Lernende Systeme
DIE PLATTFORM FÜR KÜNSTLICHE INTELLIGENZ

Künstliche Intelligenz und IT-Sicherheit

Whitepaper von Jörn Müller-Quade et al.
Arbeitsgruppe IT-Sicherheit, Privacy, Recht und Ethik

Kurzfassung

Technologien, die auf Künstlicher Intelligenz (KI) basieren, durchdringen zunehmend alle Lebensbereiche. Ihr Beitrag zu einer verbesserten Sicherheit von IT-Systemen wie auch die Sicherheit von KI-Systemen selbst sind essenziell, damit Bürgerinnen und Bürger, Unternehmen, Politik und Behörden die Vorteile der fortschreitenden Digitalisierung nutzen können.

KI-Systeme spielen zukünftig eine wichtige Rolle zur Erhöhung der IT-Sicherheit. Methoden des maschinellen Lernens können beispielsweise eingesetzt werden, um die Fähigkeit von Angriffserkennungssystemen zu verbessern oder um in Netzwerken normale von verdächtigen Aktivitäten zu unterscheiden. KI-Systeme können Fachkräfte für IT-Sicherheit wirksam unterstützen und damit kurzfristig die Auswirkungen des Fachkräftemangels in der IT-Sicherheit abmildern.

KI-Technologien bergen jedoch auch im Bereich der IT-Sicherheit ein Dual-Use-Potenzial. Methoden des maschinellen Lernens, mit denen sich bisher unbekannte Sicherheitslücken in Netzwerken oder Softwaresystemen identifizieren lassen, können auch von Angreifern verwendet werden. Diese können mithilfe von Methoden und Verfahren der KI ihre Angriffsstrategien optimieren oder neue Bedrohungen entwickeln. Das Bedrohungsrisiko sollte zwar nicht überzeichnet werden, ist jedoch eine zusätzliche Motivation dafür, sich auf diesem Einsatzfeld der KI einen technologischen Vorsprung zu erarbeiten und als Entwickler und Anwender ein Bewusstsein für das Dual-Use-Potenzial zu entwickeln.

KI-Systeme werden zunehmend in Prozesse integriert, bei denen Sicherheit und Datenschutz eine zentrale Rolle spielen. Deshalb gilt es, die KI-Systeme selbst vor Angriffen zu schützen, ihre Robustheit gegenüber möglichen Manipulationen zu erhöhen und entsprechende Schutzmaßnahmen zu implementieren.

Angesichts der neuen Dynamik, die KI-Systeme in den Bereich der IT-Sicherheit bringen, ergeben sich unterschiedliche Handlungsfelder – von der Unterstützung kleinerer und mittlerer Unternehmen über den Kompetenzaufbau zu

KI und IT-Sicherheit bis hin zu Entwicklung und Design der Systeme selbst. Die Autorinnen und Autoren dieses Papiers bringen für diese Handlungsfelder erste Lösungsansätze ein, die sie in einem nächsten Schritt ausdifferenzieren und weiterentwickeln werden.

Die nachfolgend skizzierten Lösungsansätze geben eine erste Einschätzung der Autoren dieses Papiers wieder und werden innerhalb der Plattform Lernende Systeme weiterentwickelt.

Allgemeine Handlungsfelder

- Der zukünftige Einsatz von Lernenden Systemen in sicherheitskritischen Anwendungen verlangt möglicherweise besondere Sorgfalt und die **Integration von spezifischen Schutz- und Abwehrmaßnahmen**. Dabei sollten grundsätzliche Prinzipien der Sicherheit beachtet und neuartige Verteidigungskonzepte vorangetrieben werden, die speziell Lernalgorithmen schützen und langfristig einen sicheren Betrieb von KI ermöglichen. Die Methodik der beweisbaren Sicherheit könnte dabei auch im Zusammenhang mit KI-Systemen angewendet werden.

- KI-Systeme sollten mit einer **technischen Rückfallebene ausgestattet** werden – für den Fall von Fehlfunktionen, Angriffen auf das System oder falls das System selbst sicherheitskritisches Verhalten zeigt. Der sichere Betrieb des Gesamtsystems darf dadurch nicht gefährdet werden.

- Mit Blick auf das Dual-Use-Potenzial empfiehlt es sich, unterschiedliche Möglichkeiten zu prüfen, wie die **Zweckentfremdung von KI-Systemen bestmöglich verhindert** werden kann. Dafür könnten zunächst gesellschaftliche und rechtliche Maßstäbe entwickelt werden.

- **Im Bereich von Spezialanwendungen** wie der Seitenkanalanalyse erscheint es eher unwahrscheinlich, dass in absehbarer Zeit ausgereifte Produkte zur Verfügung stehen. Vielmehr sollten Nutzer solcher Anwendungen (z. B. Wirtschaft, Universitäten und Behörden) bei Bedarf **eigene Expertise aufbauen**.

