Künstliche Intelligenz (KI) zur Abwehr von Cyber-Angriffen und Cyber-Angriffe auf KI

Unter Künstlicher Intelligenz stellt man sich gerne Computer oder gar Roboter vor, die in allen möglichen Situationen wie Menschen lernen, planen und kreative Entscheidungen treffen können. Dabei können sie sogar eigene Interessen verfolgen. Diese Art „starker künstlicher Intelligenz" ist aber bis heute Science Fiction, wie etwa jener legendäre Computer HAL aus Stanley Kubricks Kultfilm „2001: Odyssee im Weltraum" von 1968.

„Schwache Künstliche Intelligenz" dagegen beruht auf Computerprogrammen, die mithilfe von „Maschinellem Lernen" Daten analysieren und daraus Antworten auf konkrete Anwendungsfragen ableiten. „Maschinelles Lernen" ist Gegenstand aktueller Forschung. Es gibt bereits viele erfolgreiche Anwendungen, darunter Mustererkennung in Texten und Bildern, Vorhersagen zur Kreditwürdigkeit von Kunden und Prognosen zur Kündigungsbereitschaft von Verträgen. Auch *Data Mining* zur Kategorisierung von Daten und zur Herstellung von Beziehungen unter Daten gehört zur realisierbaren Künstlichen Intelligenz. Ein Anwendungsbeispiel dafür ist die Analyse von Warenkörben und Kundentypen, die zu individuellen Kaufvorschlägen führt. Andere Beispiele sind Aktienmarktanalyse oder die Erkennung von Kreditkartenbetrug. Weiterhin gibt es Software, die für eng begrenzte Aufgaben aus kontinuierlich eingehenden Sensordaten dynamisch Entscheidungen für weiteres Verhalten ableiten kann wie zum Beispiel zur Fahrzeug- oder Robotersteuerung.

Solche Lernverfahren werden auch zur Abwehr von Cyberangriffen genutzt, etwa die automatische Erkennung von Kommunikationsanomalien im Netz, das Herausfiltern von Bildern mit Kinderpornographie aus massenweisem Bildmaterial und die Unterscheidung von krimineller von harmloser Kommunikation. Umgekehrt kann Künstliche Intelligenz leider auch dazu genutzt werden, um immer raffiniertere Cyberangriffe zu steuern, etwa zum dynamischen Fälschen von Text-, Bild- und Videomaterial, zum Einstreuen von Propaganda in politische Internetforen oder zur selbst-adaptierenden Suche von Schwachstellen im Netz.

Gegenstand dieses Schwerpunktheftes ist das Potenzial von Künstlicher Intelligenz, einerseits Cyber-Angriffe zu unterstützen, andererseits aber auch, Angriffe abzuwehren. Dazu werden hier exemplarisch drei Einsatzbereiche vorgestellt. M. Sc. Raphael Antonius Frick, Prof. Dr. Martin Steinebach und Dr. Sascha Zmudzinski vom Fraunhofer-Institut SIT zeigen in ihrem Beitrag zu „Deepfakes, Dall-E & Co" auf, wie KI-gestützte Manipulationsmethoden von Bildern, Videos und Tonspuren funktionieren, wie man das erkennen und was man dagegen tun kann. Dr. Oren Halvani von der Zentralen Stelle für Informationstechnik im Sicherheitsbereich (ZITiS) beschreibt Möglichkeiten zur Erkennung von Hassreden und die damit verbundenen Herausforderungen. Privatdozent Dr.-Ing. Christian Riess und M. Sc. Anatol Maier vom Fachbereich Informatik der Universität Erlangen zeigen die Grenzen der Zuverlässigkeit von Deep-Learning-Methoden in der Multimedia-Forensik auf und machen Vorschläge zu ihrer Verbesserung.

In einem weiteren Beitrag dieses Schwerpunktheftes geht der Jurist Marcel Kohpeiß vom Wissenschaftlichen Zentrum für Informationstechnikgestaltung der Universität Kassel der Frage nach, ob das IT-Sicherheitsgesetz 2.0 Organisationen der Kritischen Infrastruktur dazu verpflichtet, ab dem 1. Mai 2023 nicht nur – wie bisher – einfach Systeme zur Angriffserkennung zu implementieren, sondern sogar solche Erkennungssysteme einzusetzen, die KI-gestützt sind.

Rüdiger Grimm[1]

1 Prof. i.R. Dr. Rüdiger Grimm (war Professor für IT-Risk-Managementik in der Universität Koblenz und) ist wissenschaftlicher Berater und Ombudsmann für gute wissenschaftliche Praxis im Fraunhofer-Institut Sichere Informationstechik SIT in Darmstadt. E-Mail: grimm@uni-koblenz.de