Bundesamt
für Sicherheit in der
Informationstechnik

Deutschland
Digital•Sicher•BSI•

# Herausforderungen der Künstlichen Intelligenz im Kontext der Cybersicherheitsentwicklung

Claudia Plattner, Präsidentin BSI

19. April 2024

# Kompetenzzentrum Künstliche Intelligenz des BSI

## IT-Sicherheit für KI

Wir untersuchen neuartige Bedrohungen für KI-Systeme und entwickeln geeignete Gegenmaßnahmen

## Angriffe durch KI

Wir verfolgen neue KI-gesteuerte und KI-unterstützte Angriffsmethoden gegen IT-Systeme und Infrastrukturen und entwickeln geeignete Gegenmaßnahmen


## IT-Sicherheit durch KI

Wir ermöglichen die Nutzung von KI-Methoden zur Verbesserung der IT-Sicherheit, z. B. zur Prävention, Detektion und Reaktion bei Cyber-Angriffen

## KI und digitaler Verbraucherschutz

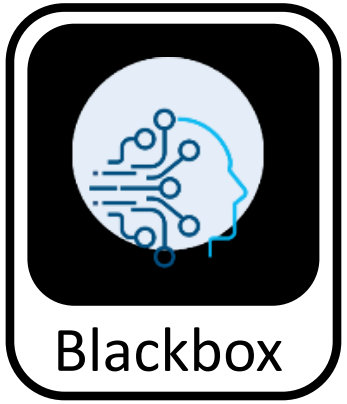
Wir fördern den sicheren und transparenten Einsatz von KI-Methoden in Verbraucherprodukten und steigern die Beurteilungsfähigkeit der Verbraucherinnen und Verbraucher für KI-basierte Systeme

## Normen und Standards für KI

Wir entwickeln und bewerten Prüfkriterien, Prüfmethoden und Prüfwerkzeuge für nachweisbar sichere und vertrauenswürdige KI-Systeme mit dem Ziel, Normen und Standards für diese Systeme zu entwickeln

# Herausforderungen im Bereich KI

## Technologische Herausforderungen




- Spezielle KI-Eigenschaften sowie Komplexität erschweren die Sicherheitsbewertung
- Hochdynamisches Entwicklungsfeld macht stetige Entwicklung und Anpassung von Gegenmaßnahmen für Angriffe notwendig
- Ein sicherer Betrieb erfordert etablierte Best Practices und Standards

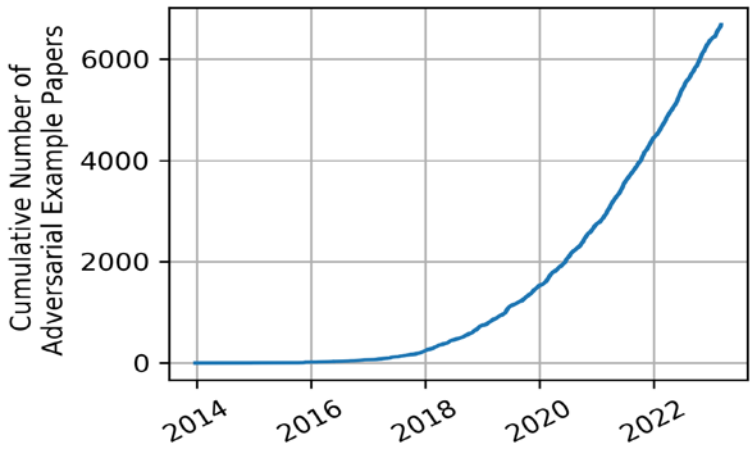
**Sicherheit**, Transparenz, Erklärbarkeit, Verlässlichkeit, Robustheit und Nicht-Diskriminierung stellen beim Einsatz von KI generell eine Herausforderung dar

## Geopolitische Herausforderungen

- Kritische Abhängigkeiten im Bereich Software und Hardware von Anbietern aus nicht EU-Staaten
- Staatlicher Einsatz von generativer KI im Bereich Desinformation

