



**Westfälische
Hochschule**

Gelsenkirchen Bocholt Recklinghausen
University of Applied Sciences

KI und IT-Sicherheit

→ Mehr Sicherheit, mehr Gefährdungen

Prof. Dr. (TU NN)

Norbert Pohlmann

Professor für Cyber-Sicherheit und Leiter des Instituts für Internet-Sicherheit – if(is), Westfälische Hochschule, Gelsenkirchen

Vorstandsvorsitzender Bundesverband IT-Sicherheit - TeleTrust

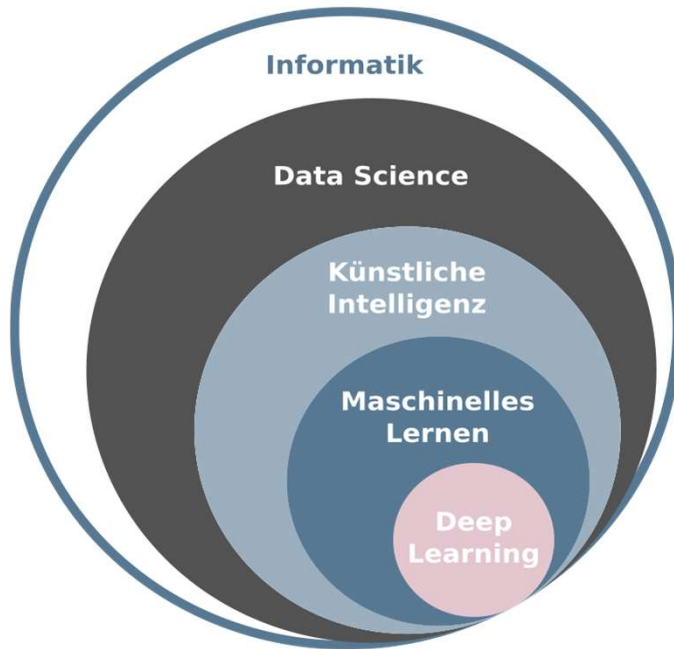
Vorstand im Verband der Internetwirtschaft - eco

if(is)
internet-sicherheit.

- **Einordnung, Definitionen, Begriffe und Prinzipien**
- **Künstliche Intelligenz für IT-Sicherheit**
- **Angreifer nutzen Künstliche Intelligenz**
- **IT-Sicherheit für Künstliche Intelligenz**
- **Künstliche Intelligenz braucht Vertrauen**
- **Zusammenfassung und Ausblick**

Einordnung

→ (Künstliche Intelligenz) **Maschinelles Lernen**



- **Data Science** bezeichnet generell die **Extraktion von Wissen** aus Daten.
- **Starke „Künstliche Intelligenz“ (Zukunft)** soll automatisiert „**menschenähnliche Intelligenz**“ nachbilden. **Singularität, Artificial General Intelligence (AGI)** - („Maschinen“ verbessern sich selbst, *sind „intelligenter“ als Menschen*)
- **Schwache „Künstliche Intelligenz“ (heute sehr erfolgreich)** **Maschinelles Lernen (ML)** ist ein Begriff für die **Generierung von Wissen aus Erfahrung** (in Daten) durch Computer.
- **Deep Learning** → Methode des MLs → **verbesserte Ergebnisse**
- **Large Language Modell (LLM)** Basismodell für generative KI (**Stochastischer Papagei**)
- **Generative KI (GenAI)** ist eine Form der künstlichen Intelligenz, die *gestützt auf ihren Trainingsdaten* Texte, Bilder, Code und verschiedene andere **Inhalte produzieren** kann.

Singularität

→ Artificial General Intelligence (AGI)

- Singularität bezieht sich auf einen **hypothetischen Zeitpunkt**, bei dem sich die KI selbstständig so rasant entwickelt, dass **wesentliche / unumkehrbar Veränderungen** für die **Menschheit** entstehen.
- **Folge**: Der Mensch verliert **die Kontrolle über die KI-Technologie und die Zukunft wird ungewiss** (*Innovationen, Werte, Wirtschaft*).
- **Chancen**: Es werden **Lösungen zu fundamentalen Fragen des Lebens geschaffen**.
- **Risiken**: Im Bereich der **Ethik, Unvorhersehbarkeit** und potenziell negative **Auswirkungen** des schnellen technologischen Fortschritts auf die Gesellschaft.

