Internationale Kooperationen bei KI zur Bewältigung gemeinsamer Aufgaben

Dr.-Ing. Markus Kühn (BSI), 05.11.2025

Vis!t 2025, Johannes Kepler Universität Linz

Agenda

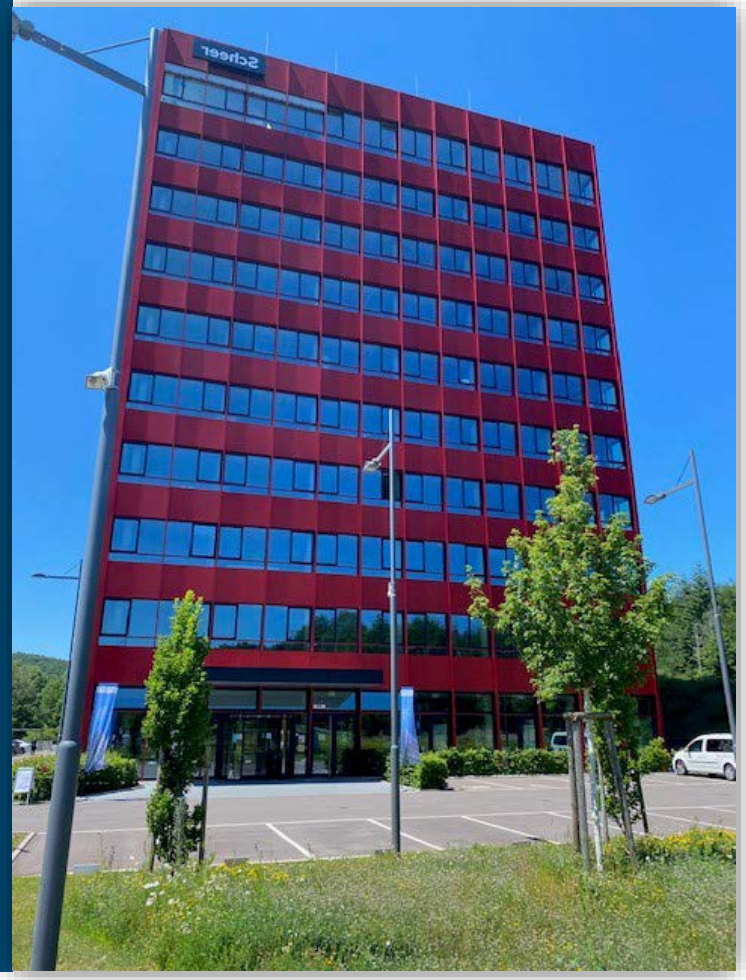
- KI im BSI
- Gemeinsame Publikationen
- Vorträge bei Partnern
- Werbung für Kooperation

BSI Saarbrücken

- seit 2021
- am Campus der Universität des Saarlandes
- aktuell ca. 30 Mitarbeiter/-innen
- 2 KI-Referate:
  - Sicherheit in der Künstlichen Intelligenz
  - Bewertungsverfahren und technische Unterstützung des Digitalen Verbraucherschutzes in der Künstlichen Intelligenz

Ziele:

- Sicherheit der KI in Deutschland und Europa mitgestalten
- Intensiver Austausch mit Universitäten und Forschungszentren
- Aufbau eines Netzwerks auf nationaler und internationaler Ebene
- Kooperation mit internationalen Partnern



Grundsatz und Strategie

Wir sind zuständig für KI-Grundsatzangelegenheiten und unterstützen die Leitung bei der Entwicklung der KI-Strategie im BSI

Normen und Standards für KI

Wir entwickeln und bewerten Prüfkriterien, Prüfmethoden und Prüfwerkzeuge für nachweisbar sichere und vertrauenswürdige KI-Systeme

IT-Sicherheit für KI

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

Bewertungsverfahren

Wir untersuchen Techniken, die KI-Systeme erklärbar und transparent machen

Angriffe durch KI

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

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

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

Internationaler Austausch

Wir verfolgen internationale Kooperationen, insbesondere mit Frankreich, der EU und Nato mit aktiver Beteiligung an Fachgremien zu KI Themen