## Personelle Herausforderungen

- Qualifiziertes Fachpersonal ist begrenzt verfügbar
- Wachsender KI-Einsatz steigert Bedarf an Experten
- Austausch und Vernetzung von Spezialisten sowie Bündelung der Kräfte werden immer wichtiger




Quelle: Nicholas Carlini, Scientist @ Google DeepMind

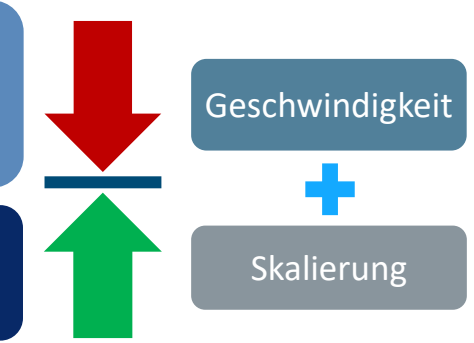
# Einfluss von KI auf die Cyberbedrohungslage
## Chancen für Angreifer und Verteidiger

### Beispiel: KI-assistierte Ransomware-Angriffe

- KI-Programmierassistenten
  - **beschleunigen** Erstellung von Schadcode
  - verringern benötigtes Vorwissen
- KI **beschleunigt** die Analyse und Ausnutzung von Schwachstellen
- KI ermöglicht **skalierbare** Verteilung von Schadsoftware durch hochwertiges & individualisiertes Social Engineering
- KI **beschleunigt** Auswertung exfiltrierter Daten
- KI „unterstützt“ Betroffene bei Bezahlung des Lösegelds in Kryptowährung

**Kein Zukunftsthema: Angreifer nutzen KI bereits heute**
-> KI ermöglicht **höhere Produktivität** (Geschwindigkeit von Angreifern)
-> KI ermöglicht **skalierbare Angriffe** (über Sprachbarrieren hinweg)

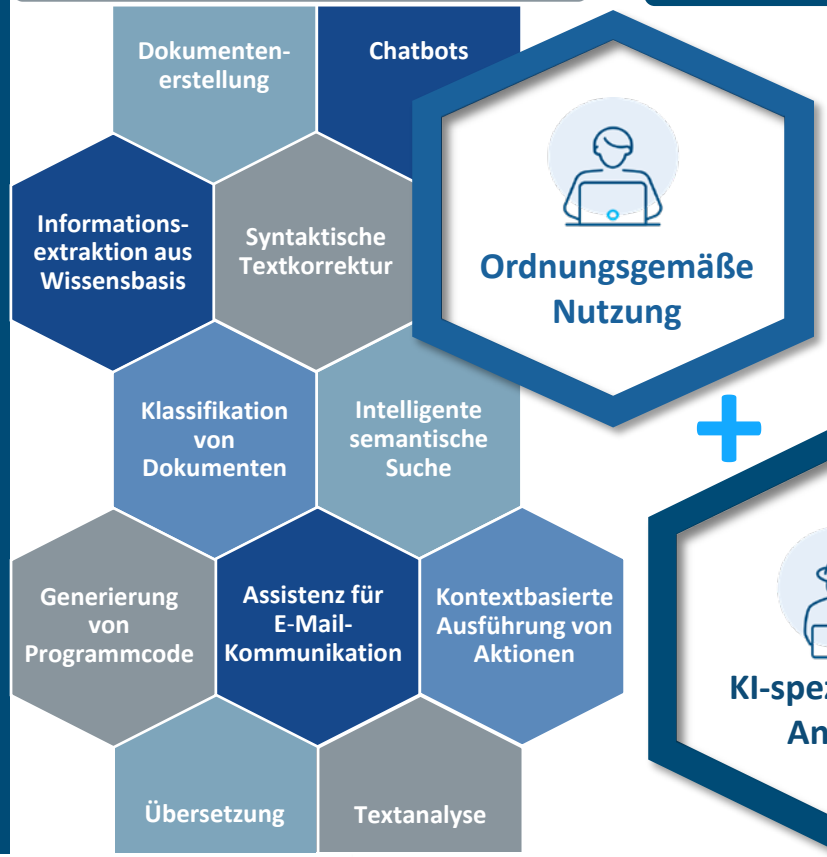
**Cybersicherheit muss Top-Priorität sein**
**KI-Produktivitätsgewinn auch für Verteidigung nutzen**