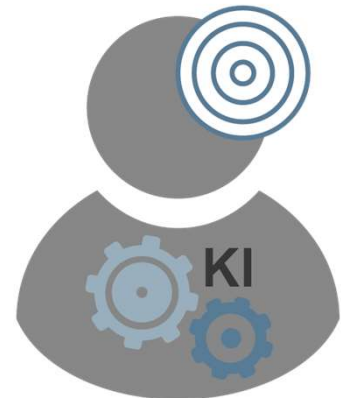
Singularität ist eine interdisziplinäre Herausforderung

- **Vollständigkeit** der genutzten der Daten / des (KI-)Wissens
- **Fehler** bei der Künstlichen Intelligenz (Algorithmus, Modell, Daten)
- **Physikalische Grenzen** der Leistungsfähigkeit, die „KI“ zu rechnen
- **Menschliche Restriktion**: genetisches Repertoire (Lernen / logischen Denken)

Singularität sollte nicht unser Ziel sein

→ **der Mensch muss die Kontrolle behalten**

→ **Superalignment** (KI agiert im Einklang mit menschlichen Werten und Zielen)



Ein Prinzip von Künstlicher Intelligenz (KI)

→ Garbage in – Garbage out

Paradigma



Standards für die Datenqualität:

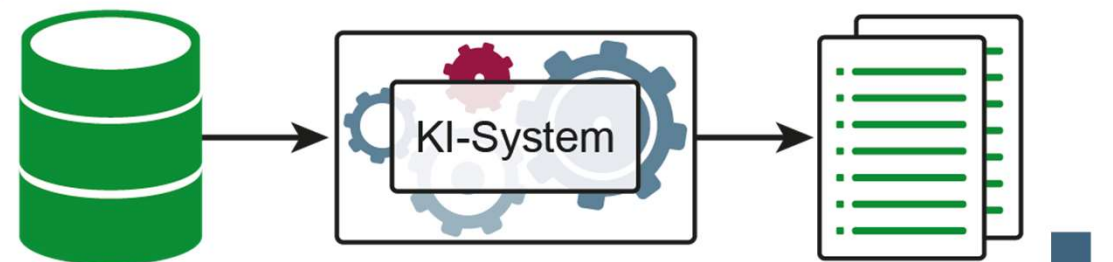
- **Inalthöhe** der Daten und **Korrektheit**
- **Nachvollziehbarkeit** (Datenquellen)
- **Vollständigkeit** und **Repräsentativität**
- **Verfügbarkeit** und **Aktualität**

Qualitativ hochwertige und sichere **Sensoren** motivieren

Weitere Aspekte zur Erhöhung der Qualität:

- Datenpools etablieren
- Austausch von Daten fördern
- Interoperabilität schaffen
- Open Data-Strategie puschen

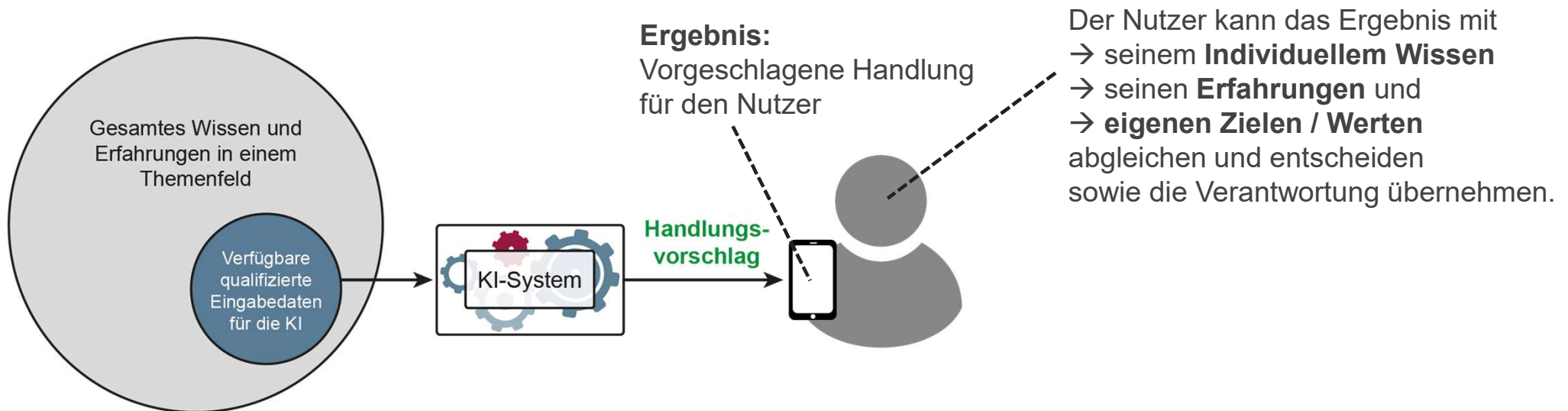
hochqualitative Daten



Vertrauenswürdigkeit von Ergebnissen

→ Umgang mit den KI-Ergebnissen

- „Keep the human in the loop“
 - KI-Ergebnisse werden als **Handlungsvorschlag** für den Nutzer verstanden.
 - *Dieses fördert die **Selbstbestimmtheit** des Nutzers und erhöht das Vertrauen*