Grundsatz und Strategie

Wir sind zuständig für KI-Grundsatzangelegenheiten und unterstützen die Leitung bei der Entwicklung der KI-Strategie im 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

Normen und Standards für KI

Wir entwickeln und bewerten Prüfkriterien, Prüfmethoden und Prüfwerkzeuge für nachweisbar sichere und vertrauenswürdige KI-Systeme

Bewertungsverfahren

Wir untersuchen Techniken, die KI-Systeme erklärbar und transparent machen

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

Internationaler Austausch

Wir verfolgen internationale Kooperationen, insbesondere mit Frankreich, der EU und Nato mit aktiver Beteiligung an Fachgremien zu KI Themen

Grundsatz und Strategie

Wir sind zuständig für KI-Grundsatzangelegenheiten und unterstützen die Leitung bei der Entwicklung der KI-Strategie im 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

Normen und Standards für KI

Wir entwickeln und bewerten Prüfkriterien, Prüfmethoden und Prüfwerkzeuge für nachweisbar sichere und vertrauenswürdige KI-Systeme

Bewertungsverfahren

Wir untersuchen Techniken, die KI-Systeme erklärbar und transparent machen

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

Internationaler Austausch

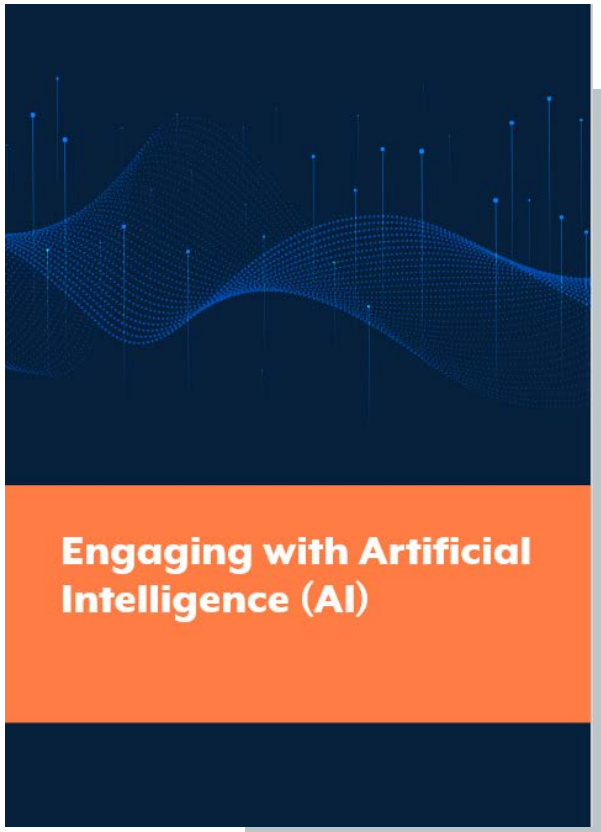
Wir verfolgen internationale Kooperationen mit Partnerländern, der EU und NATO mit aktiver Beteiligung an Fachgremien zu KI Themen

Leitlinien für die Entwicklung sicherer KI-Systeme (2023)

- Leitung: UK NCSC, US CISA
- Fokus: Anbieter / jedes System, das KI nutzt
- Ziel: **Entwicklung sicherer KI-Systeme**
- Lebenszyklusphasen: Risiken & Abhilfen
- Zusammen mit Praktiken für
  - ❖ *Cybersicherheit*
  - ❖ *Risikomanagement*
  - ❖ *Notfallmaßnahmen*

