Westfälische Hochschule
Gelsenkirchen Bocholt Recklinghausen
University of Applied Sciences

AI for IT security *and* IT security for AI

Prof. Dr. (TU NN)

Norbert Pohlmann

Professor** for **Cyber Security

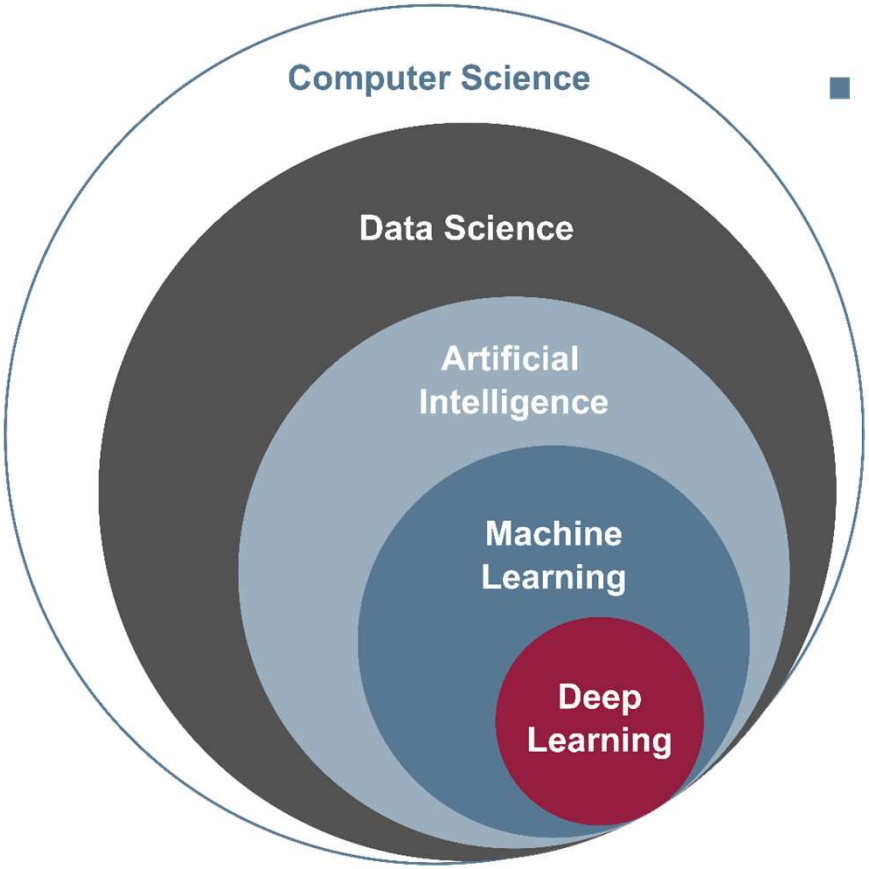
Director** of the **Institute for Internet Security – if(is)

Chairman of the board of the IT Security Association TeleTrusT

Member of the board** of the **Internet industry association eco.

if(is) internet security.

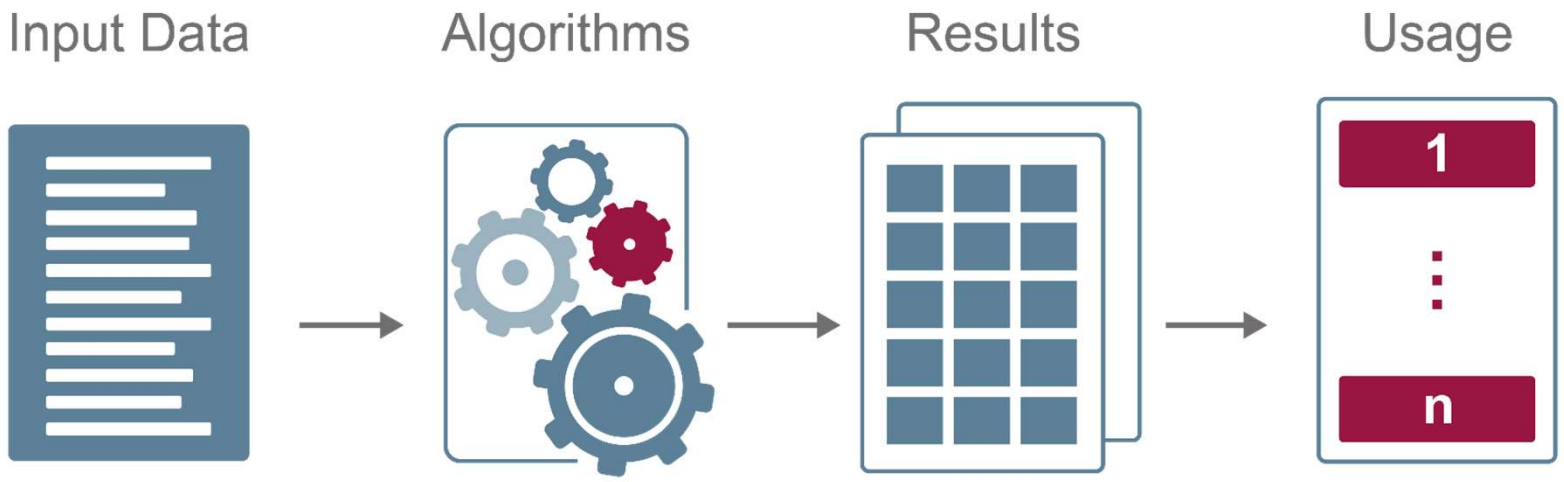
Classification
→ Artificial intelligence (AI)



- **Data science** generally refers to the **extraction of knowledge from data.**
- **Artificial intelligence** translates intelligent behavior into algorithms.
  - **Strong "Artificial Intelligence"** *automatically replicate „human-like intelligence“.*
  - Superintelligence, **Singularity** *(“Machine” improves itself, is more intelligent than humans … **future**)*
    - **Weak “artificial intelligence”** *(machine learning – successfully implemented today)*
    - **Machine learning** is "artificial" **generation of knowledge from experience (in data)** by computer.
      - **Deep learning** is an important **improvement** of machine learning

Large Language Model (LLM) like ChatGPT

Machine learning
→ Workflow



Input Data
Quality of the input data: completeness, representativeness, timeliness …

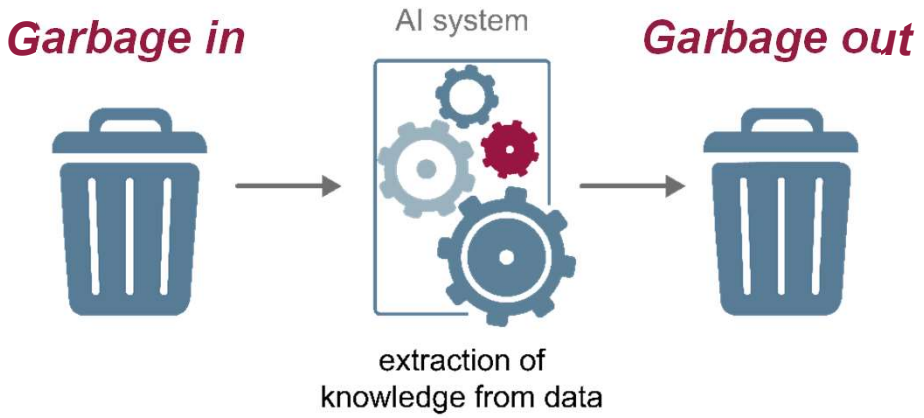
Algorithms (ML)
Support Vector Machine (SVM), k-Nearest Neighbor (kNN) ... Deep Learning

Results
Results from the processing (AI algorithm) of the input data.