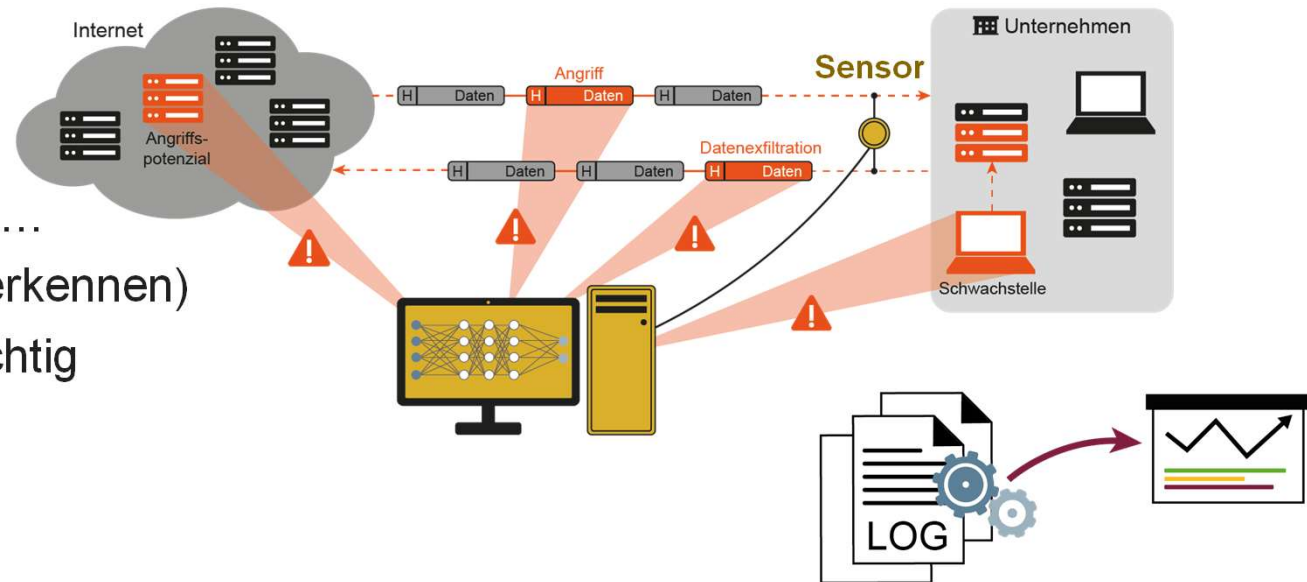
- **Automatisierte Anwendungen wie z. B. autonomes Fahren**
 - **Hoher Aufwand:** Simulation, Test und Validierung
 - **Umgang mit dem Restrisiko:** Verantwortung, Haftung und Versicherung

Künstliche Intelligenz für IT-Sicherheit

→ Erkennen von Angriffen

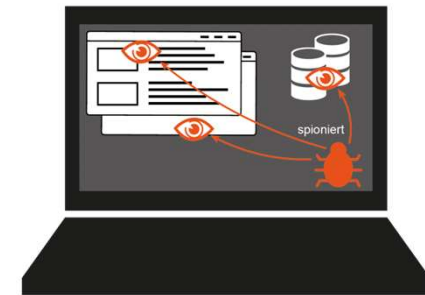
Erhöhung der Erkennungsrate von Angriffen

- Netzwerk, *IT-Endgeräte*, Server, *IoT-Geräte*, Cloudanwendungen ...
- adaptive Modelle (neue Angriffe erkennen)
- Anomalien: normal *versus* verdächtig



Weitere Bereiche, bei denen die verbesserte Erkennung eine wichtige Rolle spielt:

- Netze / Internet
- *Malware*, Spam ...
- Fake-News
- *Deepfake*
- ...



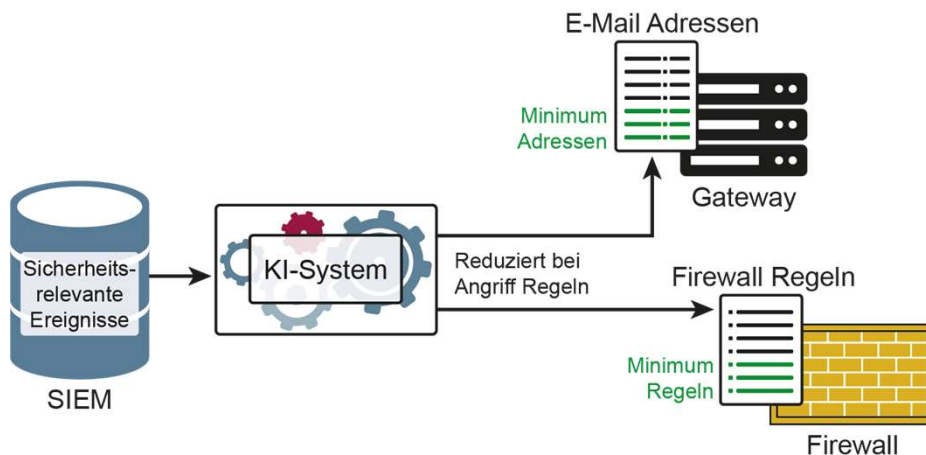
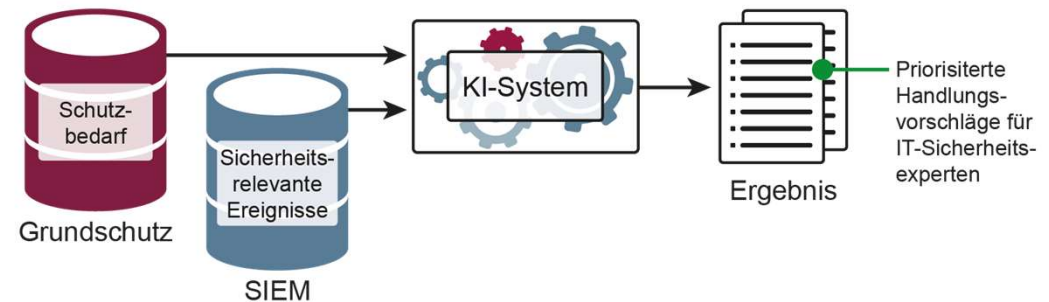
Künstliche Intelligenz für IT-Sicherheit