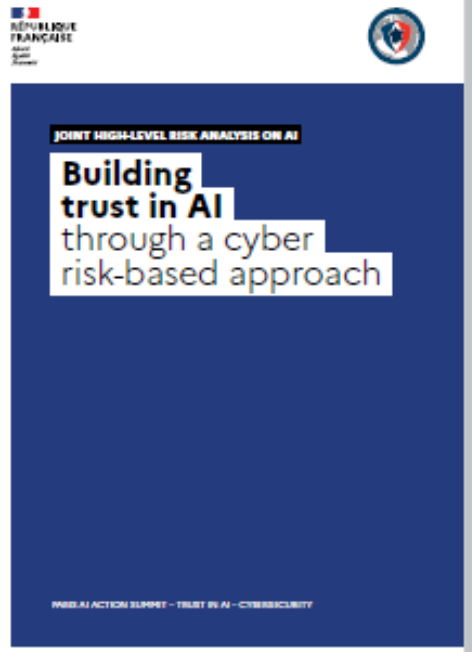
Umgang mit KI (2024)

- Leitung: Australian Cyber Security Centre (ACSC)
- Fokus: Organisationen
- Ziel: **Sichere Nutzung von KI-Systemen**
- Wichtige Bedrohungen & Abhilfeüberlegungen



Aufbau von Vertrauen in KI durch einen auf Cyberrisiken basierenden Ansatz (2025)

- Leitung: ANSSI
- Fokus: Übergeordnete Analyse von Cyberrisiken
- Ziel: Steigerung der Sicherheit von KI-Systemen inkl. Lieferkette
- Risiken & Abhilfen
- basiert auf Guidelines for Sec. AI System Development
- Anhang: Checklisten, Referenzen



Prinzipien für die sichere Integration von KI in industrieller IT (2025)

In Bearbeitung...

KI Programmierassistenten

- Zielgruppe: Manager und Entwickler
- LLM-basierte KI Programmierassistenten für die (teilweise) Automatisierung der Quellcode-Generierung
- auf großen Textmengen trainiert und anschließendes Finetuning mit Quellcode oder direkt mit einer großen Menge an Quellcode trainiert
- oft über ein Plug-in für IDE aufgerufen
- Empfehlungen für eine sichere Nutzung von KI-Codierungsassistenten unter Berücksichtigung der Chancen sowie der Risiken. Konkrete Maßnahmen zur Risikominderung werden dargelegt.

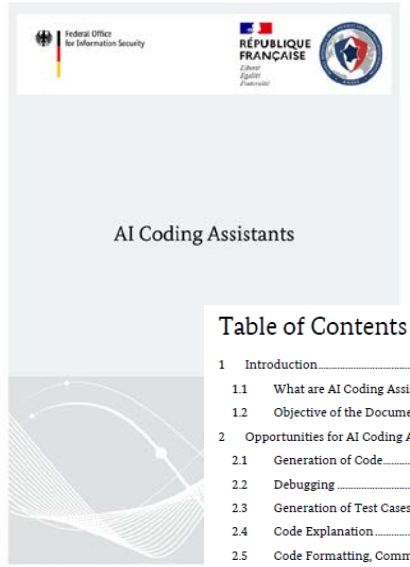


Table of Contents

1 Introduction ... 5
1.1 What are AI Coding Assistants? ... 5
1.2 Objective of the Document ... 5
2 Opportunities for AI Coding Assistants ... 6
2.1 Generation of Code ... 6
2.2 Debugging ... 6
2.3 Generation of Test Cases ... 6
2.4 Code Explanation ... 6
2.5 Code Formatting, Commenting and Documentation ... 7
2.6 Automated Code Translation ... 7
2.7 Increased Productivity and Employee Satisfaction ... 7
3 Risks associated with AI Coding Assistants ... 8
3.1 Missing Confidentiality of Inputs ... 8
3.2 Automation Bias ... 8
3.3 Lack of Output Quality and Security ... 9
3.4 Supply Chain Attacks and Malicious Code ... 9
3.4.1 Hallucinations of Methods and Packages ... 10
3.4.2 Indirect Prompt Injections ... 10
3.4.3 Data and Model Poisoning ... 11
3.4.4 Extensions for Coding Assistants ... 11
3.5 Misuse for Attacks ... 11
4 Conclusion and Recommendations ... 12
4.1 Management ... 12
4.2 Development ... 12
4.3 Research Agenda ... 13
Bibliography ... 14