Usage
The application decides how to use results (trust).

Trustworthiness of AI
→ Quality of the data

Paradigm



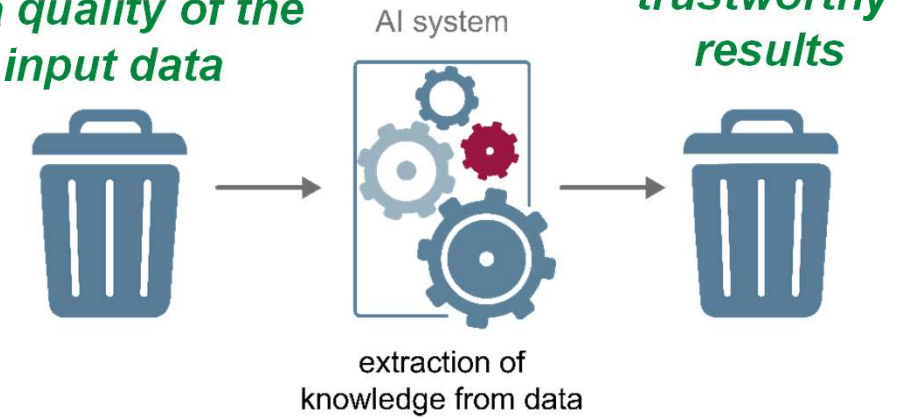
Standards for data quality:

→ Content of the data and correctness
→ Traceability of data (including data sources)
→ Completeness and representativeness
→ Availability and timeliness

Motivate high quality and secure **sensors**

high data quality of the input data

qualitative, trustworthy results

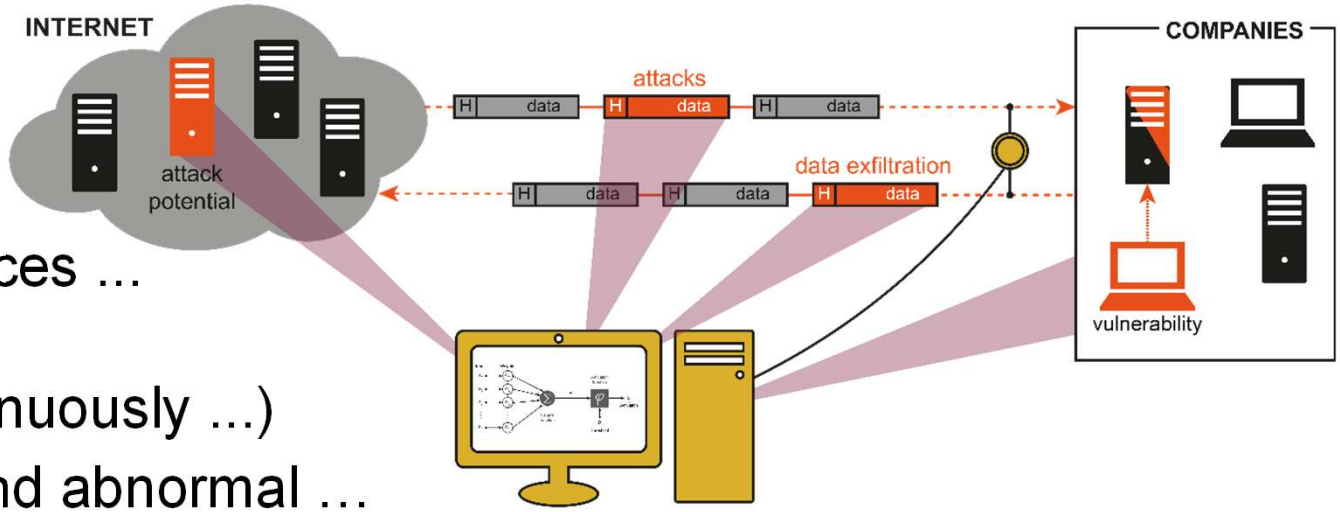


Other Ideas:

→ Establish data pools
→ Promote exchange of data
→ Create interoperability
→ Push open data strategy

Artificial intelligence
→ for IT security

- Increasing the **detection rate** of attacks
  - Network, IT end devices ...
  - adaptive models (independently, continuously ...)
  - Difference: normal and abnormal …



innovative detection of malicious network traffic

- **Support / Relief from IT security experts** (of whom we do not have enough)
  - Finding **important** security-relevant events (prioritization)
  - **(Partial) autonomy** in response … resilience ...

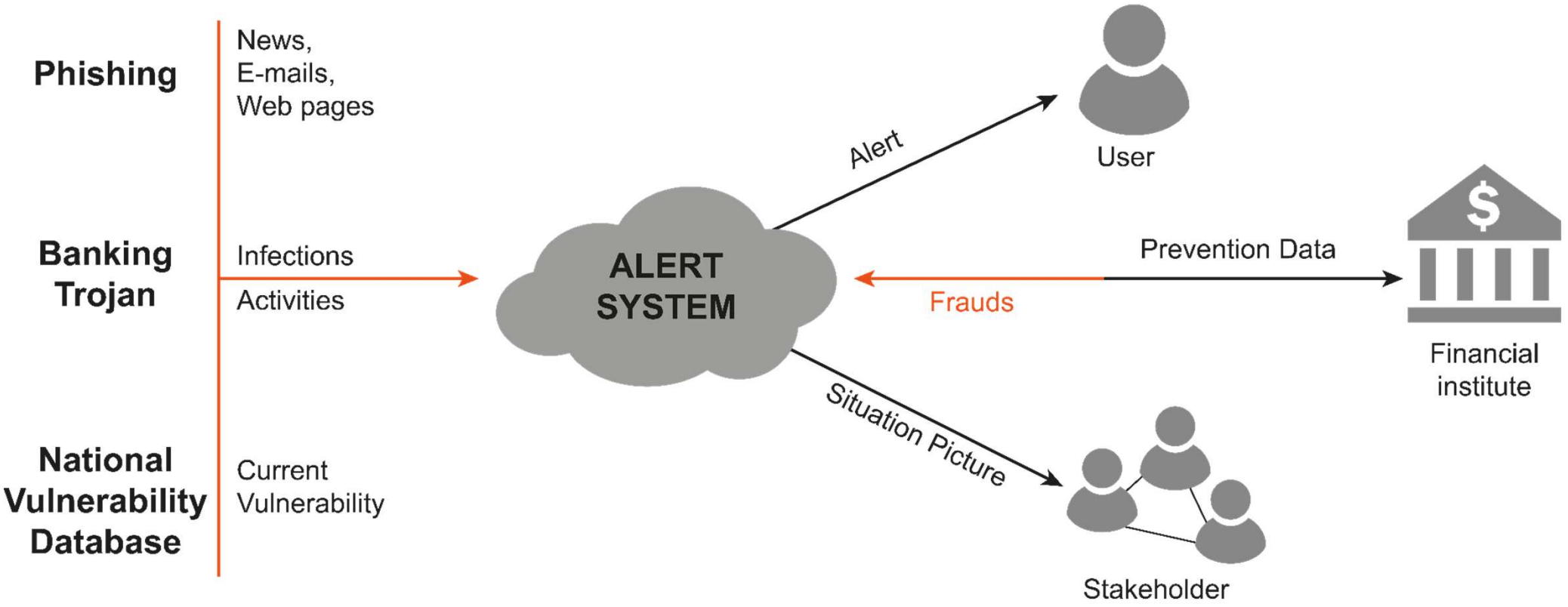
- **Improvements** to existing **IT security solutions**
  - AI contributes to increased impact and robustness
  - For example: risk-based and adaptive authentication