→ Unterstützung von IT-Sicherheitsexperten

Erkennen von **wichtigen sicherheitsrelevanten Ereignissen** mit einer **Priorisierung**

- **Idee:** IT-Sicherheitsexperten von **zeitaufwendiger Analysearbeit** zu entlasten.
- *Ein KI-System analysiert hunderte / tausende sicherheitsrelevante Ereignisse und zeigt, nach welchen **Prioritäten** diese **abgearbeitet** werden müssen.*

→ Dadurch wird der höchste Schutz **in der aktuellen Situation** für die Organisation erzielt.



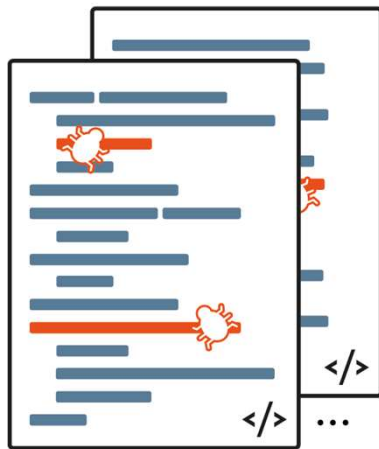
(Teil-)Autonomie bei Reaktionen

- **Idee:** Wenn ein Angriff oder eine besondere Bedrohung erkannt wird, werden sofort die **Firewall- und E-Mail-Regeln automatisch reduziert**.
 - *Die **wichtigen Prozesse** für eine Organisation bleiben **erhalten** (Minimum Regeln).*
- Dadurch wird **Angriffsfläche** für die Angreifer deutlich **reduziert** und damit Schäden verhindert.

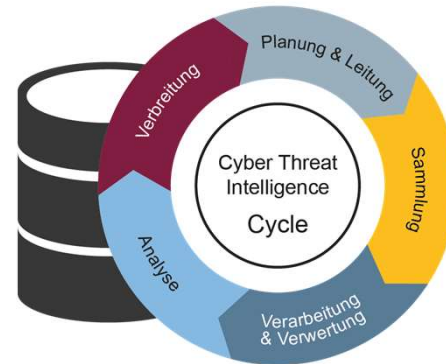
Künstliche Intelligenz für IT-Sicherheit

→ Weitere Bereiche

Sichere Softwareentwicklung



Threat Intelligence



IT-Forensik



... und vieles mehr ...

Angreifer nutzen Künstliche Intelligenz → ChatBots wie ChatGPT

Social Engineering+++

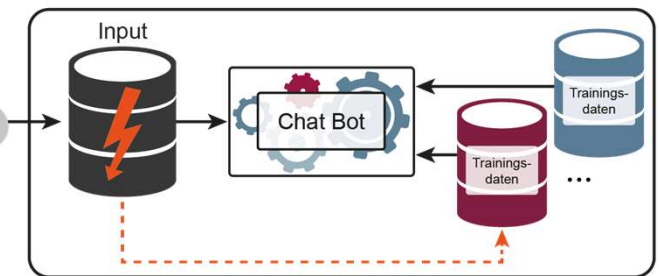
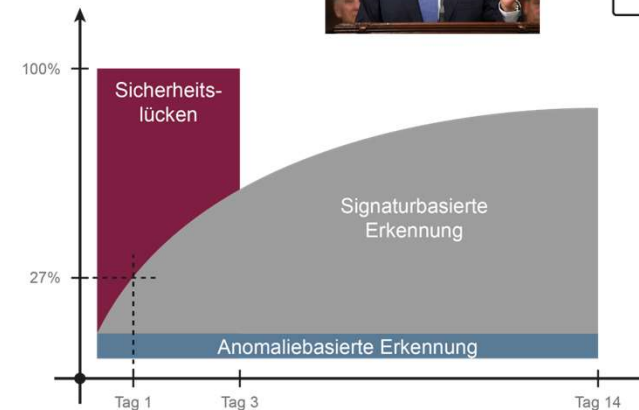
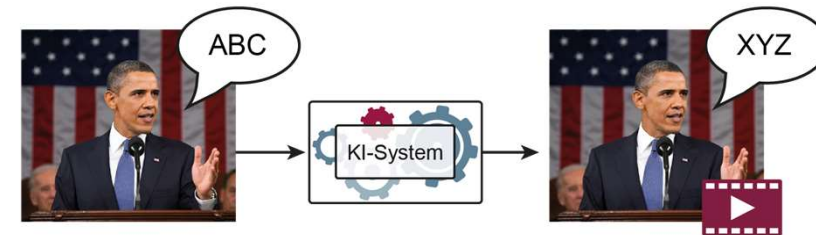
- Spear-Phishing (Erster Schritt für die meisten Angriffe)
- Fake-News (Manipulation Menschen / Gesellschaften)
- Deepfake (CEO-Fraud)

