Julius-Maximilians-
UNIVERSITÄT
WÜRZBURG

Institut für Informatik
Lehrstuhl für Kommunikationsnetze
Prof. Dr. T. Hoßfeld

# Handlungsempfehlungen (MVP): Sichere und digital souveräne Nutzung generativer KI-Systeme

Ralf Schweifler, Prof. Dr. Harald Wehnes

Stand: 22.01.2024

ARBEITSKREIS
OPEN SOURCE SOFTWARE

# Agenda

- Motivation und Ziel
- Herausforderungen und Risiken
- Strategische Management-Entscheidungen
- Auswahl sicherer und digital souveräner KI-Systeme
- Handlungsempfehlungen für Unternehmen, staatliche Einrichtungen und Individuen (MVP)
- Ausblick
- Diskussion

# Referenten

## Ralf Schweifler

**Masterstudent der Informatik am Institut für Informatik der Universität Würzburg**
**Masterarbeit (in Bearbeitung)**
**„Digitale Nachhaltigkeit: Betrachtung und Evaluation von digitaler Souveränität in Verbindung mit der Nutzung von KI“**

## Prof. Dr. Harald Wehnes

**Institut für Informatik der Universität Würzburg**
**Lehr- und Forschungsschwerpunkte: Digitale Nachhaltigkeit, Modernes Projektmanagement, KI in der Projektwirtschaft, Digitale Startups**

**Seit 1.1.2024 Mitglied des Präsidiums der GI**

# Motivation und Ziel

Ziel der Masterarbeit:

**Sicherer und digital souveräner Einsatz von KI-Systemen in Deutschland**

→ Hilfestellung für Unternehmen, staatliche Einrichtungen und Individuen

KÜNSTLICHE INTELLIGENZ

**Samsung-Ingenieure geben ChatGPT vertrauliche Daten preis**

Samsung hat ermittelt, dass drei seiner Ingenieure binnen kurzer Zeit mindestens dreimal bei der Nutzung von ChatGPT geheime Daten durchsickern lassen haben.

in Pocket speichern | merken

5. April 2023, 7:15 Uhr, Andreas Donath

**Angst vor Diebstahl von Daten und Code:**
Selbst große Tech-Player schränken den Einsatz von **ChatGPT und anderen KI-Tools** in ihren eigenen Unternehmen für ihre Mitarbeiter ein oder untersagen ihn sogar !!!

Generative AI

COMPUTERWOCHE
VOICE OF DIGITAL

Generative AI | Hintergrund | Ratgeber | Bilder | Video | News

Weltwirtschaftsforum 2024

der Tech-Monopole

**Fehlinformation durch KI größtes globales Risiko**

15.01.2024 Von Jürgen Hill (Chefreporter Future Technologies) FOLGEN

**Auf dem Weltwirtschaftsforum in Davos zeigt sich ein ambivalentes Verhältnis zu Thema KI. Die einen sehen in KI eines der größten globalen Risiken, andere loben die positiven Auswirkungen von KI auf die Wirtschaft.**

BUSINESS INSIDER

TECH

**Amazon, Apple, and 12 other major companies that have restricted employees from using ChatGPT**

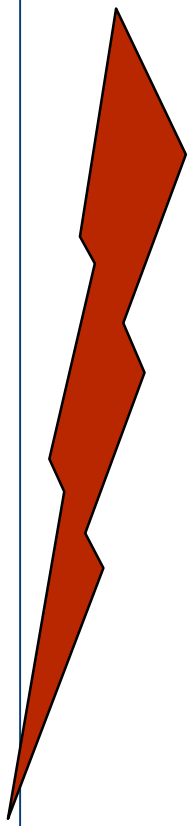
Aaron Mok Jul 11, 2023, 5:38 PM MESZ

# Herausforderungen und Risiken

- **KI bietet unzählige neue Möglichkeiten und gewaltige wirtschaftliche Potentiale**
- **Unbedachter KI-Einsatz führt zu Risiken und Gefahren, wie z.B. Abfluss und Missbrauch von Unternehmensinterna und Know-How, Verbreitung von Falschinformationen, Vertrauensverlust, Gefährdung der Demokratie, irreversible digitale Abhängigkeit etc.**

**Risiken beim Einsatz von KI-Systemen (unvollständige Liste)**

- *Halluzinationen bei generativer KI*
- *Intransparente (manipulierte?) Datenpools*
- ***Verlust hochwertiger Unternehmensdaten***
- *Kontrollverlust über die im Rahmen der Nutzung eingegebenen Daten*
- *Verlust von geistigem Eigentum und Urheberrechtsprobleme*
- *Neue „Erkenntnisse“ auf Basis falscher Informationen*
- *Erhalt und Verbreitung von Vorurteilen und Lügen, die sich wie Wahrheiten anfühlen*
- *Verlust von Datenschutz und Datensouveränität*
- *Gefährdung der Cyber security*
- ***Verstärkte digitale Abhängigkeit – insbesondere von Big-Tech-Monopolisten – → wirtschaftliches und politisches Erpressungspotential***



# Halluzination

**Du**

Wer war der erste Mensch der den Ärmelkanal zu Fuß durchlaufen hat?

