(Security) Auditing of AI-based AD/ADAS Systems

OMNISECURE 2024 | 23.01.2024 | Fabian Woitschek

ZF Technology Domains

HB-ZEF
HB-ZEF
sters

ZF AI & CS Tech Center

Responsible for Artificial Intelligence & Cyber Security on ZF corporate level

- Founded in 2019
- Global footprint
- Collaboration with universities/research centers

- Location in Saarbrücken @ University Campus
  - Goal: 100 AI & CS experts

- AI Lab Saarbrücken
  - Trustworthy AI
  - AD / ADAS
  - Industry 4.0

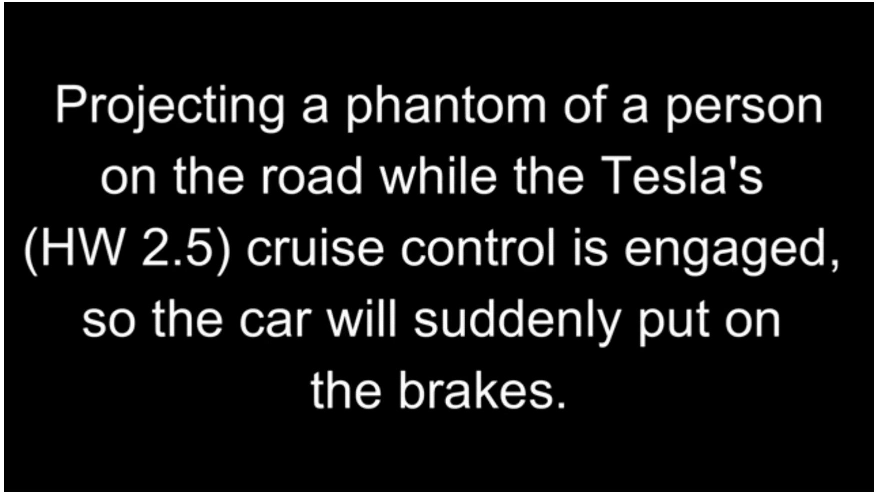
Bild: AWS Institut gGmbH

ZF AI Lab
Saarbrücken

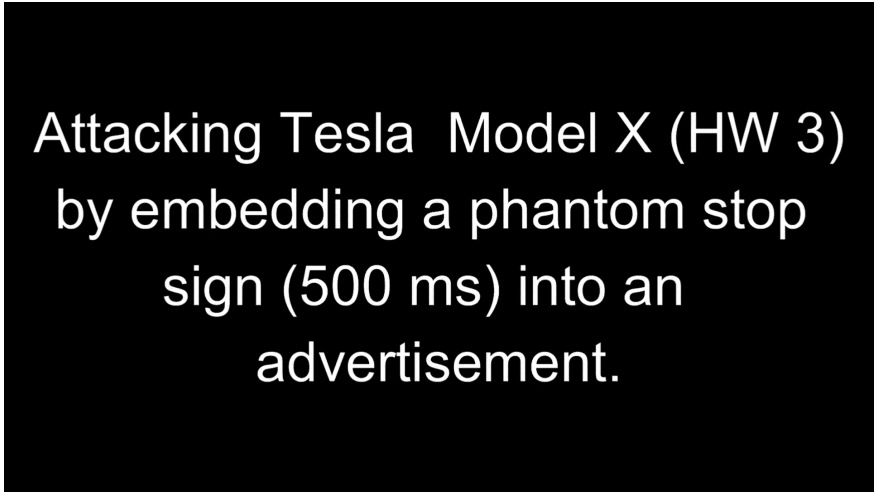
AI Security

Motivation

Phantom of the ADAS: Phantom Attacks on Driver-Assistance Systems, Nassi et al., ACM CCS, 2020



Projecting a phantom of a person on the road while the Tesla's (HW 2.5) cruise control is engaged, so the car will suddenly put on the brakes.