Beschleunigtes Entwickeln von Angriffen

- Polymorphe Malware

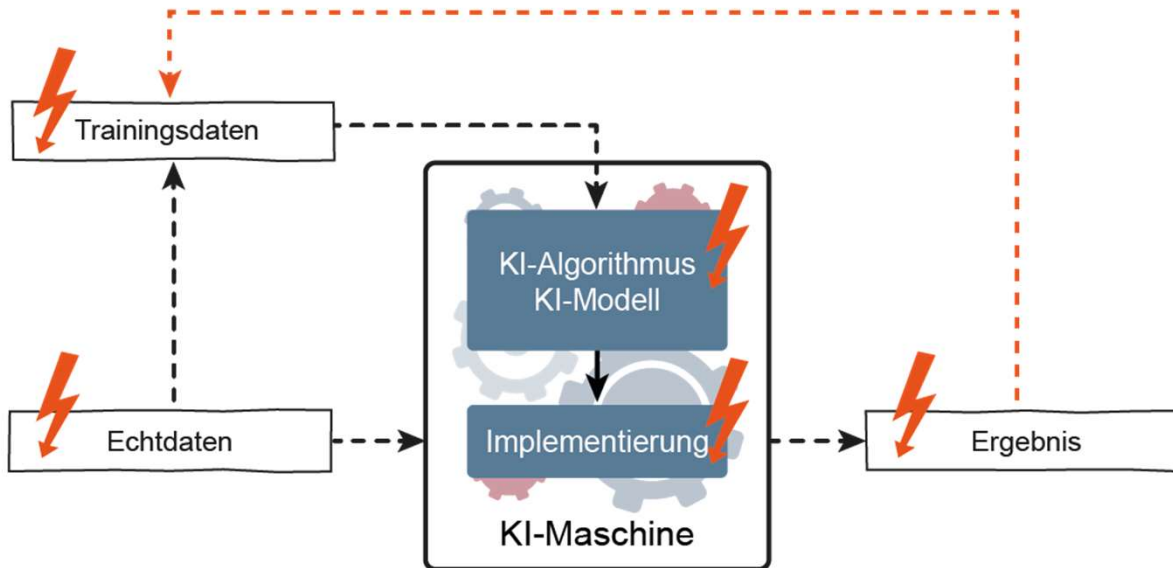
Vertrauliche Informationen als Input eines zentralen Dienstes

- Zusätzliche Angriffsfläche der vertraulichen Informationen
- Nicht-gewünschte Integration der vertraulichen Informationen in die Trainingsdaten



IT-Sicherheit für Künstliche Intelligenz

→ Angriffe



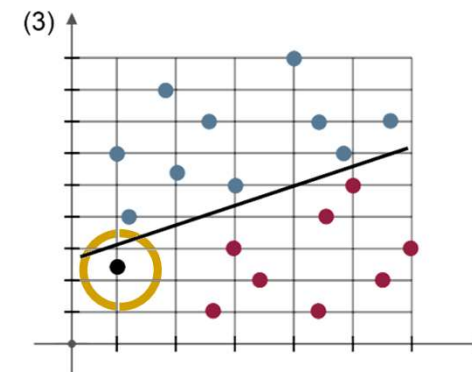
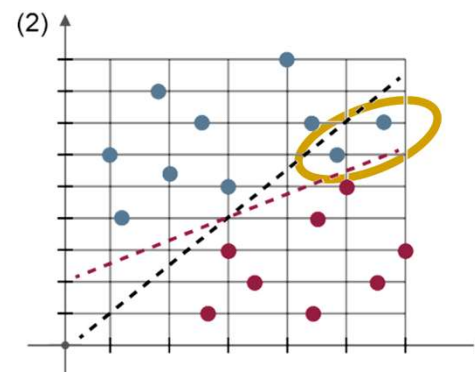
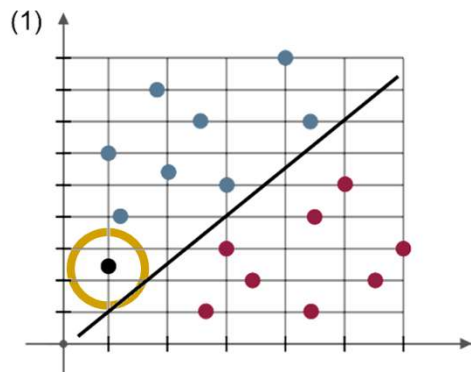
Angriffe auf KI-Systeme