**ChatGPT**

Der erste Mensch, der den Ärmelkanal zu Fuß durchquert hat, war der französische Ingenieur und Erfinder Albert Dauzat. Er führte diesen bemerkenswerten Fußmarsch im Jahr 1954 durch, indem er die Passage zwischen Sangatte in Frankreich und Dover in England wagte. Es ist jedoch wichtig zu beachten, dass solche Fußdurchquerungen des Ärmelkanals äußerst selten und gefährlich sind, da der Kanal stark befahren und die Strömungen unberechenbar sind. Es handelt sich um eine äußerst herausfordernde und gefährliche Unternehmung, die nur von erfahrenen und gut vorbereiteten Abenteurern unternommen werden sollte.

# Strategische Management-Entscheidungen

**VOR** dem Einsatz von KI-Systemen: **Strategische Entscheidungen des TOP-Managements** (Vorstände, Geschäftsführer, Aussichtsräte u.ä.) notwendig

- ✓ **Zielsetzung und Einsatzgebiete definiert**
- ✓ Organisation (KI-Verantwortlicher im Unternehmen) und Rollen mit KI-Kompetenz ernannt
- ✓ Technische Voraussetzungen – mit Verantwortlichkeiten – vorhanden
- ✓ **Auswahl KI-System(e) – Kriterien: Leistungsfähigkeit, Ethik, Sicherheit, digitale Souveränität**
- ✓ **Governance-Konzept zum KI-Einsatz und Umgang mit KI erstellt (Vorgaben zum KI-Einsatz)**
- ✓ Strategie zum Change-Management auf allen Managementebenen kommuniziert und organisiert
- ✓ **Mitarbeiter mitnehmen**: Unternehmenskommunikation und Schulung durchgeführt
- ✓ (Agiles) Vorgehensmodell zur laufenden Nachsteuerung eingerichtet

# Auswahlkriterien für sichere und digital souveräne KI-Systeme

## Auswahlkriterien

- **Monopolstellung und Wettbewerb → Der Anbieter darf keine Monopolstellung am digitalen Markt besitzen, um den Weg in die digitale Kolonie nicht weiter zu beschleunigen**
- **Quelloffenheit** → „Open Source AI“ steigert Transparenz und Vertrauen
- **Standardisierte Schnittstellen** → KI-System kann in verschiedene Anwendungen integriert werden, KI-System-Wechsel und Interoperabilität werden unterstützt
- **Rechtsstandort** → Nur Anbieter, deren juristischer Hauptsitz in der EU liegt, ermöglichen „wirkliche“ Datensouveränität!
- **On-premises Unterstützung** → Ein KI-System, das On-premises betrieben werden kann, ermöglicht Datensouveränität!

## KI-Nutzungsarten

- **Interne KI-Tools:** Das KI-System wird von der Unternehmens-IT bzw. ähnlichen IT-Organisationen für staatliche Einrichtungen „On-premises“ (in eigenen Räumlichkeiten) betrieben
- **Externe KI-Tools:** Das KI-System wird außerhalb des Unternehmens auf einer Public Cloud bereitgestellt; Beispiele: ChatGPT, BARD, MidJorneey, YOU

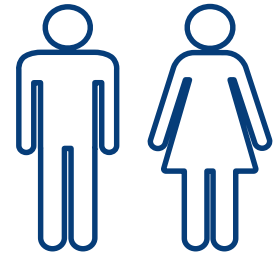
# Handlungsempfehlungen für Unternehmen (Auszug)

- **Auswahl der KI-Tools**
  - Nutzung eines internen KI-Systems → Bestmögliche Datensouveränität!
  - Externes KI-System: möglichst von europäischen Anbietern, gehosted auf Systemen, die in der EU stehen, von Providern mit juristischem Hauptsitz in der EU
- **Nutzung externer KI-Systeme**
  - Keine Eingabe sensibler Daten („Kronjuwelen“ des Unternehmens), wie Unternehmensinterna, Personaldaten, Vertrags- und Ausschreibungsdaten, Passwörter, Kundendaten, Geschäftspartnerdaten, Nutzernamen, Code, interne Dokumente etc.
- **Mitarbeiterschulungen**
  - Schaffung von Awareness und korrektem KI-Einsatz → externe KI-Systeme wie Fremdsysteme behandeln
  - **KI als sekundäre Quelle verwenden**
  - Wahrung der Unternehmensinterna, Vertraulichkeit und Datenschutz
- **Umgang mit erhaltenen KI-Antworten**
  - **Verantwortung hat der Nutzer der KI → Ergebnisse sind immer kritisch zu überprüfen!**
  - Bewusstsein für Voreingenommenheit der KI → *Halluzinations-Check der Ergebnisse*
  - Generierte Inhalte, die veröffentlicht werden sollen, müssen durch einen erfahrenen Experten geprüft werden, bspw. Senior-Entwickler für QA bei Software

