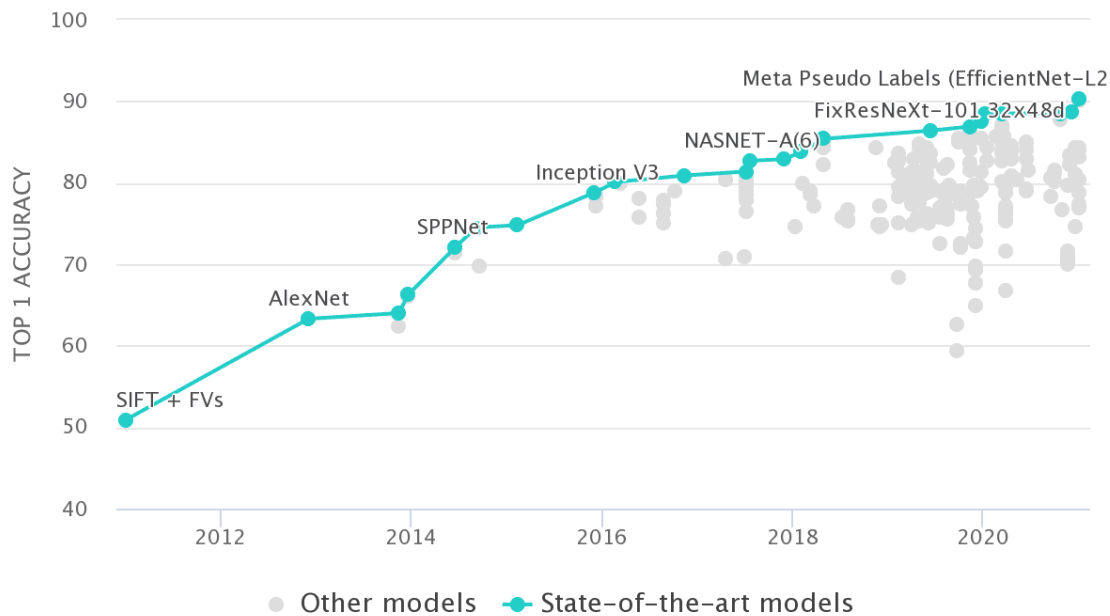
Prof. Dr. André Steimers

Hochschule Koblenz – RheinAhrCampus Remagen

Übersicht

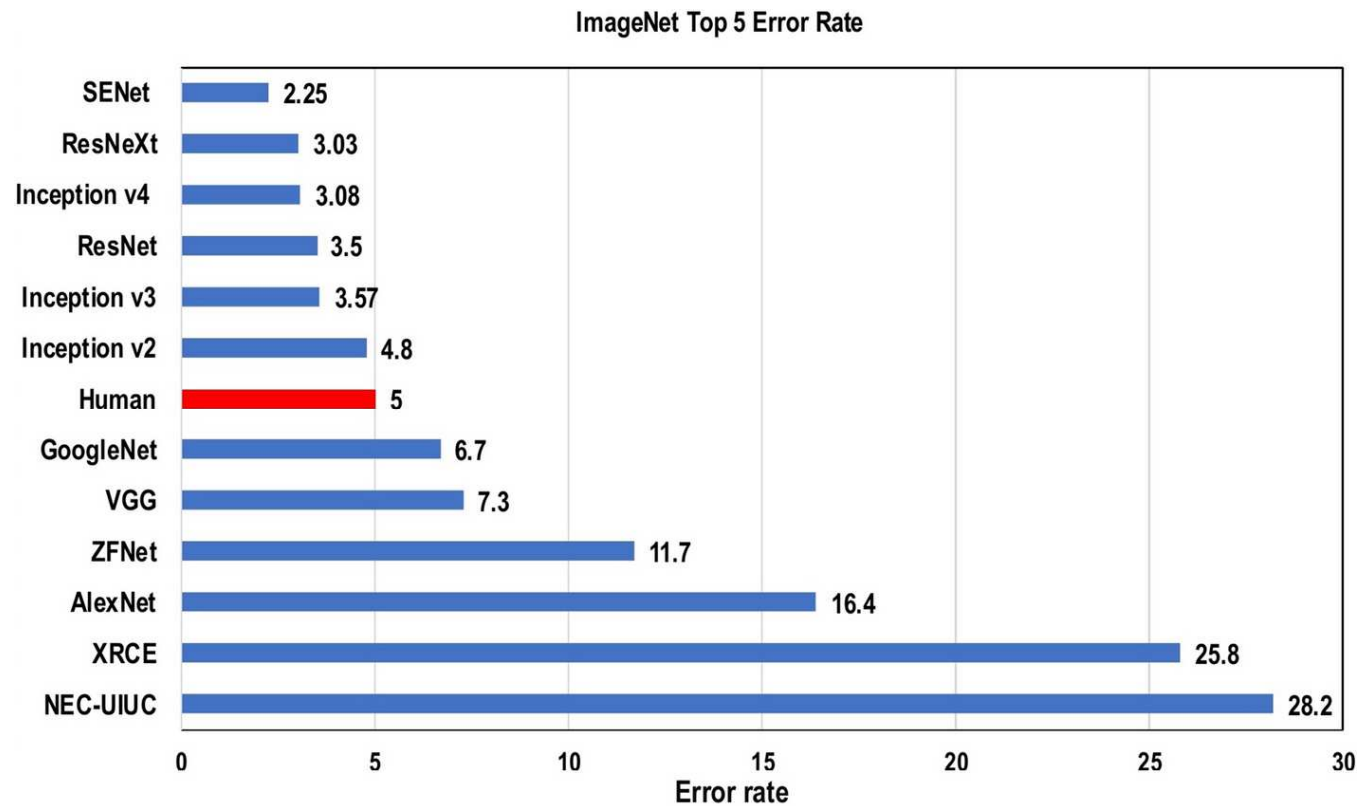


Bildverarbeitung: Klassifikation



- ImageNet
 - >14 Millionen beschriebene Bilder
 - >1 Millionen annotierte Bilder (box)
 - >20.000 Kategorien
 - >500 Bilder pro Kategorie
- Model soups / ViT-G/14 (2022):
 - 90,94 % top-1 accuracy
- Meta Pseudo Labels/EfficientNet-L2 (2021):
 - 90,2 % top-1 accuracy
 - 98,8 % top-5 accuracy

Bildverarbeitung: Klassifikation



Anwendung: Selbstfahrende Fahrzeuge

1. Öffentlich zugängliche Bereiche

- Öffentliche für jedermann zugängliche Bereiche
- Betriebliche Bereiche mit schwacher Zugangssicherung wie Schranken oder Ampeln

2. Abgeschlossene Bereiche mit begrenztem Zugang

- Betriebliche Bereiche, zu denen nur berechtigte Personen oder Fahrzeuge Zugang haben
- Personen angemessen unterwiesen oder qualifiziert