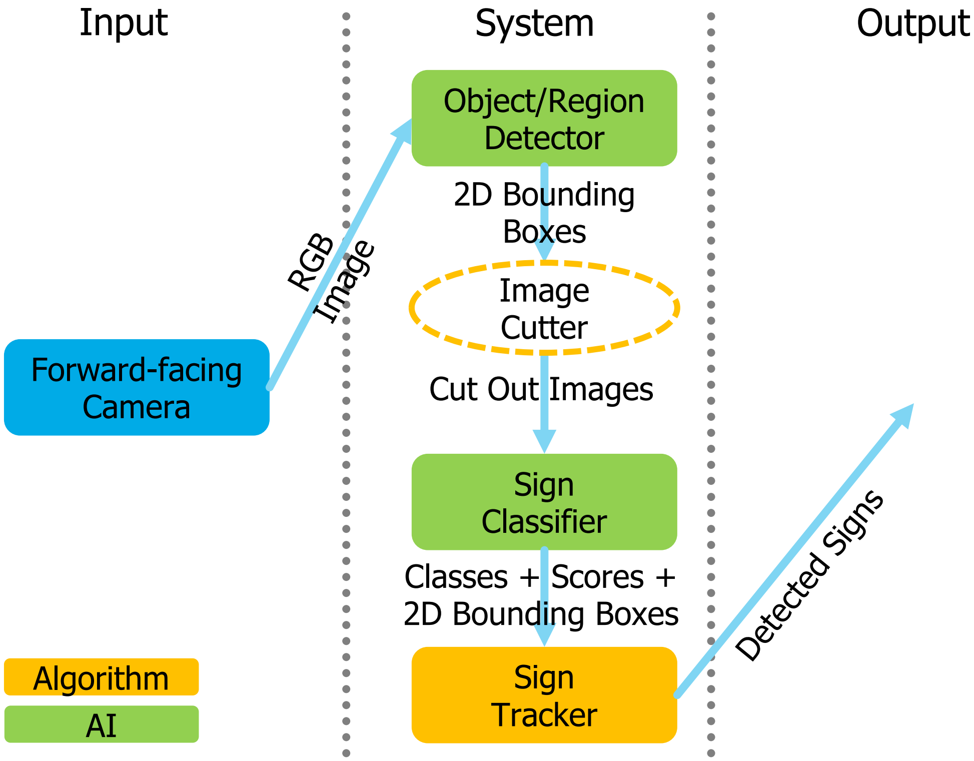


Attacking Tesla Model X (HW 3) by embedding a phantom stop sign (500 ms) into an advertisement.

➡ How applicable are adversarial attacks in reality?

Exemplary AI-based System

- Select traffic sign recognition (TSR) system as exemplary ADAS
- Use datasets for German traffic signs



Input

System

Output

Forward-facing Camera

RGB Image

Object/Region Detector

2D Bounding Boxes

Image Cutter

Cut Out Images

Sign Classifier

Classes + Scores + 2D Bounding Boxes

Sign Tracker

Detected Signs

Algorithm

AI

Adversarial Attacks Feasibility

- Report classification rate of each perturbation under 1000 different transformations

Attack Type	Classification Rate Stop \ %	Classification Rate 60 \ %

IV21 32nd IEEE INTELLIGENT VEHICLES SYMPOSIUM

https://arxiv.org/abs/2302.13570

Project: AIMobilityAudit

- How can the security & safety of AI-based systems be ensured?
- What guidelines for auditing AI-based systems should exist?
- What is needed for regulators to grant usage of AI-based systems?

ZF

TÜVIT

Federal Office for Information Security

Project report:

http://www.bsi.bund.de/dok/1079912

Publication:

https://arxiv.org/abs/2302.13567

ICISSP 2023

9th International Conference on Information Systems Security and Privacy

→ How can we test the applicability & meaningfulness of the proposed audit requirements?

Exemplary Audit

Req 7: The performance shall be compliant to the allowed worst-case error.

1. Procedure: The performance shall be compliant to an accuracy above 90% under heavy rain conditions.
2. Verdict: Failed

Tested Samples	Correct Predictions	Failed Predictions	Accuracy
2580	2031	549	78,72% < 90%

Alternative Specification

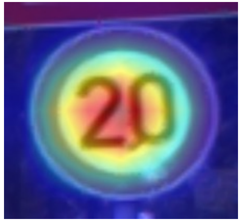
1. Procedure: The performance shall be compliant to an accuracy above 90% under a PGD attack.
2. Verdict: Failed

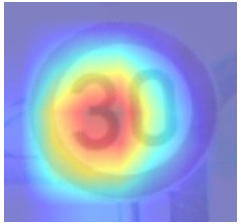
Tested Samples	Correct Predictions	Failed Predictions	Accuracy
2580	552	2028	21,40% < 90%

Exemplary Audit

Req 32: The operational design domain (ODD) requirements shall be analyzed to derive test cases for interpretable decisions of the AI model.

1. Procedure: The AI model shall not be susceptible to background information.
2. Verdict: Passed





Red regions have highest influence on the decision

Req 19: Test cases based on corner cases of the AI model shall be derived.