# Handlungsempfehlungen für staatliche Einrichtungen (Auszug)

- **Auswahl der KI-Tools**
  - Nutzung interner KI-Systeme stets bevorzugen
  - Es dürfen ausschließlich KI-Systeme europäischer Anbieter zum Einsatz kommen, gehosted auf Systemen, die in der EU stehen, von Providern mit juristischem Hauptsitz in der EU
  - AI Act und DSGVO müssen auf **allen Ebenen** eingehalten werden
- **Mitarbeiterschulungen**
  - Verpflichtende Schulung vor Verwendung von KI
  - Bewusstsein schaffen, dass Eingaben von Bürgern und Dritten über KI generiert wurden und dass diese möglicherweise Halluzinationen (vermeintliche Fakten) enthalten
- **Umgang mit erhaltenen KI-Antworten**
  - **Verantwortung hat der Nutzer der KI → Ergebnisse sind immer kritisch zu überprüfen!**
  - Bewusstsein für Voreingenommenheit der KI → *Halluzinations-Check der Ergebnisse*
  - **Generierte Inhalte**, die veröffentlicht werden sollen, **müssen validiert werden!**
  - KI als sekundäre Quelle verwenden
- Chatbots, die zur Nutzung von Bürgern eingesetzt werden, müssen vor Veröffentlichung getestet und verwendete Modelle und KI-Systeme transparent kommuniziert werden!

# Handlungsempfehlungen für Individuen (Auszug)



- **Auswahl der KI-Tools**
  - Die KI-Tools sollten mit Blick auf die eigenen Ziele gewählt werden
  - KI-Systeme europäischer Anbieter, gehosted auf Systemen, die in der EU stehen, von Providern mit juristischem Hauptsitz in der EU, sollten bevorzugt eingesetzt werden
- **Nutzung externer KI-Systeme**
  - **Zu beachten**
    - Bewusstsein für Voreingenommenheit der Ergebnisse → *Halluzinations-Check der Ergebnisse*
    - KI als sekundäre Quelle verwenden
    - Keine Weitergabe falscher Informationen, Propaganda u.ä.
    - Rassistische, diskriminierende oder unpassende Inhalte stets den KI-Betreibern melden
  - **Potenziell nützlicher Einsatz**
    - Recherchen, Informationsbeschaffung, Ideensammlung, Brainstorming, Formulierungshilfen, Erstellung von Entwürfen. Zusammenfassungen, etc.
    - Weiterbildung: KI-Tool als „Tutor“, z.B. Erklärung von Inhalten
  - **Nicht empfehlenswert**
    - Übermäßige und unkritische Nutzung von KI-Tools
      → Orientierungsverlust in der digitalen Welt (wie bei Navigationssystemen)
      → Gefahren: Verlust von Beurteilungs-, Handlungs- und Entscheidungskompetenz
      → **Manipulierbarkeit**

# Ausblick

- Feedback zu den „Handlungsempfehlungen für die Praxis“ gewünscht:
  - Ralf.schweifler@stud-mail.uni-wuerzburg.de
  - Wehnes@informatik.uni-wuerzburg.de
- Bereitstellung der Handlungsempfehlung ab 01.03.2024 auf Plattform https://digital-sovereignty.net/


Ausbau der Plattform:
**Marktplatz** für digital souveräne Produkte, insbesondere KI-Systeme

# Diskussion

- Weitere Handlungsempfehlungen?
- Empfehlung für KI-Tools!

# Quellen


- https://www.golem.de/news/kuenstliche-intelligenz-samsung-ingenieure-leaken-interne-daten-an-chatgpt-2304-173220.html
- https://www.businessinsider.com/chatgpt-companies-issued-bans-restrictions-openai-ai-amazon-apple-2023-7
- https://www.computerwoche.de/a/fehlinformation-durch-ki-groesstes-globales-risiko,3698241?utm_source=related_right
- https://www.golem.de/news/halluzination-chatgpt-erfindet-gerichtsakten-2305-174509.html
- https://www.zeit.de/digital/2023-10/bing-ki-microsoft-bayern-landtagswahl-falschinformationen
- https://www.zeit.de/zeit-wissen/2015/02/orientierung-verlust-navigationsgeraete

- Bernert, C.; Scheurer, S.; Wehnes, H. (2024): KI in der Projektwirtschaft, UVK Verlag, München

# KI in der Projektwirtschaft: Praxisbeiträge für Band 2 gesucht

- Band 1 seit Mitte Januar 2024 im Buchhandel erhältlich - 25 Beiträgen von 50 Autor:innen

- Call for Participation (CfP) für Band 2 läuft
- Fokus von Buch 2: Konkrete Anwendungen von KI in der Projektwirtschaft / Berichte aus der Praxis
- Wer Interesse hat, sich mit einem Praxis-Beitrag zu beteiligen, ist herzlich eingeladen
- Mail an wehnes@informatik.uni-wuerzburg.de