Designprinzipien für LLM-basierte Systeme mit Zero Trust

- Zielgruppe: Systemarchitekten, Betreiber und Behörden
- Rahmenwerk für sichere LLM-Systeme
  angelehnt an dem Zero-Trust-Prinzip
- 6 Designprinzipien mit zugehörigen Beschreibungen, Risikoszenarien
  & konkreten Gegenmaßnahmen
- Blindes Vertrauen in LLM-Ausgaben ist gefährlich,
  ein vollautonomer Betrieb ohne menschliche Aufsicht nicht
  empfehlenswert

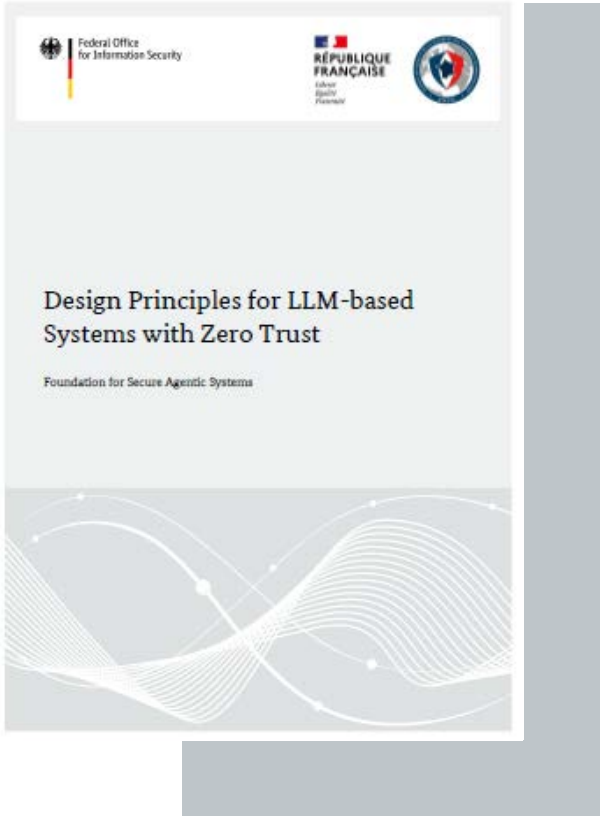


Table of Contents

1 Introduction 5
2 Design Principles for Secure LLM Systems 7
2.1 Authentication and Authorization 8
2.2 Input and Output Restrictions 10
2.3 Sandboxing 11
2.4 Monitoring, Reporting and Controlling 12
2.5 Threat Intelligence 12
2.6 Awareness 13
3 Conclusion 15
Bibliography 16

Gemeinsame G7 Vision zu SBOM für KI

- Erarbeitet mit Cybersicherheitsbehörden/–instituten der G7 Partner im Arbeitsbereich “Smarter Together: Artificial Intelligence” der G7 Cybersecurity Working Group
- Konzept zur Verbesserung der Cybersicherheit durch **Transparenz entlang der KI Lieferkette durch eine SBOM für KI**
- Notwendige **Eigenschaften** und **Minimum Elements** einer SBOM für KI, zusätzlich **Herausforderungen** und **nächste Schritte** innerhalb des G7 Arbeitsbereichs in Richtung einer praktischen Implementierung

„Transparenz von KI-Systemen ist die Bereitstellung von Informationen über den gesamten Lebenszyklus eines KI-Systems sowie über dessen Ökosystem.“ (BSI)

Mitwirkung bei internationalen Veranstaltungen

Aufruf zur Kooperation

- Fachlicher Austausch
- Gemeinsame Publikationen
- Vorträge/Workshops bei Partnern
- Evtl. Hospitationsmöglichkeit?

Vielen Dank für Ihre Aufmerksamkeit!

Kontakt

Dr.-Ing. Markus Kühn
Referent KI

markus.kuehn@bsi.bund.de

Bundesamt für Sicherheit in der Informationstechnik (BSI)
Godesberger Allee 185-189
53175 Bonn
www.bsi.bund.de