Research projects
→ Alert-System for online banking

- **How could a solution look like?**
  - Daily warnings in the event of an increased risk situation (online banking)
    → enable the bank customer and the bank to react quickly and appropriately
  - Instruct the users when there are dangers
    → so that the bank customer can behave "correctly"

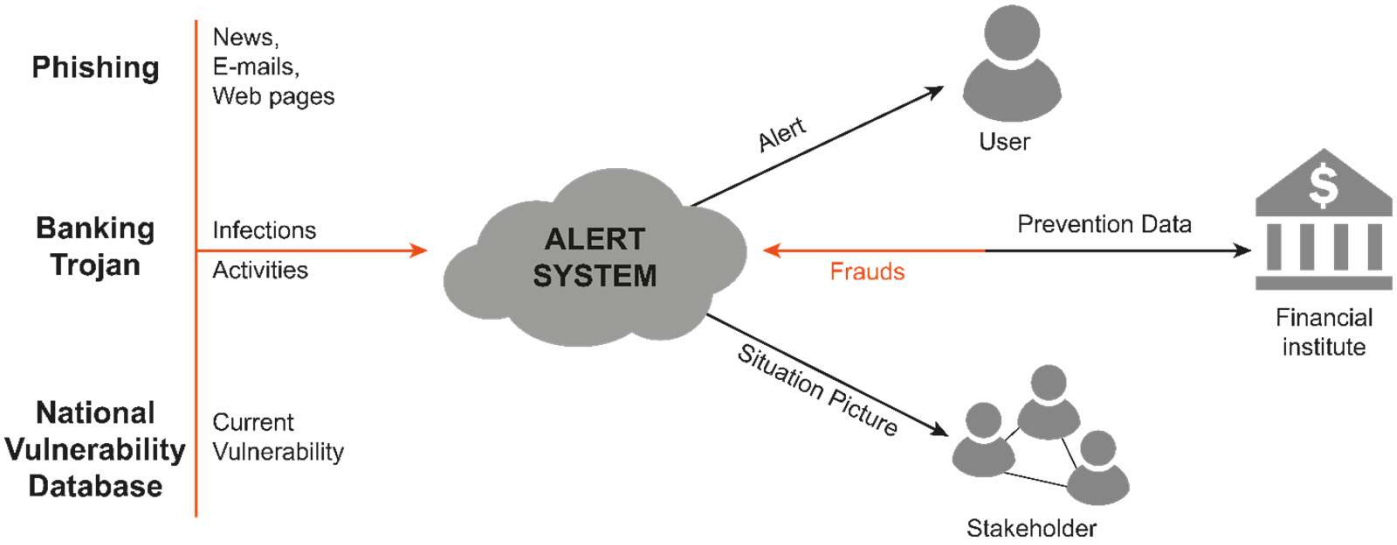
- **Approach of the alert system**
  - Identify **security metrics** for fraud
  - Determine **danger situation** with AI
  - **Warn** users and banks

Alert-System for online banking
→ Basic concept



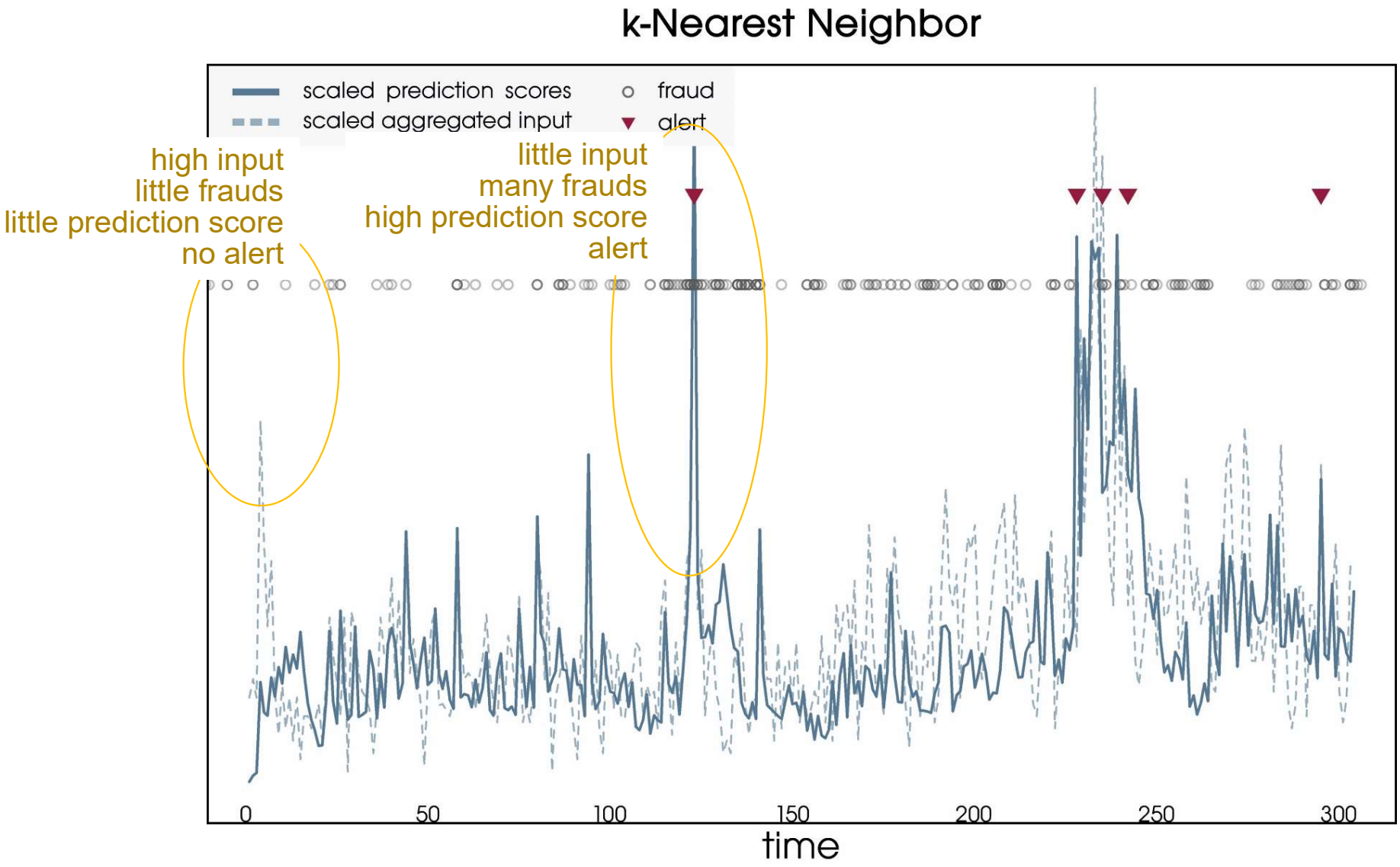
Alert-System for online banking
→ Numbers for the test period of 456 days

