Westfälische Hochschule
Gelsenkirchen Bocholt Recklinghausen
University of Applied Sciences

if(is) internet-sicherheit.

# ENQUETE-KOMMISSION KÜNSTLICHE INTELLIGENZ

## Sicherheit und Vertrauenswürdigkeit von KI-Systemen

## **Thesen** und **Handlungsempfehlungen**


Prof. Dr. (TU NN)
**Norbert Pohlmann**

***Professor** für **Informationssicherheit***
***Leiter** des **Instituts für Internet-Sicherheit - if(is)***
***Vorstandsmitglied** des **Verbands der Internetwirtschaft - eco***
***Vorstandsvorsitzender** des **Bundesverbands IT-Sicherheit - TeleTrusT***

**Institut für Internet-Sicherheit - if(is)**
Westfälische Hochschule
Neidenburger Straße 43
45877 Gelsenkirchen
E-Mail: pohlmann@internet-sicherheit.de
Web: www.internet-sicherheit.de, www.norbert-pohlmann.com
Tel: +49 173 3021 838

## 1. These: Vertrauenswürdigkeit

***Vertrauenswürdigkeit ist der Schlüssel zum zukunftsfähigen Erfolg von Künstlicher Intelligenz (KI)***

Vertrauen wird als die subjektive Überzeugung von der Richtigkeit einer Aussage und von Handlungen verstanden. Daher wird ein KI-System von uns als vertrauenswürdig eingestuft, wenn es sich für den vorgesehenen Zweck immer in der erwarteten Weise verhält.

Diese Vertrauenswürdigkeit kann bei KI-Systemen dann aufgebaut werden, wenn

1.) die Eingangsdaten der KI eine hohe Qualität für den Anwendungsfall aufweisen,
2.) die IT-Anwendung und das genutzte KI-System von KI- und Anwendungsexperten konzipiert sowie manipulationssicher und vertrauenswürdig umgesetzt,
3.) die Nachvollziehbarkeit der Ergebnisse ermöglicht und
4.) ethische Grundsätze eingehalten werden.

Vertrauenswürdigkeit kann zu einer höheren Akzeptanz von Künstlicher Intelligenz führen.