3. Abgeschlossene Bereiche ohne Zutritt

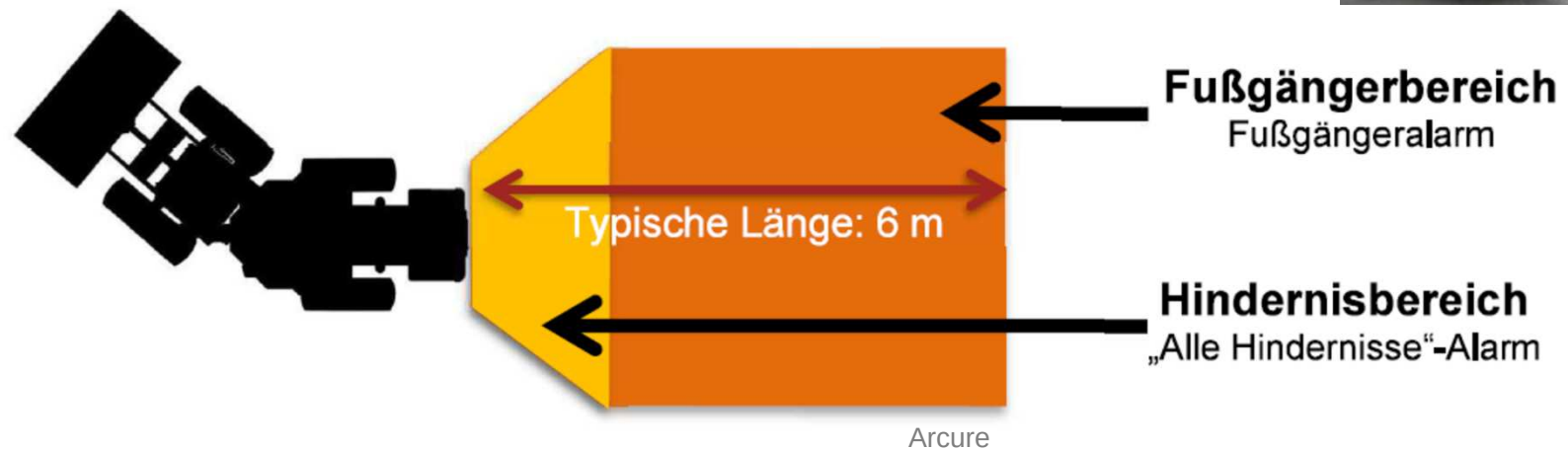
- Betriebliche Bereiche mit automatisiertem Fahrzeugverkehr, zu denen Personen im Regelfall keinen Zugang haben
- Bei Zugang (Instandhaltung), sind Maßnahmen umzusetzen (z.B. Stillstand, sichere Bereiche)



Anwendung: Rückfahrsistenzsysteme

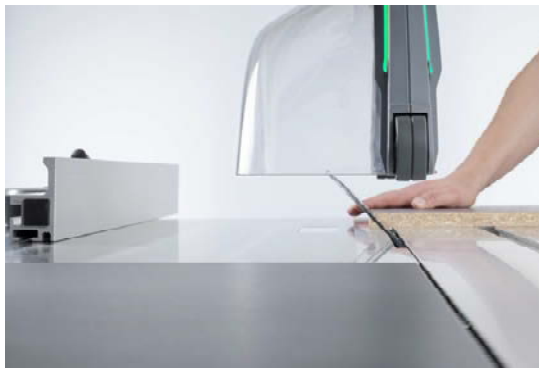
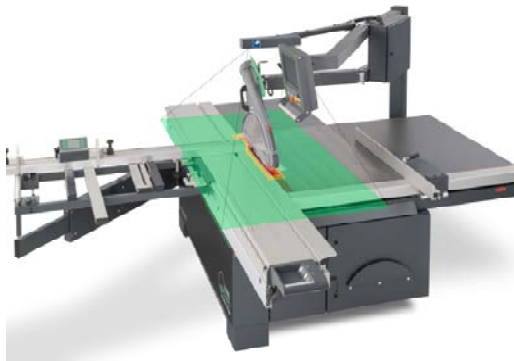
Antikollisionskamera für Fußgänger und Fahrzeuge

- intelligente Kamera, die in Echtzeit eine Person von einem anderen Hindernis unterscheiden und den Fahrer im Gefahrenfall warnen kann



Anwendung: Assistenzsystem für Formatkreissägen

Assistenzsystem das die Bewegungen der Hände eines Operators erkennt und die Maschine, in Abhängigkeit von deren Bewegungstrajektorie, in einen sichereren Zustand versetzt.