- 1.904 News (phishing attack) – “Stackoverflow Network”
- 5.589 **E-mail** (phishing attack) – „Spam Archive“
- 2.776 Phishing **websites** – „PhishTank“
- 23.184 **infections** of banking Trojans (malware) - Anti-malware companies
- 875 relevant **vulnerabilities** (NVD)
- 459 successful **fraud cases** in online banking - banking group



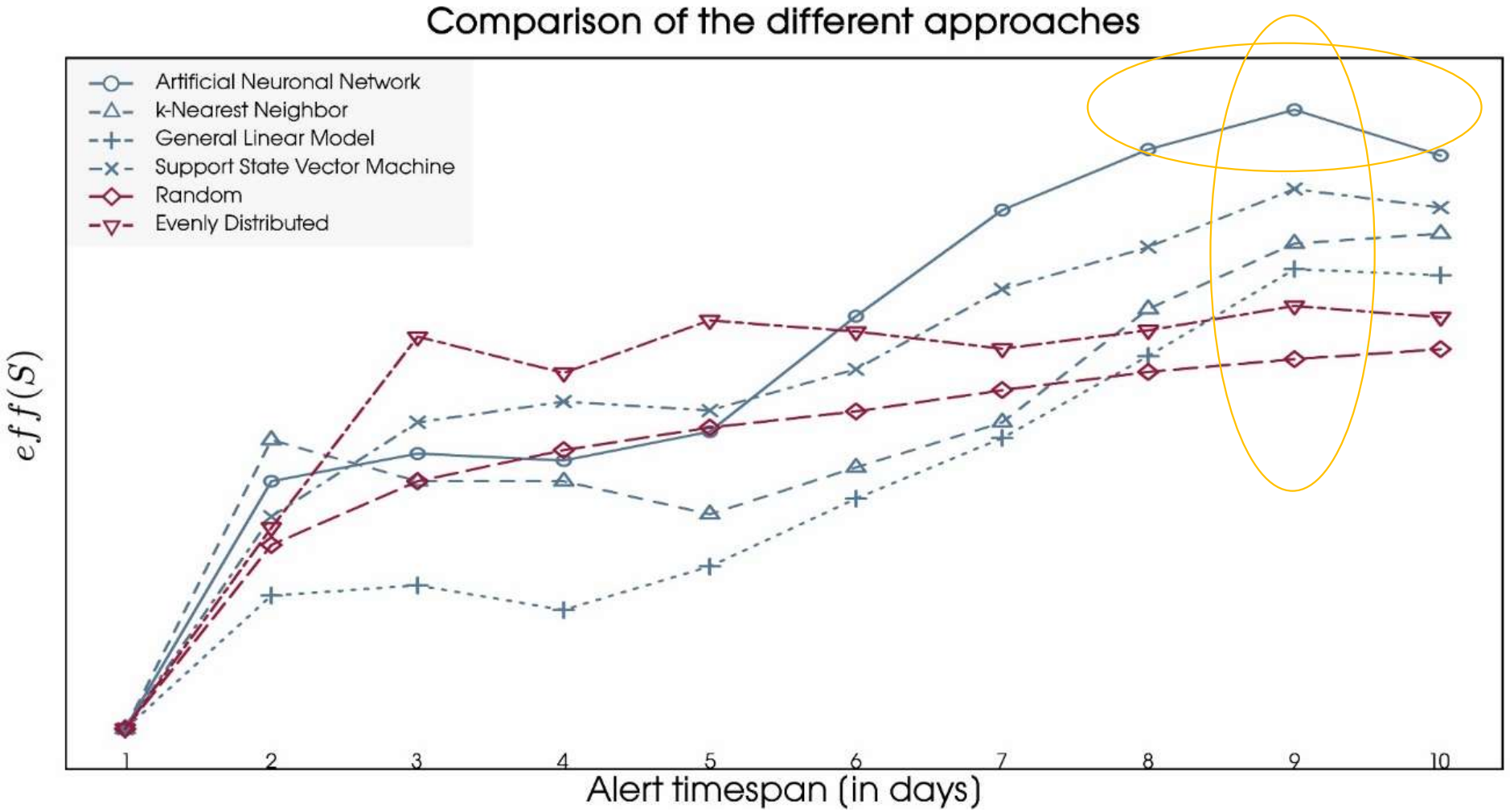
1/3 for the training period (152 days) 2/3 for evaluation period (304 days)

Assess the result
→ k-Nearest Neighbor



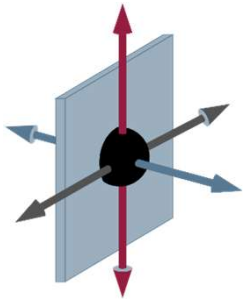
Results
→ Comparison of the different methods

„But three times as much time for training for Artificial Neural Networks“

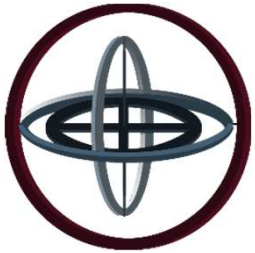


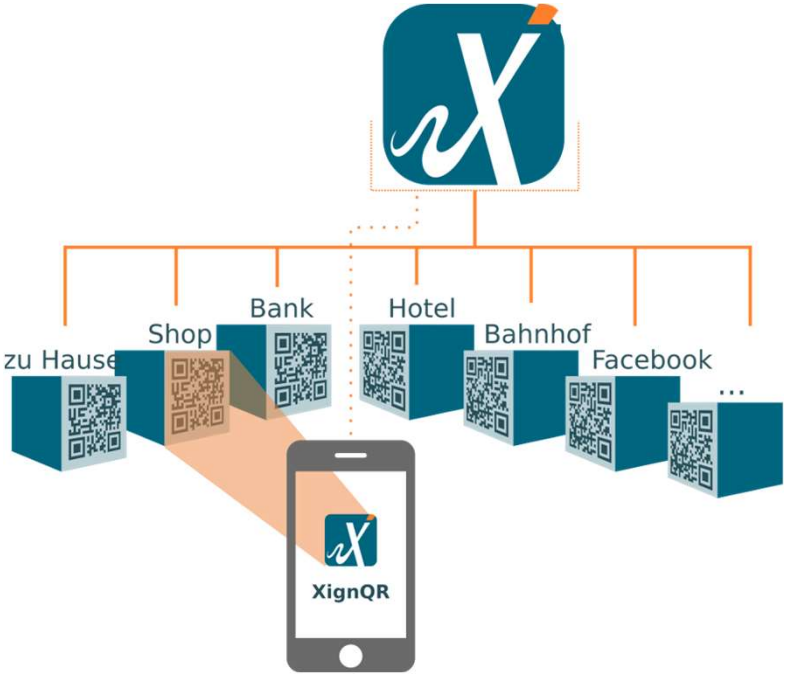
Research projects
→ Passive Authentication