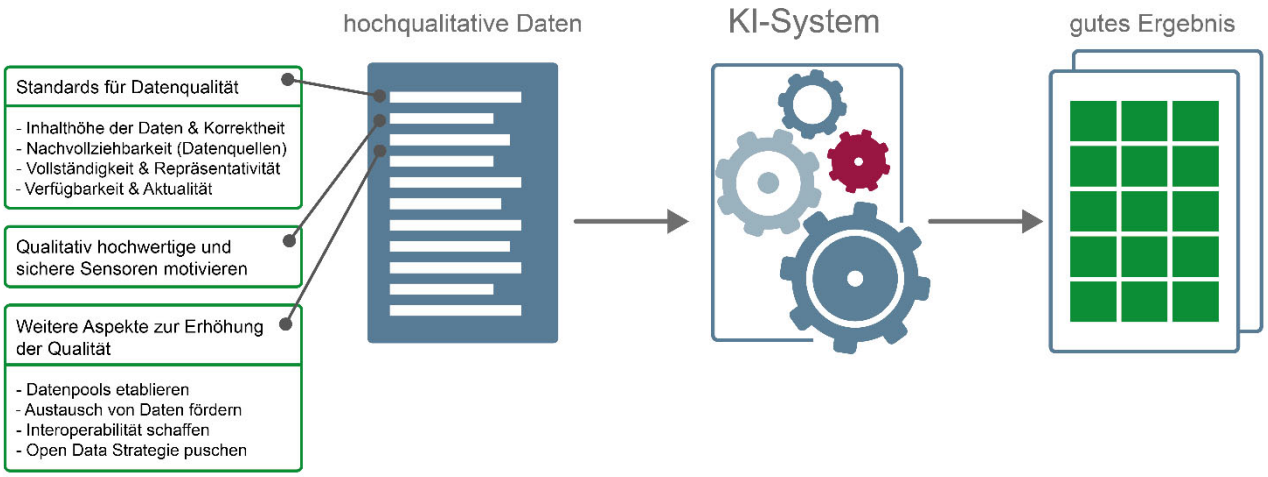


Abb. 1: Qualitative KI-Anwendung

## Handlungsempfehlungen: Vertrauenswürdigkeit

### 1.) Aufklärung

Wir brauchen für die Bildung der Vertrauenswürdigkeit eine **sachliche Aufklärungsarbeit** bezüglich der Schaffung von Verständnis durch Transparenz der Chancen und Risiken der KI-Technologie.

### 2.) Qualität der genutzten Daten

Wir müssen **Standards für die Datenqualität** von KI-Systemen definieren und etablieren. Kriterien für Datenqualität sind unter anderem Vollständigkeit, Repräsentativität, Verfügbarkeit, Nachvollziehbarkeit, Aktualität und Korrektheit.

Zur Umsetzung einer hohen Datenqualität ist auch die **Förderung** von qualitativ hochwertigen und sicheren **Sensoren** notwendig. Auch Daten aus anderen Quellen müssen sorgfältig und möglichst nachvollziehbar erhoben und vor der Nutzung verifiziert werden.

Darüber hinaus müssen wir, in einer KMU-orientierten Wirtschaft, hochwertige **Datenpools** etablieren, den **Austausch von Daten** fördern, eine **Interoperabilität von Daten** schaffen und Open Data Strategien vorantreiben, um bessere Ergebnisse und damit eine höhere Vertrauenswürdigkeit zu erreichen.

### 3.) Sichere und vertrauenswürdige Umsetzung von KI-Systemen

Wir brauchen **qualitativ hochwertige KI-Technologien**, um das Vertrauen in die Nutzung zu steigern (z.B. eigene Technologie „Made in Germany“).

Wir müssen die **Zusammenarbeit von erfahrenen Experten** der jeweiligen **Anwendungsdomäne** sowie der **KI-Entwicklung** und **IT-Sicherheit** fördern, die für das entsprechende Anwendungsgebiet (Autonomes Fahren, personal Assistent, Cyber-Sicherheit, …) KI-Systeme vertrauenswürdig konzipieren und umsetzen können.

Außerdem müssen wir aktiv den **Stand der Technik** an **IT-Sicherheitsmaßnahmen** für Integrität, Vertraulichkeit, Datenschutz und Verfügbarkeit definieren und nachhaltig umsetzen, um die Manipulationsmöglichkeiten und den Missbrauch der KI-Anwendungen und der genutzten Daten zu minimieren.

### 4.) Nachvollziehbarkeit von Entscheidungen

Wir sollten den **Menschen als kontrollierenden Faktor** in den Kreislauf der KI einbinden („Keep the human in the loop“). Das Ergebnis von KI sollte als Handlungsempfehlung verstanden werden. Der Mensch kann immer noch entscheiden, ob er dieser folgt oder auch nicht. Damit wird die Selbstbestimmtheit der Nutzer gefördert und die Vertrauenswürdigkeit erhöht.

Wenn KI in automatisierte IT-Systeme eingebunden wird, müssen **Tests, Simulationen und Validierungen** einen **höheren Stellenwert** und die Definition der **Verantwortung** sowie die daraus resultierende **Haftung** eine **besondere Bedeutung** haben.

## 2. These: Souveränität

***Wir brauchen eine leistungsfähige KI-Infrastruktur zur Aufrechterhaltung der digitalen Souveränität.***

KI ist eine **Schlüsseltechnologie** für das zukünftige **Wirtschaftswachstum** in allen Branchen und Bereichen. Dabei ist die **Datenhoheit** ein entscheidender Faktor bei der Verwendung von KI-getriebenen Technologien. Außerdem ist die Verfügbarkeit von **leistungsstarken KI-Infrastrukturen** wichtig, um die KI-Anwendungen erfolgreich, sicher, qualitativ und souverän umsetzen zu können. Aber auch die Motivation eines KI-Ökosystems durch KMU/Startup Unterstützung/Förderung, Ausbildungsinitiativen, Internationalisierungskampanien, usw. sind wichtig für die Aufrechterhaltung der digitalen Souveränität.