Altendorf

Beispiele einiger KI Fehler

- 07/2016 Wachroboter verletzt Kind in Kaufhaus
- 11/2016 Roboter Xiao-Pang verletzt Messebesucher
- 05/2017 Auffahrunfall Tesla Model S Feuerwehrfahrzeug
- 01/2018 Auffahrunfall Tesla Model S Feuerwehrfahrzeug
- 05/2018 Auffahrunfall Tesla Model S Polizeifahrzeug
- 01/2019 Kollision mit Gegenverkehr Tesla Model 3
- 08/2019 Auffahrunfall Tesla Modell S Abschleppwagen
- 12/2020 Fehlfunktion eines Serviceroboters im Kaufhaus

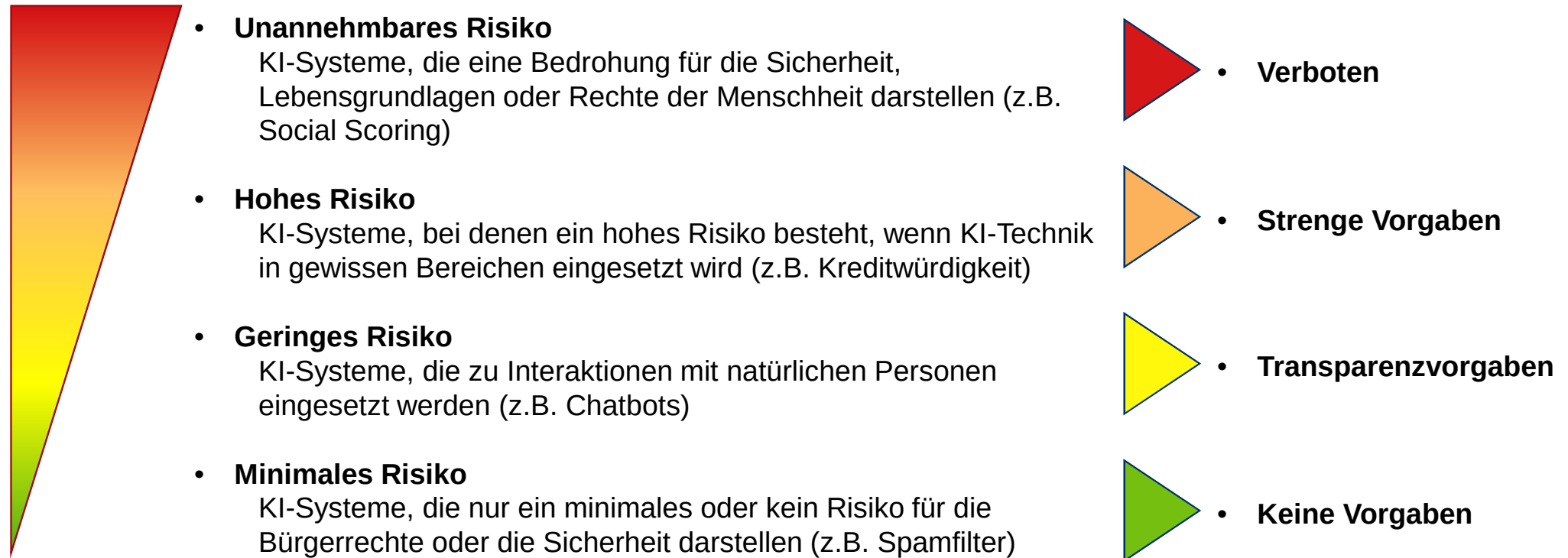


Beispiele einiger KI Fehler

- 01/2016 China, Tesla Model S, 1 Fahrer[†]
- 05/2016 Florida, Tesla Model S, 1 Fahrer[†]
- 03/2018 Arizona, automatisiertes Uber Taxi, 1 Fußgängerin[†]
- 03/2018 Kalifornien, Tesla Model X, 1 Fahrer[†]
- 04/2018 Japan, Tesla Model X, 1 Fußgänger[†]
- 03/2019 Florida, Tesla Model 3, 1 Fahrer[†]
- 04/2019 Florida, Tesla Model S, 1 Fußgänger[†]
- 12/2020 Kalifornien, Tesla Model S, 2 Personen Honda Civic[†]
- 05/2020 Norwegen, Tesla Model X, 1 Fußgänger[†]
- 04/2021 Texas, Tesla Model X, 2 Insassen[†]