→ Poisoning Attack

→ Evasion Attack

→ Exploratory Attack

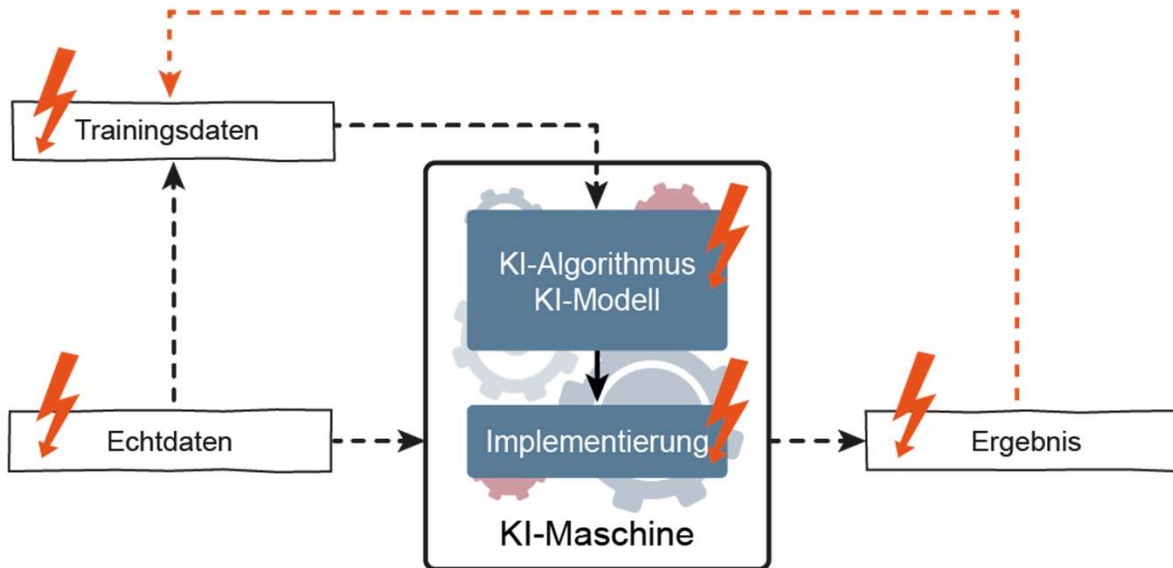
→ ...



Poisoning Attack

IT-Sicherheit für Künstliche Intelligenz

→ IT-Sicherheitsmaßnahmen



Schutzziele:

- **Integrität**
(Erkennen von Manipulation der Daten)
- **Vertraulichkeit**
(Wahrung von Geschäftsgeheimnissen)
- **Datenschutz**
(Schutz von personenbezogenen Daten)
- **Verfügbarkeit**
(der Anwendung und Ergebnisse)

Nutzung einer qualitativ hochwertigen KI-Technologie

Stand der Technik bei den **IT-Sicherheitsmaßnahmen** nutzen

- der **Daten** (Trainingsdaten, Echtdaten/Inputdaten, Ergebnisse),
- der **KI-Maschine** (Algorithmen, Modelle, Implementierung) und
- der **Anwendung**

Der Einsatz von KI basiert auf Vertrauen

→ Vertrauenswürdigkeit schafft Vertrauen

Status Quo: Aufgrund der Digitalisierung erhöht sich der Grad an Komplexität, wodurch es für den Anwender zunehmend schwieriger wird, einzelne **KI-/IT-Sicherheit-Lösungen** und deren Hintergründe **verstehen** und **bewerten** zu können.

Folge: Die **Unsicherheit** macht den Menschen Angst und **schränkt** sie in ihrer **Handlungsfähigkeit ein**.

Fazit: Vertrauen ist unerlässlich, um handlungsfähig sein zu können.

Wichtig: Anwender wollen Vertrauen können.