- Die heutige Fokussierung auf Betreiber kritischer Infrastrukturen sollte mit Blick auf vernetzte Geräte mit KI-Komponenten, die bei Bürgerinnen und Bürgern oder Behörden im Einsatz sind, ergänzt werden. Dabei sollten auch in Forschung und Entwicklung von Systemen mit KI-Komponenten die Prinzipien **Security by Design** und **Security by Default** beachtet werden.

- Angesichts der notwendigen Vertrauenswürdigkeit von KI-Systemen, u. a. auch in sicherheitskritischen Kontexten, ist eine forschungs- und industriepolitische Pointierung des Anspruchs der **Digitalen Souveränität**[1] notwendig.

Handlungsfelder für Politik und Behörden

- Um die Fachkräftelücke im Bereich der IT-Sicherheit zu schließen, sind die Anstrengungen im Bereich der **Ausbildung und Gewinnung von Fachkräften in der IT-Sicherheit** zu verstärken. Dabei könnten der Umgang mit KI-Systemen für die IT-Sicherheit in der Aus- und Weiterbildung der

1 Die Digitale Souveränität umfasst „die vollständige Kontrolle über gespeicherte und verarbeitete Daten sowie die unabhängige Entscheidung darüber, wer darauf zugreifen darf. Sie umfasst weiterhin die Fähigkeit, technologische Komponenten und Systeme eigenständig zu entwickeln, zu verändern, zu kontrollieren und durch andere Komponenten zu ergänzen. Digitale Souveränität ist deswegen einerseits wichtige Grundlage für vertrauenswürdige Systeme und andererseits unverzichtbare Voraussetzung für unabhängiges staatliches Handeln." (Plattform Innovative Digitalisierung der Wirtschaft 2018: 3.)

Fachkräfte berücksichtigt und entsprechend Lehrpläne auf dem Stand der Technik fortlaufend aktualisiert werden.

- Ein **Grundverständnis für IT-Sicherheit** sollte in angemessener Form auch in Disziplinen integriert werden, in denen das Thema im Zuge der fortschreitenden Digitalisierung zunehmend an Bedeutung gewinnt, etwa im Maschinenbau.

- Für KMU empfiehlt es sich, beispielsweise im Rahmen der bestehenden Kompetenzzentren Angebote zu schaffen oder auszubauen, mit deren Hilfe sie ihre **Kompetenzen im Bereich der IT-Sicherheit im Hinblick auf den Einsatz von KI-Systemen erweitern** können. Dabei sollten auch entsprechende Beratungsangebote berücksichtigt werden. Ein Navigationssystem zu IT-Sicherheit im KI-Kontext könnte KMU unterstützen und dazu beitragen, bestehende Angebote übersichtlicher und zugänglicher zu gestalten.

- Bei der **Integration von KI-Systemen in die öffentliche Verwaltung** und in Dienste für Bürgerinnen und Bürger spielt die Sicherheit der Systeme eine wichtige Rolle. Setzen Behörden beispielsweise Chatbots für ihre Services ein, sollten Maßnahmen erarbeitet und ergriffen werden, die verhindern, dass Angreifer durch Überlisten der für diese Dienste eingesetzten KI-Komponenten Zugriff auf personenbezogene Daten erhalten.

- Der Einsatz von KI-Systemen verdeutlicht den Bedarf einer **kohärenten, möglichst globalen IT-Sicherheitspolitik**. Internationale Initiativen sollten im Ergebnis auch auf globaler Ebene verantwortliches staatliches Handeln in diesem Bereich fördern, etwa ein entschlossenes Vorgehen gegen Hackerangriffe vom eigenen Territorium.

Handlungsfelder für Unternehmen

- Konzerne sollten Werkzeuge wie **Angriffserkennungssysteme mit KI-Funktionalität** am Markt kaufen und selbst betreiben, während kleinere Firmen Dienstleistungen beziehen könnten. Wichtig ist, dass entsprechende Angebote für KMU gemacht werden.

- Der Aufbau technischer Fähigkeiten und Kompetenzen für den Umgang mit KI im Bereich der IT-Sicherheit ist möglicherweise erfolgskritisch. Marktfähige **KI-unterstützte Sicherheitslösungen und deren ständige Weiterentwicklung** sind wichtige Voraussetzungen für die IT-Sicherheit der deutschen Industrie.

- Unternehmen sollten ihre **Maßnahmen und Kompetenzen in der IT-Sicherheit** mit Blick auf den zukünftigen Einsatz von KI in diesem Bereich überprüfen und gegebenenfalls Anstrengungen unternehmen, entsprechende Kompetenzen aufzubauen.