Entwurf einer KI-Verordnung der Europäischen Kommission



Entwurf einer KI-Verordnung der Europäischen Kommission

KI-Anwendungsbereiche mit hohem Risiko

Kritische Infrastrukturen

- KI-Systeme, durch die das Leben und die Gesundheit gefährdet werden kann (z.B. im Verkehr)

Schul- und Berufsausbildung

- KI-Systeme, durch die der Zugang zur Bildung und zum Berufsleben beeinträchtigt werden kann (z. B. Bewertung von Prüfungen)

Sicherheitskomponenten von Produkten

- KI-Systeme, durch die die Sicherheit von Produkten beeinträchtigt werden kann (z.B. in der robotisierten Chirurgie)

Beschäftigung, Personalmanagement und Zugang zu selbstständiger Tätigkeit

- KI-Systeme, durch die die Chancengleichheit beeinträchtigt werden kann (z.B. Auswertung von Lebensläufen)

Wichtige private und öffentliche Dienstleistungen

- KI-Systeme, durch die der Zugang zu wichtigen Dienstleistungen beeinträchtigt werden kann (z.B. Bewertung der Kreditwürdigkeit)

Strafverfolgung

- KI-Systeme, die in die Grundrechte der Menschen eingreifen können (z.B. Bewertung der Verlässlichkeit von Beweismitteln)

Migration, Asyl und Grenzkontrolle

- KI-Systeme, die in die Bewegungsfreiheit der Menschen eingreifen können (z.B. Überprüfung von Reisedokumenten)

Rechtspflege und demokratische Prozesse

- KI-Systeme, durch die das Prinzip des Rechtsstaats beeinträchtigt werden kann (z.B. Anwendung von Rechtsvorschriften)

Entwurf einer KI-Verordnung der Europäischen Kommission

Vorgaben für KI-Anwendungen mit **hohem Risiko**

1. Angemessene **Risikobewertungs- und Risikominderungssysteme**
2. Hohe **Qualität der Datensätze**, die in das System eingespeist werden
3. Technische **Dokumentation**
4. Ausführliche **Aufzeichnungen** mit allen erforderlichen Informationen über das System und seinen Zweck
5. Klare und angemessene **Informationen für die Nutzer**
6. Angemessene **menschliche Aufsicht**
7. Hohes Maß an **Robustheit, Sicherheit und Genauigkeit**



Trustworthiness Analysis: TR 5469 Risk Factors

- **Vertrauenswürdigkeit der KI**

- Abhängig von den Risikoquellen des gewählten KI-Verfahrens
- Siehe: ISO/IEC WD TR 5496 AI – Functional Safety and AI-based system (Clause 8)

1. Fairness
2. Privacy
3. Automatisierungsgrad und Kontrolle

} Ethische Aspekte

4. Komplexität der Aufgabe und Verwendungsumgebung
5. Grad der Transparenz und Erklärbarkeit
6. Security
7. System-Hardware
8. Technologische Ausgereiftheit