- A user is automatically detected by the way of scanning the QR code.
- Throughout the process, passive biometric movement data is measured.
- Data collection by
  - **Accelerometer**

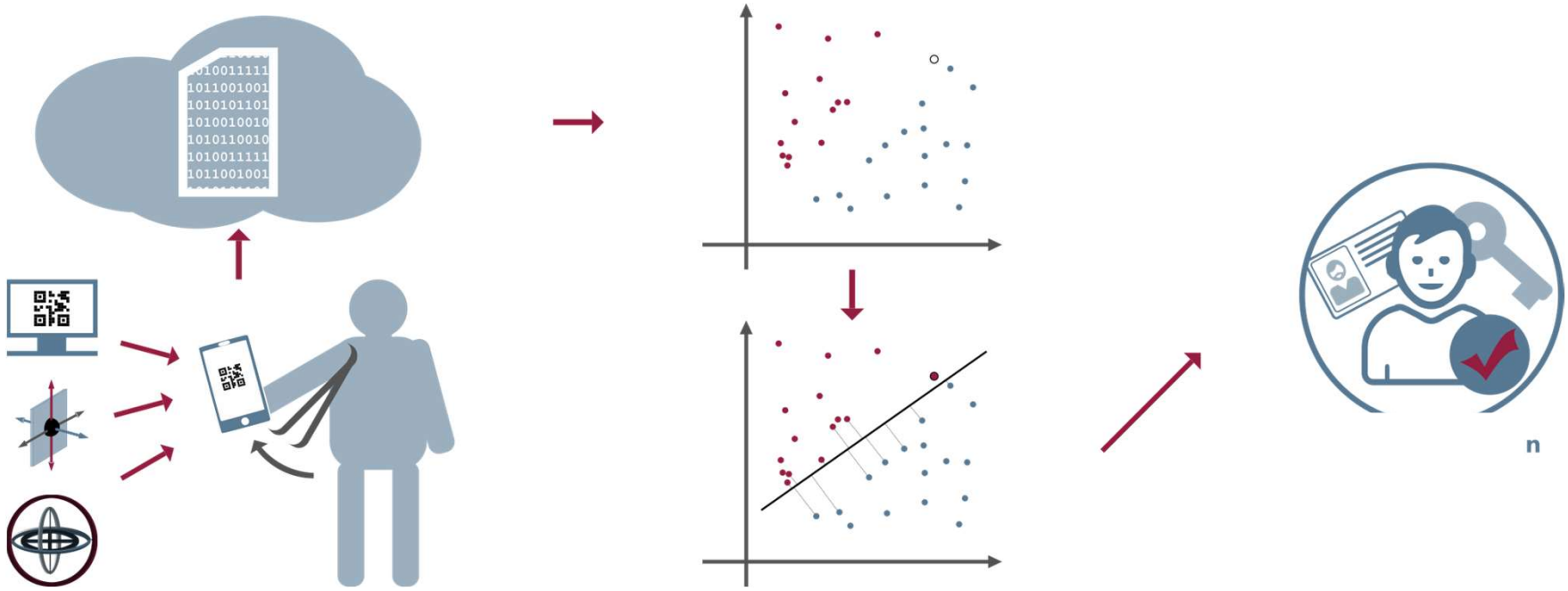


  - **Position sensor**



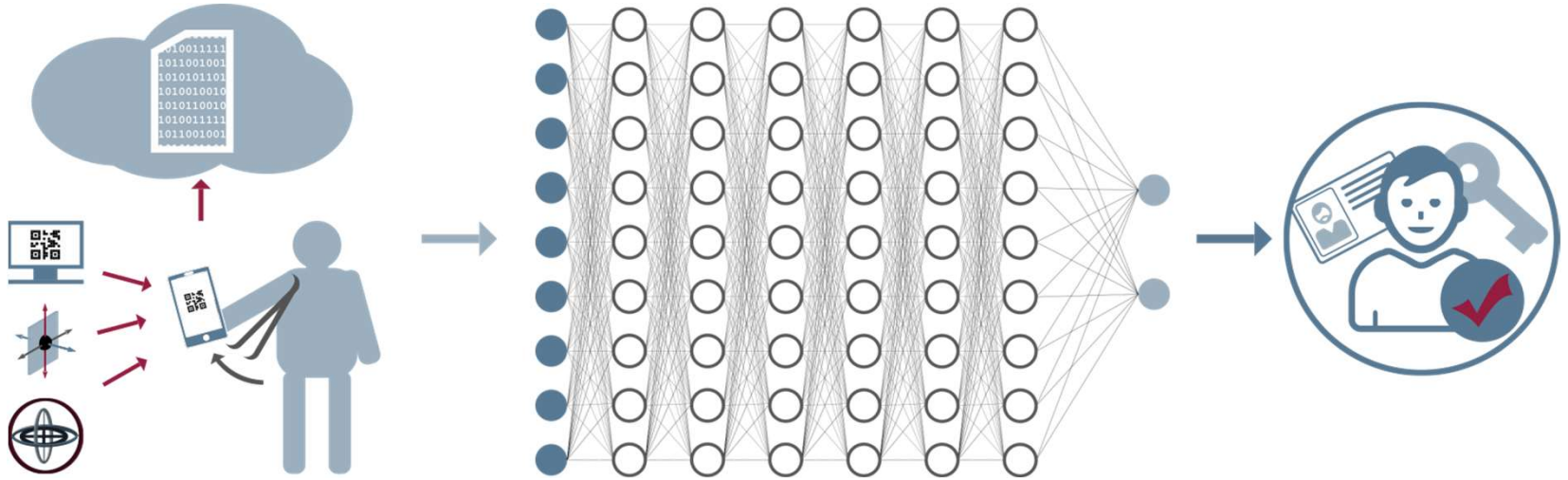


Passive Authentication
→ Support-Vector-Machine (SVM)



- **Input data:**
  - User takes the smartphone from pocket
  - Measure **location** and **acceleration** of the smartphone
- **ML algorithm:**
  - Data is classified by a model
  - red match is **positive** classification
  - blue a **negative** classification (e.g. other users)
- **Output:**
  - Authentication is either successful or fails (**95 %**)

Passive Authentication
→ Artificial Neural Networks



- **Input data:**
  - Location and acceleration data of the user are generated

```
time, type, x, y, z
271, Accelerometer, -0.07606506, 9.173798, 3.6333618
277, Accelerometer, 1.0681152E-4, 9.146423, 3.5619507
279, Gyroscope, 0.027664185, 0.06774902, 0.02182006
...
```

- **ML algorithm:**
  - Input data is processed in the artificial neurons in the layers

- **Output:**

User	Accordance
0	0,059 %
1	99,85 %
2	0,087 %

```
[[5.9110398e-04 9.9853361e-01 8.7528664e-04]]

Predicted Class [1]

Predicted Person: Sandra Kreis
```