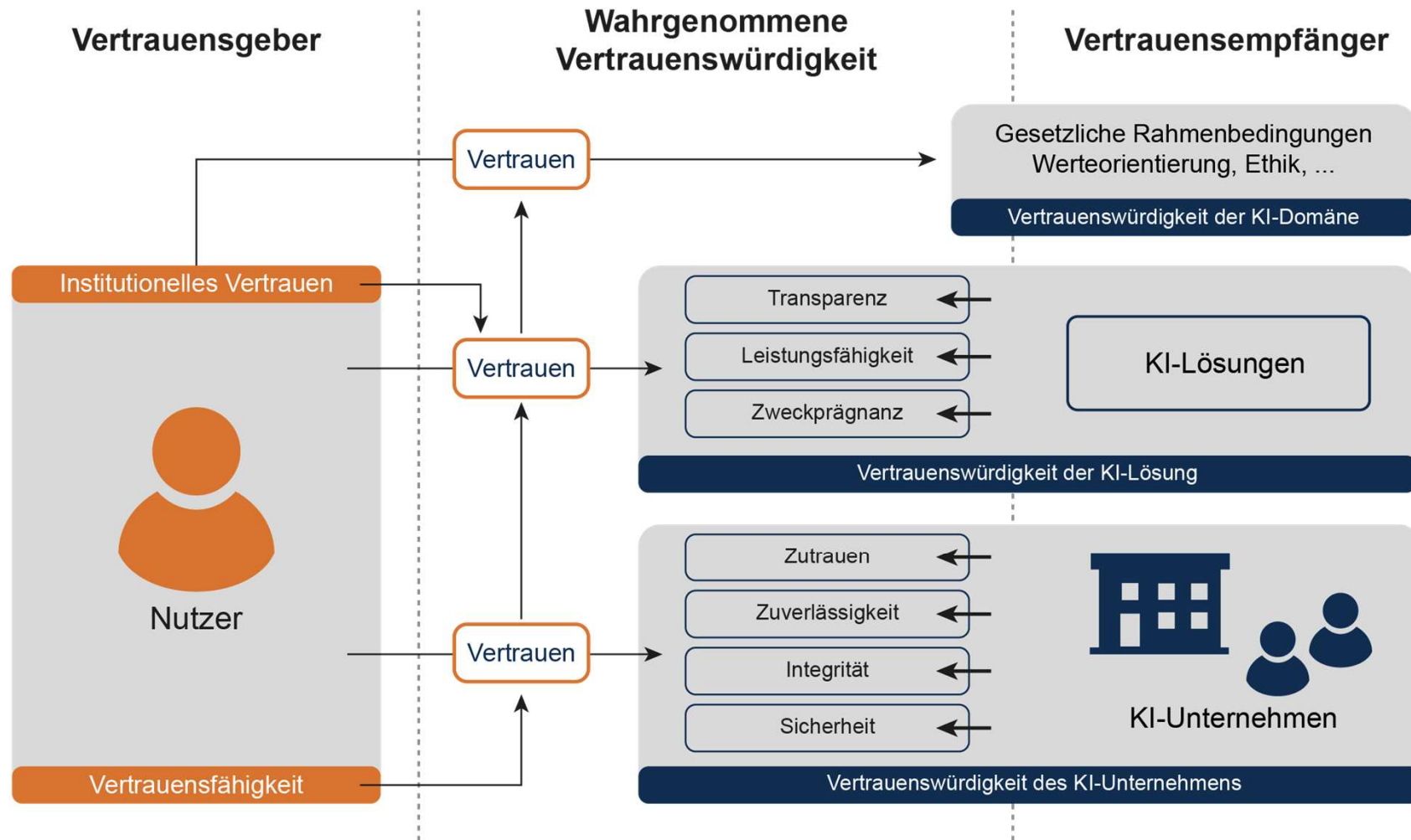
Das bedeutet: KI-/IT-Sicherheit-Hersteller müssen vertrauenswürdig agieren.

Fazit: Damit der Anwender ein gerechtfertigtes Vertrauen in **KI-/IT-Sicherheit-Hersteller** sowie deren **KI-/IT-Sicherheit-Lösung** aufbauen kann, müssen diese alles tun, um ihre **Zuverlässigkeit, Ehrlichkeit und ihre Qualifikation unter Beweis zu stellen**.