} Zuverlässigkeit und Robustheit

1. Fairness



Quelle: www.jobrapido.com



Quelle: www.embedica.ai/fairness-2



Quelle: Joy Buolamwini, M.I.T. Media Lab

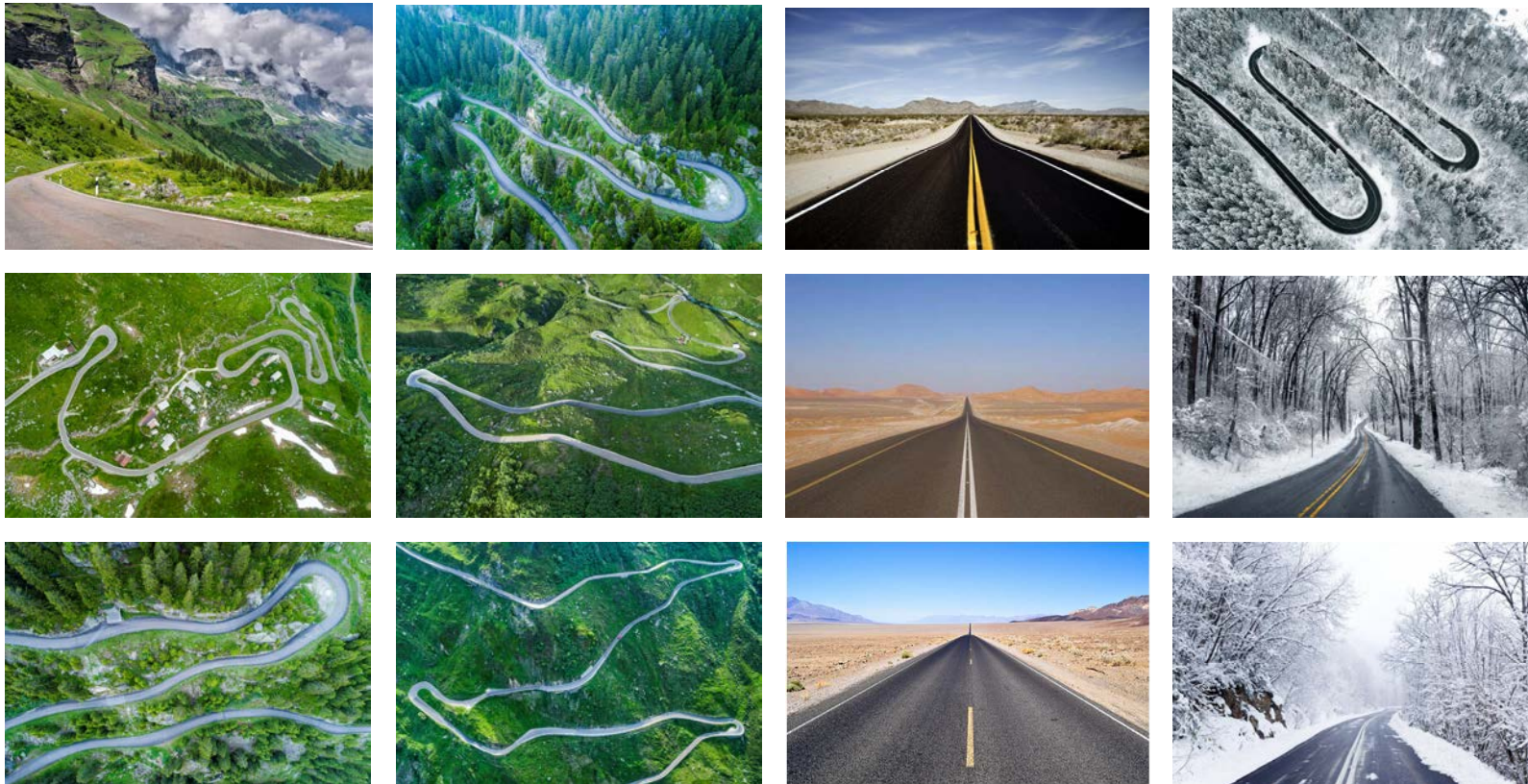
- **Rekrutierungs-Tool**
diskriminiert Frauen
- **Historischer Bias**
ML-Modell kann negative Korrelation lernen, da Männer in der Vergangenheit oft systematisch bevorzugt wurden
- **Gesichtserkennung**
schlechtere Performance bei farbigen Menschen
- **Daten Bias**
Unterrepräsentierte Gruppen in den Trainingsdaten führen zu höheren Fehlerraten dieser Gruppen im ML-Modell