AI for IT Security
→ Further examples

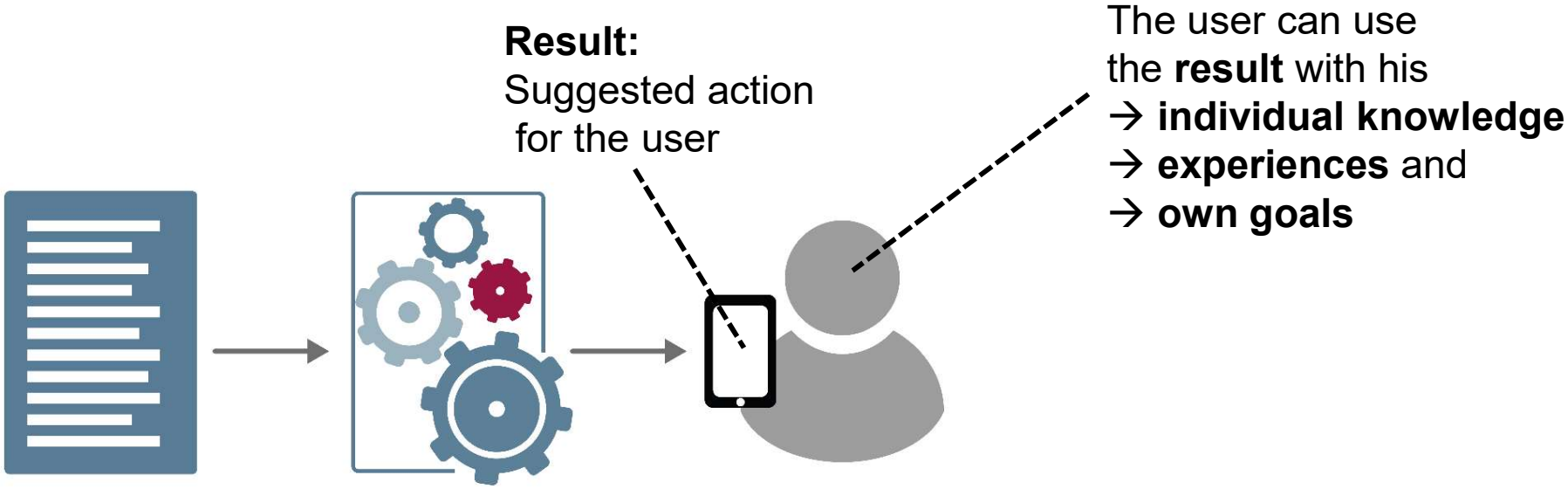
- Log analysis
- Malware detection
- Security Information and Event Management (SIEM)
- Threat Intelligence
- Voice recognition
- Image recognition (ID card, VideoIdent ...)
- Authentication method
- Fake News
- IT Forensics
- Secure software development (new possibilities with ChatGPT)
- ...

Artificial intelligence
→ Opportunities and risks

- When it comes to **individual knowledge** and **complexity of thinking,** humans are still superior to algorithms! **+**
- **Algorithms can more quickly generate knowledge** from existing data! **+**
- Individual knowledge + algorithms knowledge = **+++**
- **Practical Problem Medicine / Watson**
  - Diagnostics (machine)
  - Liability (human)

Trustworthiness
→ types of validation of results

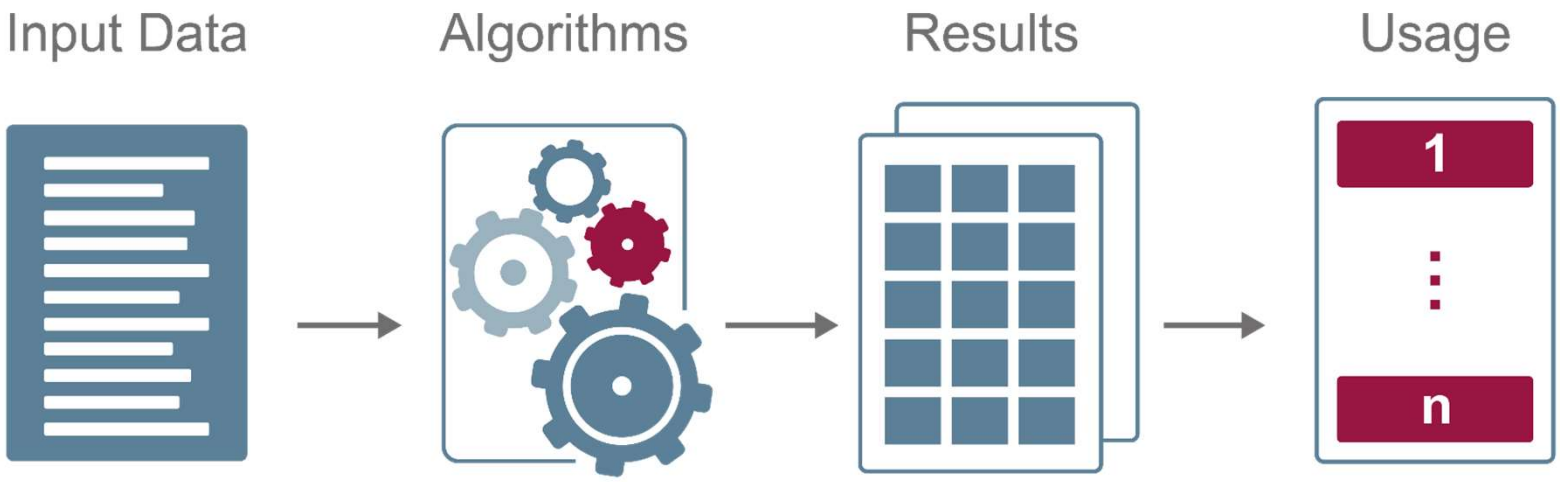
- **„Keep the human in the loop“**
  - AI result must be understood as a **recommendation for the** user.
  - This promotes the **self-determination** of users and increases their trustworthiness.



- **Automated applications** (e.g., autonomous driving)
  - Simulation, test and **validation**
  - Responsibility, **liability** and insurance

Attacks
→ on machine learning (AI)

Hackers attack and manipulate the workflow (“result”)

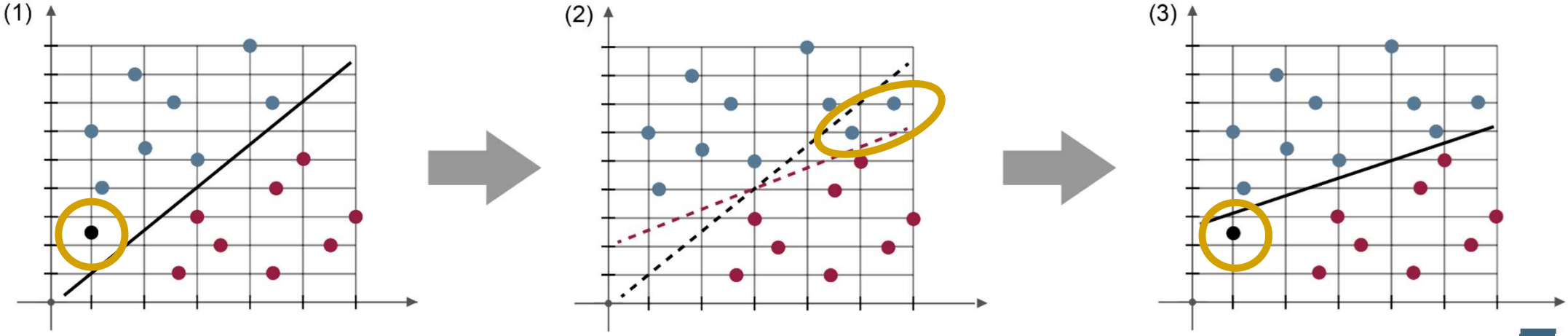


- ***Input data (input)***
- ***Algorithms / Models***
- ***Results (output)***
- ***Usage***