### Einflussfaktoren

# Sicherheitsherausforderungen in Bezug auf Nutzung von großen KI-Sprachmodellen

**Vielfältiges Nutzungspotential**




**Sicherheitsherausforderungen**

**Ordnungsgemäße Nutzung**

- Fehlende Faktizität und Halluzinieren
- Fehlende Erklärbarkeit
- Eingeschränkte Reproduzierbarkeit
- Vertraulichkeit eingegebener Daten

+

**KI-spezifischer Angriff**

- Adversarial attacks
- (Indirect) prompt injections
- Poisoning attacks
- Privacy attacks

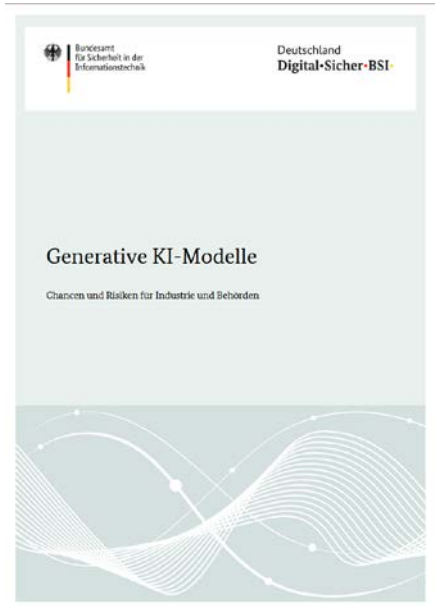
Bundesamt für Sicherheit in der Informationstechnik

Nationales IT-Lagezentrum BSI

SCHWACHSTELLE | GEFÄHRDUNG | VORFALL | IT-ASSETS

Indirect Prompt Injections - Intrinsische Schwachstelle in anwendungsintegrierten KI-Sprachmodellen

*CSW-Nr. 2023-249034-1032, Version 1.0, 18.07.2023*




**Technologiebedingte Schwachstelle**

**Systematische Risikoanalyse durchführen**

**Mitigations-maßnahmen beachten**

# Zusammenarbeit und Unterstützungsmöglichkeiten

**Unterstützungsmöglichkeiten des BSI für Landesverwaltungen**

**Auf Dauer angelegte Kooperation**

- **Verstetigte** und **vereinheitliche Unterstützung** der Länder zu hochdynamischen **KI-Entwicklungen**
- Bereitstellung von **Informationen** und **Tools**
- Projektbegleitende **KI-Sicherheitsberatung**
- **Austausch** von Best Practices und **Einbezug** und **Übersicht** zu Entwicklungen aus der **Forschung**
- Bereitstellung und Entwicklung von praxisnahen **Bewertungsverfahren** für KI
- Länderübergreifende Koordination zur **Mitgestaltung** von kommenden **KI-Normen und KI-Standards**

**Fehlender Rechtsrahmen**

- **Bündelung von Expertise** (**KI**, Cloud, Rechenzentrumssicherheit, Hochverfügbarkeit)
- **Lagebilder** für Länder
- **Austausch-Plattform** in Echtzeit
- Prüfung Hard-/Softwareprodukte
- **Detektion** für Ländernetze
- Koordinierung **IT-Sicherheitsvorfälle**

**Zentralstelle**

# Vielen Dank für Ihre Aufmerksamkeit!

**Kontakt**

Claudia Plattner
Präsidentin

Claudia.Plattner@bsi.bund.de
https://de.linkedin.com/in/claudiaplattner

Bundesamt für Sicherheit in der Informationstechnik (BSI)
Godesberger Allee 87
53175 Bonn
www.bsi.bund.de


Bundesamt
für Sicherheit in der
Informationstechnik