3. Grad der Automatisierung und Kontrolle

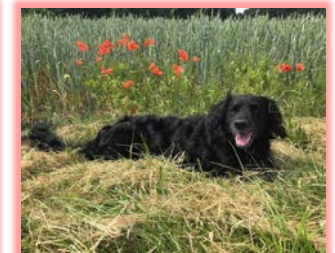
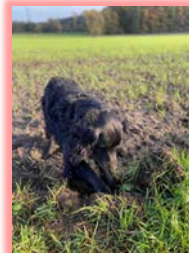
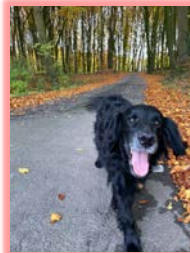
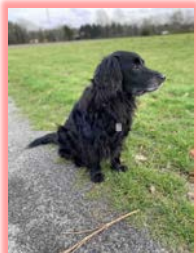
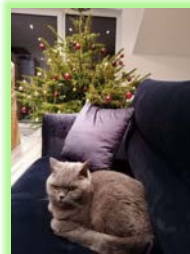
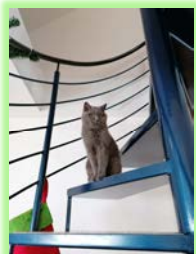
System	Grad der Automatisierung	Grad der Kontrolle	Kommentare
Autonomie	Autonom	Human out of the loop	Das System ist in der Lage, seinen Betriebsbereich oder seine Ziele ohne Eingreifen, Kontrolle oder Aufsicht von außen zu ändern.
Heteronomie	Vollautomatisiert	Human on the loop Human out of the loop	Das System ist in der Lage, seine gesamte Aufgabe ohne Eingriffe von außen zu erfüllen.
	Hochautomatisiert	Human on the loop	Das System führt Teile seiner Aufgaben ohne externen Eingriff aus.
	Bedingte Automatisierung	Human on the loop	Nachhaltige und spezifische Leistung eines Systems, wobei ein externer Agent bereit ist, bei Bedarf einzuspringen.
	Teilautomatisierung	Human in the loop	Einige Teilfunktionen des Systems sind vollständig automatisiert, während das System unter der Kontrolle eines externen Agenten bleibt.
	Assistenzfunktion	Human in the loop	Das System unterstützt den Bediener.
	Keine Automatisierung	Human in the loop	Der Bediener hat die volle Kontrolle über das System.