Attacks on machine learning

→ **Manipulation of training data** (Poisoning Attack)

- (1) **Normal classification** of a new input.
  (new black dot belongs to the blue class)
- **(2) Example: manipulation of training data**
  - Incorrectly classified data will be injected into the training phase as an attack (two more blue dots).
  - This manipulates the straight line of the model for classification (straight line becomes flatter).
- (3) This can be used by an attacker to create **wrong classifications**.
  (now the new black dot belongs to the red class)



Attacks on machine learning

→ **Manipulation of Input Data** (Evasion Attack)

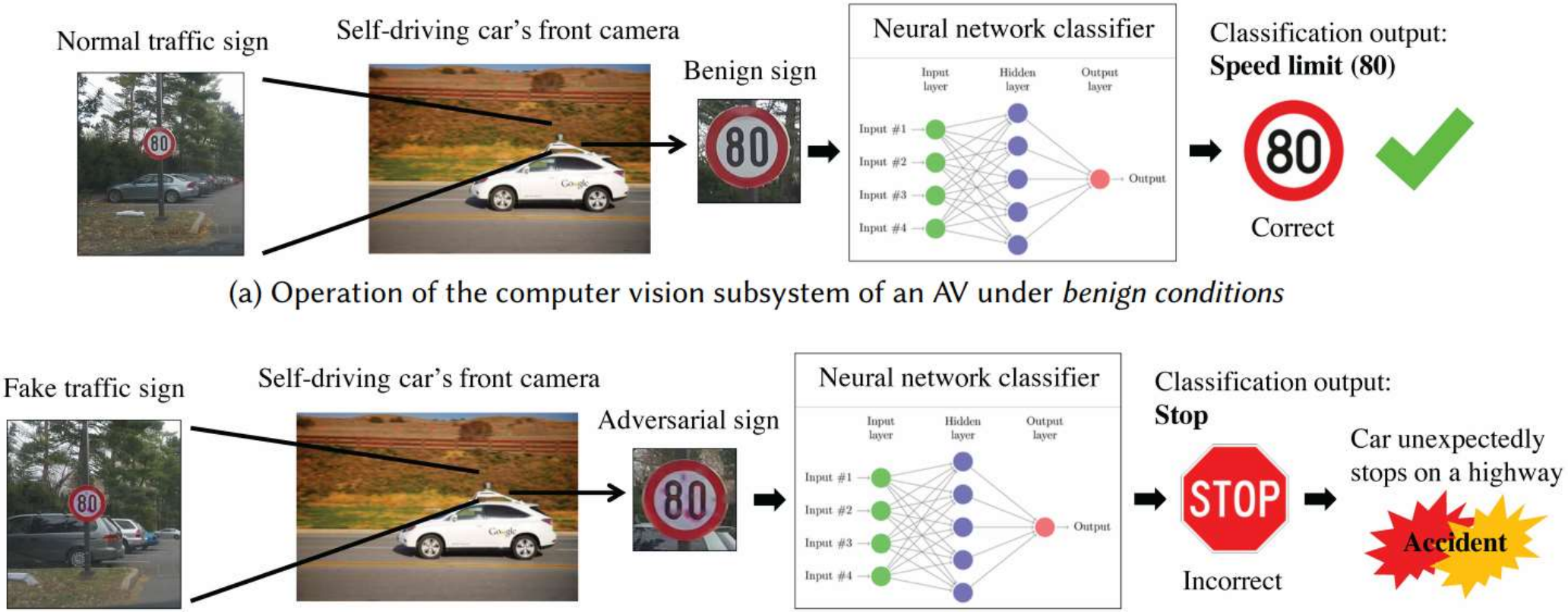
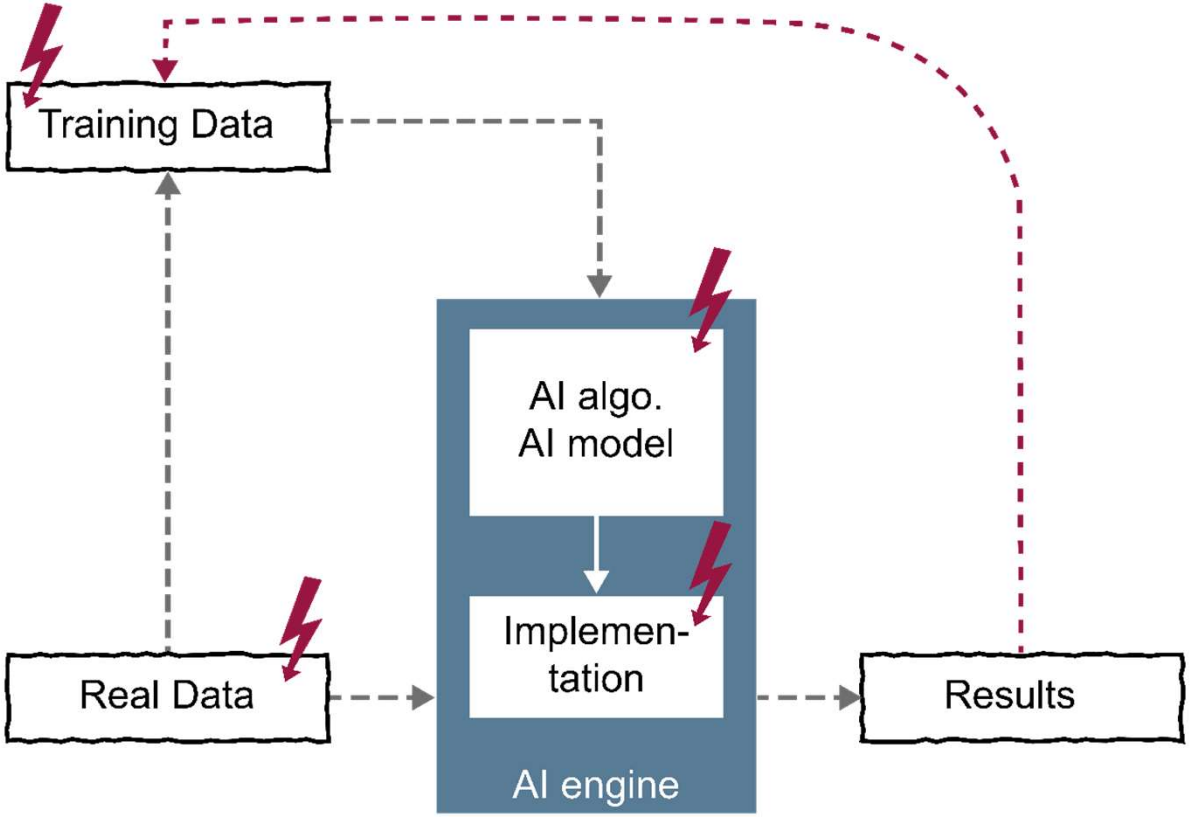


Fig. 1. **Difference in operation of autonomous cars under benign and adversarial conditions**. Figure 1b shows the classification result for a drive-by test for a physically robust adversarial example generated using our Adversarial Traffic Sign attack.