1. Procedure: On high brightness data the AI model should have an accuracy comparable to normal data.
2. Verdict: Passed

Tested Samples	Correct Predictions	Failed Predictions	Accuracy
2580	2548	32	98,76% ~ 99,19%

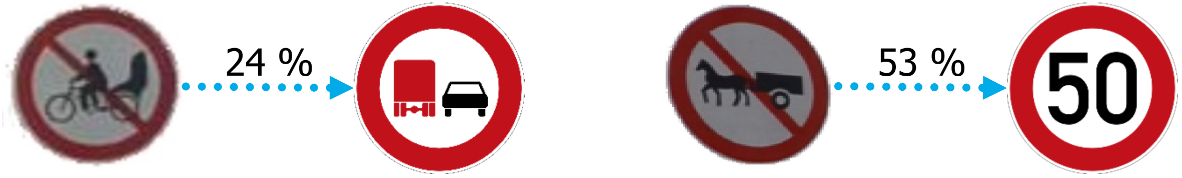
Exemplary Audit

Req 14: The development process shall be tracked.

1. Procedure: No specification required.
2. Verdict: Passed
   - Development of system is tracked using Git
   - Development of AI model is tracked using MLflow

Req 18: The AI model shall be tested against out-of-distribution data.

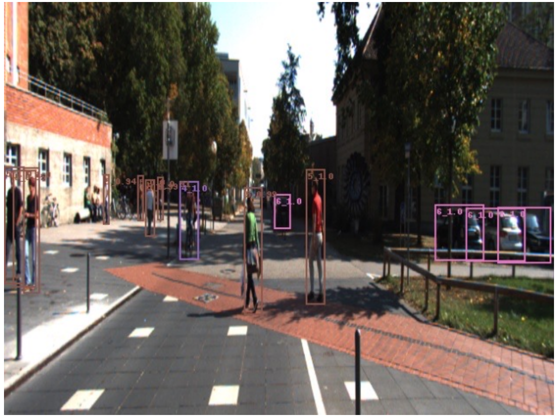
1. Procedure: The AI model shall classify Chinese traffic signs with an accuracy below 50%.
2. Verdict: Failed

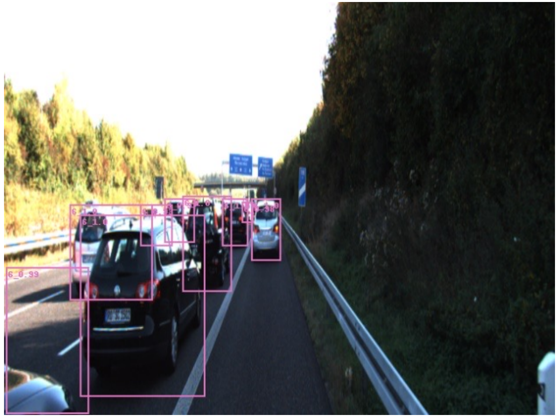


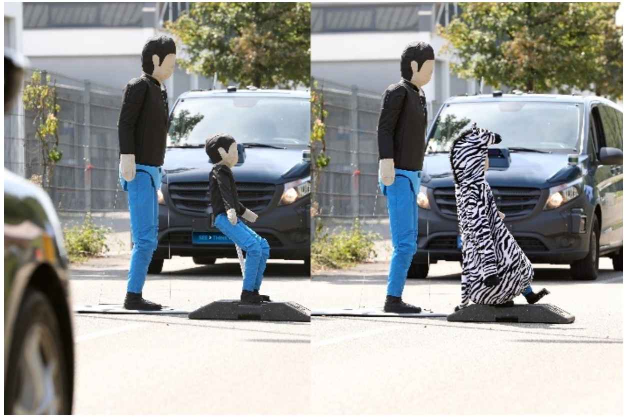
Practical Tests

- Investigate real industry systems on test track

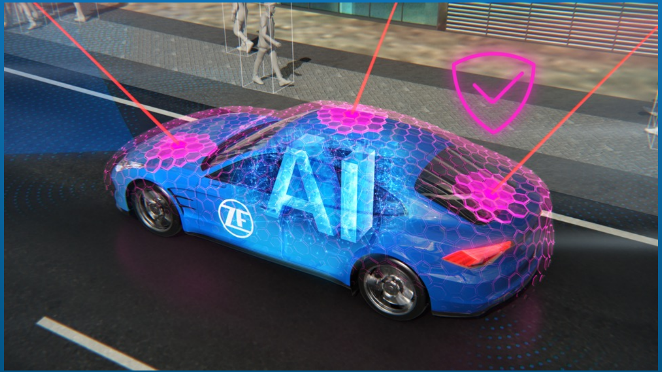
- Use road user detection (RUD) system as 2nd exemplary ADAS







Conclusion



- Summary
  - Important to ensure trustworthiness of AI-based systems
  - Apply proposed audit requirements to industry grade systems
  - Perform practical tests in the real-world
  - Cooperation between industry, auditors & regulators to find common basis for deployment

- Outlook
  - Obtain practical insights, limitations & feedback for requirements
  - Use obtained results as blueprint for standardization activities

Questions?

fabian.woitschek@zf.com

ZF AI Lab
Saarbrücken