4. Komplexität der Aufgabe und Verwendungsumgebung

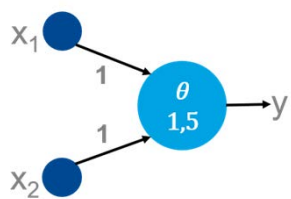


5. Grad der Transparenz und Erklärbarkeit

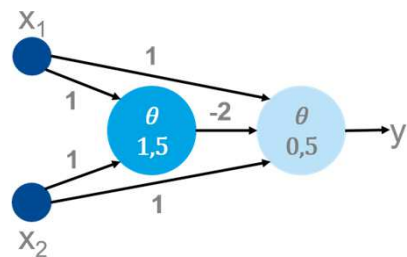


5. Grad der Transparenz und Erklärbarkeit

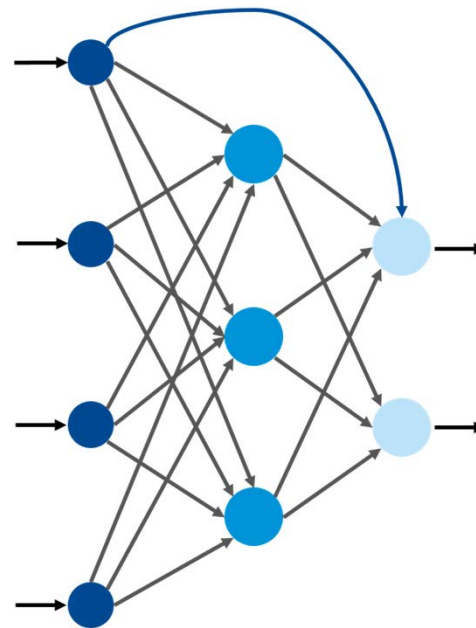
Inception Net V3



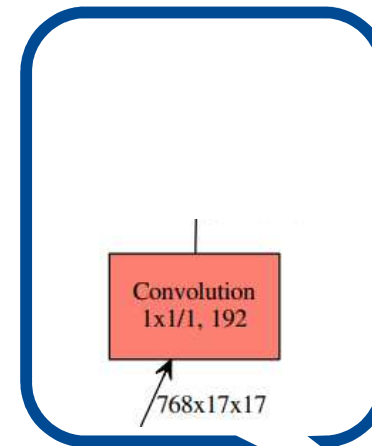
1 Neuron



2 Neurone



5 Neurone



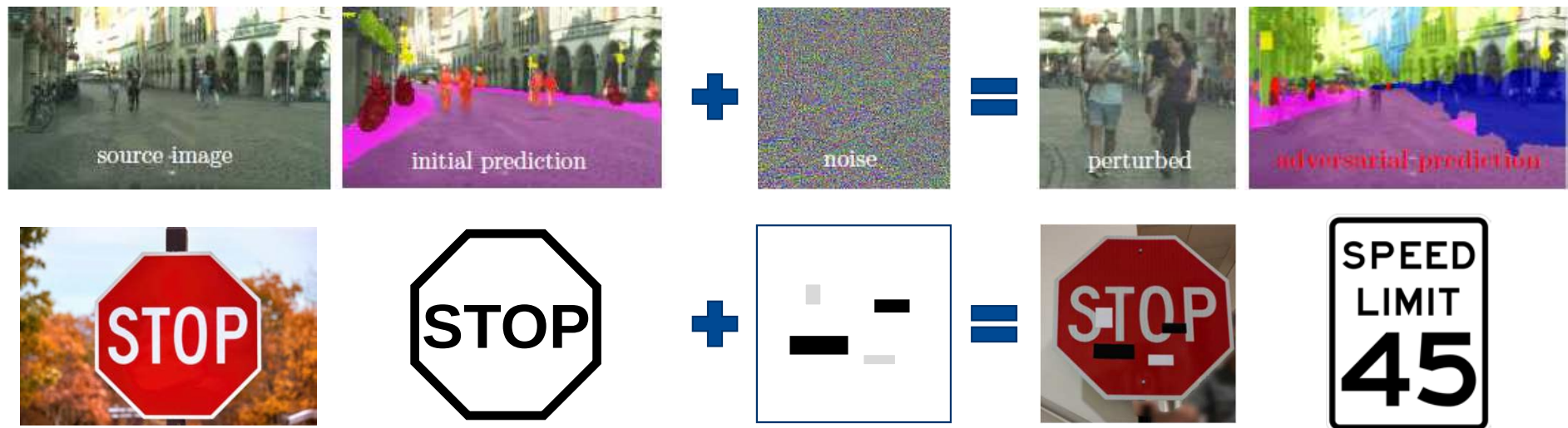
768x17x17 Neurone



6. Security

Adversarial Attacks

- Einem gültigen Modell werden gestörte Eingangsdaten geliefert um dieses zu täuschen



Quellen: Koopman et. al., Challenges in autonomous vehicle testing and validation, SCAV 17, 2017
Eykholt et. al., Robust Physical-World Attacks on Deep Learning Visual Classification, CVPR, 2018

Vielen Dank für Ihre Aufmerksamkeit.

Prof. Dr. André Steimers

Hochschule Koblenz, RheinAhrCampus Remagen

steimers@hs-koblenz.de

More Information:

Steimers A, Schneider M. **Sources of Risk of AI Systems.**
Int J Environ Res Public Health. 2022 Mar 18;19(6):3641.
doi: 10.3390/ijerph19063641