Quelle: https://arxiv.org/abs/1802.06430

Secure AI
→ Protection of the implementation and data

State of the art
IT security measures
for protection
→ *the **data** (training, real, result),*
→ *the **AI engine** and*
→ *the **application***



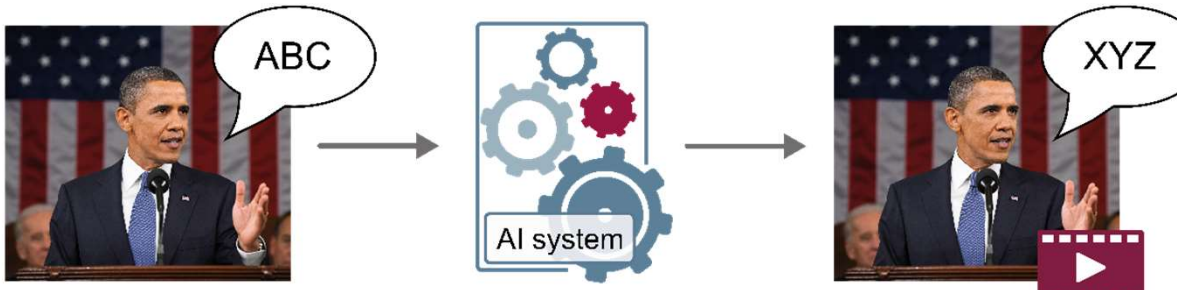
Use of a high quality AI technology

Security goals:
→ ***Integrity***
(detection of data manipulation)
→ ***Confidentiality***
(protection of business secrets)
→ ***Data protection***
(protection of personal data)
→ ***Availability***
(of the application and results)

Artificial intelligence
→ Attackers use AI

Hacker also uses AI for their own purposes (dual-use)

- Vulnerability search (faster attack, new attack vectors …)
- Social engineering (chat bots like ChatGPT ...)
- Password cracker
- Polymorphic malware (programming with ChatGPT)
- New attack structures and procedures
- Video manipulation (deep fake)
  - "Fake Obama Video„
  - "Make Putin Smile Video"



AI for IT security - IT security for AI
→ Result and outlook

- **AI / ML is an important technology in the flied of IT security**
  - Detect threats, vulnerabilities, attacks ...
  - Support of IT security experts
  - Secure software development
  - …
- **We need to secure our AI to be able to produce trustworthy results**
  - Hackers attack and manipulate data, algorithm/models and results
  - …
- **Balance of power for the future between attacker and defender**
  - The **attackers** use AI for their attacks **very successfully**
  - The defenders should do this more and also together
  - …

Westfälische Hochschule
Gelsenkirchen Bocholt Recklinghausen
University of Applied Sciences

AI for IT security *and* *IT security for AI*

*With secured **Artificial Intelligence***
into a more secure and trustworthy digital future!

Prof. Dr. (TU NN)

Norbert Pohlmann

Professor** for **Cyber Security
Director** of the **Institute for Internet Security – if(is)

Chairman of the board of the IT Security Association TeleTrusT
Member of the board** of the **Internet industry association eco.

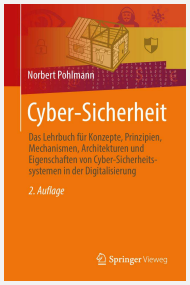
if(is) internet security.

Anhang / Credits

Wir empfehlen

Cyber-Sicherheit

Das **Lehrbuch** für Konzepte, Mechanismen, Architekturen und Eigenschaften von Cyber-Sicherheitssystemen in der Digitalisierung“, Springer Vieweg Verlag, Wiesbaden 2022
https://norbert-pohlmann.com/cyber-sicherheit/



7. Sinn im Internet (Cyberschutzraum)

https://www.youtube.com/cyberschutzraum

CYBER SCHUTZRAUM

Master Internet-Sicherheit

https://it-sicherheit.de/master-studieren/

(M) MASTER internet-sicherheit.

Glossar Cyber-Sicherheit

https://norbert-pohlmann.com/category/glossar-cyber-sicherheit/

Cyber-Sicherheit GLOSSAR

It‘s all about Trust!

https://vertrauenswürdigkeit.com/

Besuchen und abonnieren Sie uns :-)

WWW
https://www.internet-sicherheit.de

Facebook
https://www.facebook.com/Internet.Sicherheit.ifis

Twitter
https://twitter.com/_ifis

https://twitter.com/ProfPohlmann

YouTube
https://www.youtube.com/user/InternetSicherheitDE/

Prof. Norbert Pohlmann
https://norbert-pohlmann.com/

Der Marktplatz IT-Sicherheit

(IT-Sicherheits-) Anbieter, Lösungen, Jobs, Veranstaltungen und Hilfestellungen (Ratgeber, IT-Sicherheitstipps, Glossar, u.v.m.) leicht & einfach finden.
https://www.it-sicherheit.de/

Quellen Bildmaterial

Eingebettete Piktogramme: Institut für Internet-Sicherheit – if(is)

Literature

N. Pohlmann, S. Schmidt: „Der Virtuelle IT-Sicherheitsberater – Künstliche Intelligenz (KI) ergänzt statische Anomalien-Erkennung und signaturbasierte Intrusion Detection“, IT-Sicherheit – Management und Praxis, DATAKONTEXT-Fachverlag, 05/2009

D. Petersen, N. Pohlmann: "Ideales Internet-Frühwarnsystem", DuD Datenschutz und Datensicherheit – Recht und Sicherheit in Informationsverarbeitung und Kommunikation, Vieweg Verlag, 02/2011

U. Coester, N. Pohlmann: „Diskriminierung und weniger Selbstbestimmung? Die Schattenseiten der Algorithmen“, tec4u, 12/17

N. Pohlmann: „Künstliche Intelligenz und Cybersicherheit - Unausgegoren aber notwendig“, IT-Sicherheit – Fachmagazin für Informationssicherheit und Compliance, DATAKONTEXT-Fachverlag, 1/2019

U. Coester, N. Pohlmann: „Wie können wir der KI vertrauen? - Mechanismus für gute Ergebnisse“, IT & Production – Zeitschrift für erfolgreiche Produktion, Technik-Dokumentations-Verlag, Ausgabe 2020/21

D. Adler, N. Demir, N. Pohlmann: „Angriffe auf die Künstliche Intelligenz – Bedrohungen und Schutzmaßnahmen“, IT-Sicherheit – Mittelstandsmagazin für Informationssicherheit und Datenschutz, DATAKONTEXT-Fachverlag, 1/2023

P. Farwick, Pohlmann: „Chancen und Risiken von ChatGPT – Vom angemessenen Umgang mit künstlicher Sprachintelligenz“, IT-Sicherheit – Mittelstandsmagazin für Informationssicherheit und Datenschutz, DATAKONTEXT-Fachverlag, 4/2023

N. Pohlmann: Lehrbuch „Cyber-Sicherheit“, Springer Vieweg Verlag, Wiesbaden 2022 Druckausgabe (ISBN 978-3-658-36242-3) und eBook (ISBN 978-3-658-36243-0).

See more articles: https://norbert-pohlmann.com/artikel/