- In Verbindung mit KI können künftige Angriffe auf die IT-Systeme in Büros und in der Produktion gezielter und deutlich intelligenter ausgeübt werden. Unternehmen könnten deshalb neben den klassischen Schwachstellen- und Bedrohungsanalysen **KI-unterstützte und lernende Überwachungssysteme** implementieren.

- Wie auch in der IT-Sicherheit schreitet der Stand der Technik auch bei KI-Systemen voran. Erforderlich ist daher eine revolvierende **Überprüfung der bereits eingesetzten intelligenten Abwehrmaßnahmen** inklusive der verwendeten KI.

Handlungsfelder im Bereich Forschung

- Es besteht Forschungsbedarf, wie sich KI-Systeme für die Unterstützung und Verbesserung der IT-Sicherheit nutzen lassen, aber auch mit Blick auf die Sicherheit und den Schutz der KI-Systeme selbst. Dazu sollten entsprechende Forschungsaktivitäten in diesem Bereich angestoßen oder ausgebaut werden. Sie könnten unterstützt werden durch eine **stärkere Vernetzung von Einrichtungen mit Forschungsschwerpunkten** im KI-Bereich und/oder im Bereich der IT-Sicherheit in Verbindung mit einem weiteren Kompetenzaufbau in Deutschland.

- Die Möglichkeiten, die der Einsatz von KI-Systemen für Bürgerinnen und Bürger, Unternehmen oder die öffentliche Verwaltung in unterschiedlichen Anwendungsbereichen bietet, lassen sich nur nutzen, wenn die Systeme möglichst gut gegen Manipulationen geschützt sind, insbesondere gegen Angriffe auf die Vorhersage oder Angriffe auf den Lernprozess. Hier lässt eine Intensivierung der Forschung zur **Resilienz von KI-Systemen gegen Manipulationen** entscheidende Fortschritte erwarten.

- Die Sicherheit von KI-Systemen lässt sich insbesondere durch **Tests gegen Sonderfälle** erhöhen. Die Forschung zu Techniken, die beispielsweise gezielt ungewöhnliche Eingaben automatisiert generieren und damit potenzielle Angriffe simulieren, könnte hier einen wichtigen Beitrag leisten.

- Auch beim Einsatz von KI-Systemen muss der Schutz personenbezogener Daten sichergestellt werden, insbesondere in Anwendungsfeldern, in denen sensible Daten für den Lernprozess der Systeme verwendet werden, wie beispielsweise im medizinischen Bereich. Die fortlaufende Erforschung und **Weiterentwicklung datenschutzerhaltender Lernalgorithmen**, die eine Extraktion und Rekonstruktion personenbezogener Daten aus den Modellen der Lernenden Systeme verhindern oder erschweren, kann dazu beitragen, den Einsatz von KI-Systemen in unterschiedlichen Anwendungsbereichen unter Wahrung des Datenschutzes zu befördern.

- In der IT-Sicherheit, wie in anderen Anwendungsfeldern von KI-Systemen, kann die **Erklärbarkeit der Entscheidungen der Systeme** zu einem wichtigen Faktor für deren Anwendbarkeit werden. Das gilt etwa in Anwendungsbereichen, in denen der Mensch, der mit den Systemen interagiert, die Einflussfaktoren für deren Entscheidungen nachvollziehen und beurteilen können muss. Dies betrifft insbesondere komplexe neuronale Netze. Die Forschung zu Möglichkeiten der Erklärbarkeit sollte vorangetrieben werden, um einen sicheren und transparenten Einsatz der Systeme zu ermöglichen.

Impressum
Herausgeber: Lernende Systeme – Die Plattform für Künstliche Intelligenz | Geschäftsstelle | c/o acatech | Karolinenplatz 4 | D-80333 München | kontakt@plattform-lernende-systeme.de | www.plattform-lernende-systeme.de | Folgen Sie uns auf Twitter: @LernendeSysteme | Stand: April 2019 | Bildnachweis: matejmo / iStock

Diese Kurzfassung entstand auf Grundlage des Whitepapers *Künstliche Intelligenz und IT-Sicherheit – Bestandsaufnahme und Lösungsansätze,* München, 2019. Die Autorin und Autoren sind Mitglieder der Arbeitsgruppe IT-Sicherheit, Privacy, Recht und Ethik der Plattform Lernende Systeme. Die Originalfassung der Publikation ist online verfügbar unter: https://www.plattform-lernende-systeme.de/publikationen.html

GEFÖRDERT VOM

Bundesministerium für Bildung und Forschung

acatech
DEUTSCHE AKADEMIE DER TECHNIKWISSENSCHAFTEN