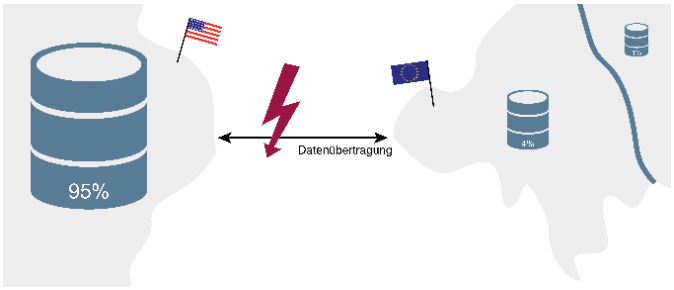


Abb. 2: Verteilung der genutzten KI-Daten

## Handlungsempfehlungen: Souveränität

### 1.) Verfügbarkeit der Daten

Es muss dafür gesorgt werden, dass die erhobenen **Daten** zunehmend im **deutschen/europäischen Raum gespeichert** und **verarbeitet** werden, sodass die aktuelle und existenzgefährdende Abhängigkeit von außereuropäischen Unternehmen auf ein vertretbares Maß reduziert wird.

Hier müssen Konzepte entwickelt werden, besonders kritische Daten identifizierbar zu machen, die ihrerseits nicht nur dezentral an europäischen Standorten gespeichert, sondern auch zusätzlich von öffentlichen Stellen vorgehalten werden. Damit kann verhindert werden, dass eine Verweigerung der Herausgabe der in anderen Hoheitsgebieten gespeicherten Daten als Druckmittel (z.B. Cyberwar) verwendet werden kann. Nur so kann die **Verfügbarkeit der Daten** unabhängig der politischen und wirtschaftlichen Lage gewährleistet werden.

### 2.) Leistungsstarke KI-Infrastruktur

Außerdem muss der Aufbau einer **leistungsstarken KI-Infrastruktur** gefördert werden, um souverän das Wirtschaftswachstum in allen Branchen und Bereichen ermöglichen und fördern zu können.

Dann sind wir auch in der Lage, KI als Schlüsseltechnologie sicher und qualitativ für den zukünftigen Erfolg unserer Gesellschaft nutzen zu können.

### 3.) Unterstützung des Mittelstands

Wir brauchen Förderungen zur **Unterstützung des Mittelstands**, damit dieser die KI optimal, qualitativ und souverän nutzen kann, um den zukünftigen Erfolg zu garantieren. Dazu sollten geeignete **KI-Kompetenzzentren** mit **leistungsfähige KI-Infrastrukturen** und **qualitativ hochwertigen Daten** etabliert werden.

### 4.) Gewinnbringende Nutzung von persönlichen Daten

Wir sollten Konzepte erarbeiten, wie die **persönlichen Daten** auf den großen IT-Plattformen der Welt (Social Media, Suchmaschinen, Einkauf, …) **nutz-** oder **gewinnbringend** für unsere Bürgerinnen und Bürger **genutzt werden können**.

## 3. These: Missbrauch

***KI-Systeme werden missbraucht, um der Gesellschaft und einzelnen Personen gezielt zu schaden.***

Angriffsvektoren, die wie Phishing auf den Faktor Mensch abzielen, können durch KI-Systeme hinsichtlich ihrer schädlichen Auswirkung von Angreifern optimiert werden. Die Opfer können weitaus effizienter als bisher analysiert und mögliches Verhalten vorhergesagt werden. Autonome Agenten, wie Chatbots können in sozialen Netzwerken selbstständig auf Ereignisse reagieren, um Meinungen von Menschen zu beeinflussen und Gesellschaften gezielt zu manipulieren. Bildgenerierende Systeme können überzeugend gestellte Videos (Deep-Fakes) erzeugen und damit Individuen oder ganze Personengruppen diffamieren, zu Gewalt aufrufen und Chaos anstiften.

