Bundesamt
für Sicherheit in der
Informationstechnik

TOP 4: Herausforderungen der Künstlichen Intelligenz im Kontext der Cybersicherheitsentwicklung

Tischvorlage

KI in der IT-Sicherheit: Verschiedene Rollen und Dimensionen von KI

- IT-Sicherheit für KI: Neuartige Bedrohungen für KI-Systeme erfordern die Entwicklung geeigneter Gegenmaßnahmen
- IT-Sicherheit durch KI: KI-Methoden zur Prävention, Detektion und Reaktion bei Cyber-Angriffen können zur Verbesserung der IT-Sicherheit genutzt werden
- Angriffe durch KI: Neue KI-gesteuerte und KI-unterstützte Angriffsmethoden, z.B. im Bereich Biometrie und Deepfakes, erfordern geeignete Gegenmaßnahmen, um IT-Systeme und Infrastrukturen zu schützen

Herausforderung und Gefahrenpotenzial: Sicherheit, Transparenz, Erklärbarkeit, Verlässlichkeit, Robustheit, Nicht-Diskriminierung von KI-Systemen

- Aktuell weitverbreitete KI-Systeme sind mitunter sehr komplex. Im Gegensatz zu klassischen Programmen wird die auszuführende Funktionalität z.T. nicht von Menschen explizit vorgegeben, sondern anhand von Daten implizit „gelernt“
- Das „Erlernte“ ist für Menschen bei vielen KI-Systemen weder lesbar noch verständlich, sodass bei diesen Systemen nicht erklärbar ist, wie sie konkret zu einem erstellten Ergebnis gelangen, ob dieses inhaltlich korrekt ist und ob dieses trotz gleicher Eingabe nicht variieren kann.
- Herstellung von Sicherheit, Transparenz, Erklärbarkeit, Verlässlichkeit, Robustheit und Nicht-Diskriminierung stellen beim Einsatz von KI eine Herausforderung dar

Herausforderung: KI ist ein hochdynamisches und personalintensives Entwicklungsfeld

- KI ist in Forschung, Entwicklung und Anwendung ein hochdynamisches Thema. Es ergeben sich täglich neue Erkenntnisse, die es jeweils sicherheitstechnisch zu bewerten und zu adressieren gilt
- Stetig neue Angriffsvektoren auf KI-Systeme erfordern eine permanente Weiterentwicklung von Gegenmaßnahmen, fortlaufende Kontrolle und Nachbesserung sowie regelmäßig aktualisierte Schulungen von Entwicklern und Anwendern.
- Etablierung und Aufrechterhaltung einer fachlich qualifizierten und größenmäßig adäquat aufgestellten Personaldecke, insbesondere für die öffentlichen Stakeholder, stellt eine große Herausforderung dar.

Nutzungsmöglichkeiten von KI als Werkzeug im Rahmen klassischer IT-Sicherheit

- KI wird etablierte Arbeitsabläufe und die Nutzung von Informationstechnik verändern und dadurch auch Einfluss auf die allgemeine Cybersicherheitsentwicklung ausüben.
- Generative KI kann auf Grundlage von großen gelernten Datenmengen neue Inhalte in Form von Text, Bildern, Audio oder Video in immer besserer Qualität erzeugen, verarbeiten und anpassen.
- Die Authentizität von medialen Inhalten und medialen Identitäten kann immer weniger nur über den Inhalt und dessen Qualität allein beurteilt werden. Phishing und Social Engineering sowie das Erstellen und Verbreiten von Falschinformationen wird durch generative KI erleichtert.
- Insbesondere große KI-Sprachmodelle können für Verteidiger und Angreifer gleichermaßen als Werkzeug dienen, um Einstiegshürden zu verringern, Effektivität und Qualität von Arbeitsergebnissen zu steigern, Programmcode zu erzeugen, Datenverkehr zu analysieren und Wirkungen zu skalieren.