Der Einsatz von KI basiert auf Vertrauen

→ Vertrauenswürdigkeitsmodell



Beispiele ethischer Herausforderungen

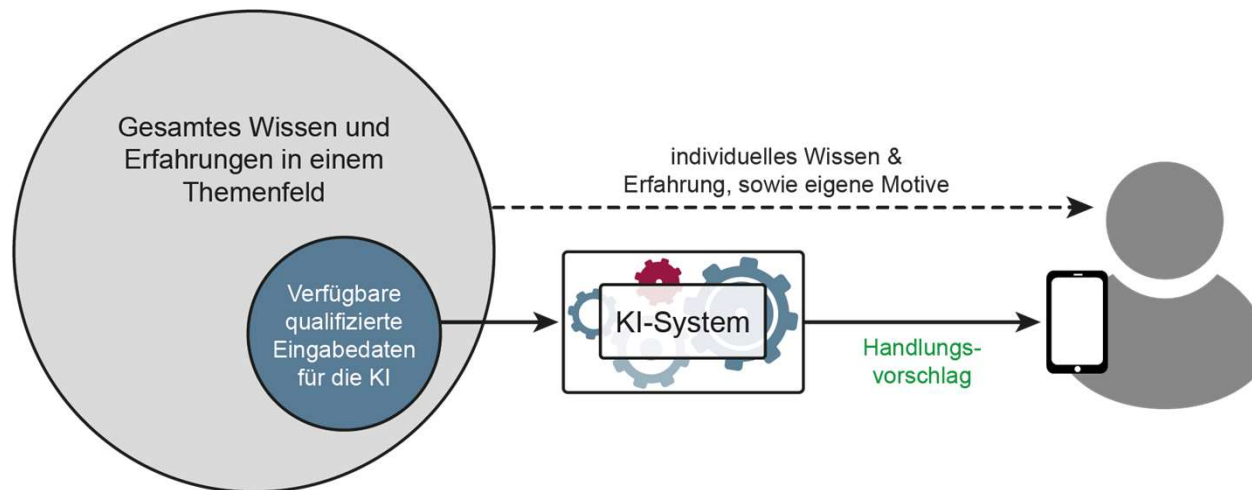
→ Privatheit vs. Allgemeinwohl

- Eine **Notwendigkeit für die Auswertung von personenbezogenen Daten** von Mitarbeitern könnte sich zum Beispiel in dem Fall ergeben, wenn ein **Energieversorger** (Kritische Infrastruktur) **angegriffen wird**.
- **Es besteht das Dilemma:**
Eine ethische Handlungsweise erfordert, dass grundlegende **Rechte wie die Privatheit** nicht zugunsten eines höheren Ziels vollständig aufgegeben werden dürfen.
- Dagegen steht gemäß **Utilitarismus** die **Prämisse des Strebens nach dem Gesamtnutzen für die Gesellschaft**.
- **Daraus resultiert die ethische Fragestellung:**
Wann ist es angeraten die **Rechte des Individuums** zugunsten des **Wohles für die Gemeinschaft** aufzuheben?

Beispiele ethischer Herausforderungen

→ Strike Back: Problem Unvollständigkeit der Daten

- Dem Konstrukt des ‚Strike Back‘ liegt die Hypothese zugrunde, dass ein **Angriff durch einen Gegenangriff beendet** werden kann.
- Die KI kann den Strike Back berechnen und automatisiert durchführen.
- Aber was ist, wenn dies gleichzeitig auch die Stromversorgung eines Krankenhauses lahmlegen würde?
- **Die ethische Frage in diesem Kontext müsste lauten:**
Ist es möglich die Vollständigkeit der Daten zu erreichen und deren Relevanz zu beurteilen damit eine KI die richtige Entscheidung treffen kann?



IT-Sicherheit *und* Künstliche Intelligenz

→ Zusammenfassung und Ausblick

- **Künstliche Intelligenz ist eine sehr wichtige Technologie im Bereich der IT-Sicherheit**
 - Erkennen von Angriffen, Bedrohungen, Schwachstellen, Fake-News ...
 - Unterstützung / Entlastung von IT-Sicherheitsexperten
 - Sichere Software-Entwicklung
 - ...
- **Wir müssen unsere KI absichern, damit sie vertrauenswürdige Ergebnisse liefern kann**
 - Hacker greifen an und manipulieren Daten, Algorithmen/Modelle und Ergebnisse
 - ...
- **Machtverhältnis zwischen Angreifern und Verteidigern muss sich verändern**
 - Die Angreifer nutzen KI sehr professionell und erfolgreich für ihre Angriffe
 - Die Verteidiger müssen KI deutlich intensiver - gemeinsam mit allen Verteidigern - umsetzen
 - Der Austausch von sicherheitsrelevanten Daten, wird helfen, besser zu werden
 - ...



**Westfälische
Hochschule**

Gelsenkirchen Bocholt Recklinghausen
University of Applied Sciences

KI und IT-Sicherheit

→ **Mehr Sicherheit, mehr Gefährdungen**

*Wir brauchen die **KI** in der **IT-Sicherheit**.*

Aber: *Wir müssen auch die **Risiken** der **KI** aktiv **reduzieren!***

Prof. Dr. (TU NN)

Norbert Pohlmann

Professor für Cyber-Sicherheit und Leiter des Instituts für Internet-Sicherheit – if(is), Westfälische Hochschule, Gelsenkirchen

Vorstandsvorsitzender Bundesverband IT-Sicherheit - TeleTrust

Vorstand im Verband der Internetwirtschaft - eco

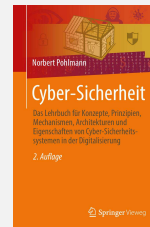
if(is)
internet-sicherheit.

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>



Master Internet-Sicherheit

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



Glossar Cyber-Sicherheit

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



It's all about Trust!

<https://vertrauenswuerdigkeit.com/>



Quellen Bildmaterial

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

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/>

Literatur

- 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
- M. Fourné, D. Petersen, N. Pohlmann: "Attack-Test and Verification Systems, Steps Towards Verifiable Anomaly Detection". In Proceedings der INFORMATIK 2013 - Informatik angepasst an Mensch, Organisation und Umwelt, Hrsg.: Matthias Horbach, GI, Bonn 2013
- U. Coester, N. Pohlmann: „Verlieren wir schleichend die Kontrolle über unser Handeln? Autonomie hat oberste Priorität“, BI-SPEKTRUM Fachzeitschrift für Business Intelligence und Data Warehousing, 05-2015
- U. Coester, N. Pohlmann: „Ethik und künstliche Intelligenz – Wer macht die Spielregeln für die KI?“, IT & Production – Zeitschrift für erfolgreiche Produktion, TeDo Verlag, 2019
- N. Pohlmann: „Künstliche Intelligenz und Cybersicherheit – Diskussionsgrundlage für den Digitalgipfel 2018“
<https://norbert-pohlmann.com/app/uploads/2018/12/Künstliche-Intelligenz-und-Cybersicherheit-Diskussionsgrundlage-für-den-Digitalgipfel-2018-Prof.-Norbert-Pohlmann.pdf>
- 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).

Weitere Artikel siehe: <https://norbert-pohlmann.com/artikel/>