ITnet Thüringen e.V., Anger 81, 99084 Erfurt
Vorstand: Heiko Kahl, Tanja Zschau, Jörg Flügge

Handlungsempfehlung zur Verwendung von Allzweck-KI

von der Arbeitsgruppe IT Sicherheit des ITnet Thüringen (Stand: 18. März 2025)

KI-Systeme mit allgemeinem Verwendungszweck, sog. General Purpose AI, können den Arbeitsalltag enorm erleichtern. Die Vorteile erkauft man sich jedoch mit einem gewissen Risiko: Die Rückgaben auf die Eingabeaufforderungen („Prompts“) können Fehler enthalten, wie dieses Beispiel (https://www.golem.de/news/ki-erfindet-fallzitate-wie-der-blamiert-sich-ein-anwalt-wegen-chatgpt-2502-192968.html) eindrucksvoll belegt. Zudem lernen die meisten KI-Systeme durch die von Nutzern bereitgestellten Daten (Eingaben und hochgeladene Dateien), was unweigerlich dazu führt, dass ein gewisses Risiko besteht, sensible Informationen wie personenbezogene Daten und Geschäftsgeheimnisse an den jeweiligen Dienst zu übertragen. Durch Hacking an den Eingabeaufforderungen, sog. Prompt-Injections können KI-Systeme wie ChatGPT dazu gebracht werden, Daten preiszugeben, die eigentlich geheim bleiben sollen. Höchste Zeit also, das richtige Verhältnis zwischen Nutzen und Risiko bzw. Komfort und Sicherheit zu finden!

Gesetzliche Vorgaben

Der EU AI Act legt fest, dass Nutzer von KI ab Februar 2025 eine gewisse Nutzungskompetenz vorweisen müssen. Diese Kompetenz kann sehr gut im Rahmen des jährlichen Datenschutz-Trainings vermittelt werden. Empfehlenswert sind eine Teilnehmerliste und eine Präsentation, die als Nachweis vorgelegt werden können. Im Idealfall lassen Sie die Teilnehmer des Trainings ein Wissensquiz in Form eines (Online-)Fragebogens ausfüllen. Zur Jahresmitte soll wohl ein offizieller Leitfaden zur Verwendung von Allzweck-KI erscheinen. (vgl. https://heise.de/-10264965)

Besser Erlauben statt Verbieten

Wir empfehlen Ihnen, den Einsatz von ChatGPT & Co. lieber unter Auflagen zu gestatten, als diese neue Technologie pauschal zu verbieten. Warum? Nutzer verwenden sonst womöglich ihre privaten Accounts – es entsteht Schatten-IT und Sie verlieren völlig die Kontrolle über den Umgang mit Ihren Geschäftsdaten.

Erwerben Sie Firmen-Lizenzen für die gewünschten Produkte und konfigurieren Sie diese gewissenhaft (z.B. Opt-out zur Verwendung Ihrer Eingaben für das Training der KI). Danach erstellen Sie eine Nutzungsrichtlinie, in der alle Dienste aufgeführt sind, die verwendet werden dürfen. Außerdem sollten Hinweise für die Nutzer enthalten sein, wie die freigegebenen Dienste zu verwenden sind. Zum Beispiel:

- Es dürfen keine personenbezogenen Daten eingegeben werden
- Es dürfen keine Firmengeheimnisse eingegeben werden
- Es dürfen keine Kommunikationsverläufe an den Dienst übermittelt werden (z.B. der Mailverlauf)

- Es dürfen keine Metadaten aus der vorherigen Kommunikation übermittelt werden (z.B. Mail-Header)
- Ausgaben des Dienstes müssen auf Richtigkeit und Plausibilität geprüft werden! KI-generierte Inhalte können grobe Fehler enthalten.
- Werden sich KI-generierte Inhalte zu eigen gemacht, haftet im Zweifel der Nutzer für den Inhalt. Möchte man sich aus der Haftung nehmen, muss der Inhalt mit dem Hinweis versehen werden, dass es sich um einen KI-generierten Inhalt handelt.
- Bei integrierter KI in Standard-Applikationen (z.B. Adobe Acrobat) oder Betriebssystemen ist der Einsatz risikobasierend zu bewerten und gegebenenfalls zu blockieren.
- KI ist derzeit nicht zu Innovation und Kreativität fähig. Sind diese Skills in einer Arbeitsaufgabe gefordert, ist ein KI-Produkt zum Stand dieser Richtlinie nicht geeignet.

Weitere Tipps:

- Bleiben Sie am Thema dran und verfolgen Sie die Berichterstattung, z.B. über den Heise Online Newsticker
- Diese Handlungsempfehlung reicht nicht mehr aus, wenn Ihr KI-System über den Status einer Allzweck-KI hinausgeht; z.B. indem Gesundheitsdaten interpretiert werden – damit steigt die Risikoklasse
- Wenn Sie sich unsicher sind, konsultieren Sie einen Fachanwalt für IT-Recht
- Bei besonders schützenwerten Daten sollte über den Einsatz einer Lokalen KI nachgedacht werden bzw. sollten nur KI-Modelle eingesetzt werden, die eine Einstellung auf „ausschließlich lokales lernen“ ermöglichen.
- Die OWASP Top 10 stellt die 10 größten Risiken von Generativer KI und Large Language Models in diesem Paper vor. Auf der Website stehen zusätzliche Ressourcen zum Thema bereit. (https://www.owasptopten.org)