Sicherheitsherausforderungen in Bezug auf Nutzung von großen KI-Sprachmodellen

- Große KI-Sprachmodelle sind in der Lage, textuelle Eingaben zu verarbeiten und darauf basierend Textausgaben in Form verschiedener Textformate, wie z.B., Tabellen, Programmcode oder natürlichsprachlichem Text in hoher menschenvergleichbarer Qualität zu erzeugen.
- Eine Vielzahl von textbasierten Aufgaben in Behörden, wie z.B. Verarbeitung, Klassifikation, Überprüfung und Erzeugung von Textdokumenten sowie das Auffinden, Extrahieren und Zusammenfassen von Informationen, können so teil- oder vollautomatisiert durch KI übernommen werden
- Große KI-Sprachmodelle weisen technologiebedingt spezielle intrinsische Schwachstellen auf, denen mit verschiedenen Sicherheitsmaßnahmen und der Durchführung einer systematischen Risikoanalyse begegnet werden muss. Die BSI-Publikation „Große KI-Sprachmodelle - Chancen und Risiken für Industrie und Behörden“ kann hierfür als Grundlage dienen.

Begegnung von Sicherheitsherausforderungen mittels KI-Prüfung, KI-Standardisierung und digitalem Verbraucherschutz

- KI ist eine Schlüsseltechnologie der Digitalisierung. Das BSI hat den Anspruch die Digitalisierung in all ihren Facetten federführend sicher zu gestalten und somit auch eine zentrale Stelle zu Fragen der Sicherheit und Prüfung von KI-Systemen im Bund zu werden.
- Das BSI entwickelt aktuell mit Partnern aus Forschung, Entwicklung, Wirtschaft und Verwaltung technologische Grundlagen und Kriterien zur Bewertung und Prüfung von KI-Systemen, um diese anschließend in Form von Prüfkatalogen und -empfehlungen in die Praxis zu überführen.
- Auf Basis dieser Grundlage wirkt das BSI bei der Entwicklung von KI-Normen und KI-Standards, auch in Hinblick auf den EU AI Act, aktiv mit und bringt seine langjährige Erfahrung und fachliche Expertise in nationalen und internationalen Standardisierungsgremien ein.
- Fähigkeiten, Limitierungen und Sicherheitseigenschaften von KI-Systemen spielen für deren vertrauensvollen und zweckmäßigen Einsatz eine wichtige Rolle. Hierbei kann auch durch die Schaffung von Transparenz und Einhaltung von Mindestanforderungen, z.B. im Sinne eines freiwilligen Sicherheitslabels für KI-Systeme, ein Mehrwert für Anwender und Verbraucher geschaffen werden.

Unterstützungsmöglichkeiten des BSI für Landesverwaltungen

- Die Unterstützungsmöglichkeiten des BSI hängen im Einzelfall von derzeitigen rechtlichen Rahmenbedingungen im Bund-Länder-Verhältnis, bestehenden BSI-Kooperationsvereinbarungen mit den Ländern sowie BSI-seitiger Ressourcenverfügbarkeit ab. Nach Prüfung der jeweiligen Rahmenbedingungen in einem Bundesland wären im Einzelfall eine allgemeine Beratung zum sicheren Einsatz von KI-Systemen sowie Bewertung in einem konkreten Anwendungsfall möglich.
- Angesichts der beschriebenen Einsatzmöglichkeiten und Potentiale von KI-Verfahren auch für den Bereich der Landesverwaltung erscheint jedoch eine einzelfallbezogene Unterstützungsleistung als nicht hinreichend, um den benannten Herausforderungen zu begegnen. Hierzu wird aus fachlicher Sicht eine dauerhafte Unterstützung durch das BSI in einer Schlüsselrolle als Stelle der zentralen Koordinierung und Bereitstellung von Informationen und Know-How als erforderlich erachtet.
- Um eine dauerhafte Unterstützungsleistung zu ermöglichen, bedarf es jedoch der Schaffung eines entsprechenden rechtlichen Rahmens.