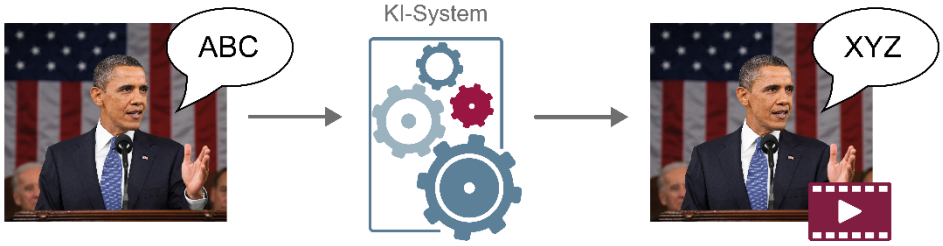


Abb. 3: Verteilung der genutzten KI-Daten

## Handlungsempfehlungen: Missbrauch

### 1.) Definition von Verantwortung und Haftung

Wir brauchen eine klare Zuordnung von Verantwortlichkeiten und Haftung. Es sollte geprüft werden, ob Regelungslücken bestehen (beispielsweise bei Manipulation und absichtlich schadhafter Verwendung von KI-Systemen). Auf Entwicklungsebene kann ein Missbrauch von KI-Technologien nicht unterbunden werden, da beispielsweise bildgenerierende Systeme für die Filmproduktion und Deep-Fake eingesetzt werden oder Verteidiger, aber auch Angreifer, KI-Technologien zunehmend nutzen (Dual-Use). Daher kann hier nur eine gesetzliche Regulierung angesetzt werden, die eine absichtlich schadhafte Verwendung entsprechender KI-Systeme unter Strafe stellt. Nur bei einer Endanwendung, bei der die Verwendung keine Zweifel über die Absichten zulässt, sollte der Hersteller dafür in Haftung genommen werden. Des Weiteren sollten Personen in Haftung treten, die KI-Systeme nutzen, die ursprünglich nicht als schädigend eingestuft wurden, aber so eingesetzt werden, dass sie Schäden anrichten.

### 2.) Aufklärungsarbeit und Medienkompetenz

Auch hier gilt, dass **Aufklärungsarbeit** geleistet und Medienkompetenz in Bezug zur Künstliche Intelligenz gefördert werden muss, um eine Schadensminimierung bei bereits in Umlauf geratenen KI-Systemen bzw. Produkten zu erzielen.

### 3.) Verifizierung der von KI-Systemen verwendeten Daten

Sinnvolle Konzepte zur **Verifizierung** von Ein- und Ausgabedaten für KI-Systeme müssen unmittelbar erforscht und implementiert werden. Diese IT-Sicherheitsmaßnahme zielt auf die verbesserte Analyse und letztendlich die IT-Sicherheit und Stabilität der KI-Systeme durch die Reduzierung der Angriffsflächen von nutzenden Menschen und Maschine ab.

### 4.) Verifizierung des Verantwortlichen für die Ergebnisse

Wir müssen IT-Sicherheitskonzepte erarbeiten und umsetzen, um den Verantwortlichen der Ergebnisse verifizierbar zu machen. Das kann auf der Basis von Vertrauensdiensten, wie PKI und Blockchain unter der Nutzung von digitalen Signaturen umgesetzt werden.

## 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

D. Petersen, N. Pohlmann: „Kommunikationslage im Blick - Gefahr erkannt, Gefahr gebannt“, IT-Sicherheit – Management und Praxis, DATAKONTEXT-Fachverlag, 4/2014

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: „Diskriminierung und weniger Selbstbestimmung? Die Schattenseiten der Algorithmen“, tec4u, 12/17

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

N. Pohlmann: Lehrbuch „Cyber-Sicherheit“, Springer Vieweg Verlag, Wiesbaden 2019
ISBN 978-3-658-25397-4s

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