Wissenschaft als Helfer für angewandte Ethik in der KI-unterstützten IT

IM SINNE DER GESELLSCHAFT

Im Kontext der Digitalisierung, und insbesondere vor dem Hintergrund der dadurch prädizierten Effizienzsteigerung, wird der Künstlichen Intelligenz (KI) eine hohe Bedeutung beigemessen. Ebenso in Bezug auf die IT-Sicherheit gehen Experten davon aus, dass KI entscheidend dazu beitragen wird, die strategische Abwehr von Cyberangriffen zu optimieren. Doch trotz aller Euphorie sollten hier neben den Chancen auch die daraus potenziell resultierenden Risiken in Betracht gezogen werden. Denn der Einsatz von KI ist mit Implikationen verbunden, die auf die gesamte Gesellschaft wirken. Von daher ist es nicht nur ratsam, sondern sogar erforderlich, Anwendungen – auch im Bereich IT-Sicherheit – unter ethischen Aspekten zu analysieren und bewerten. Ein am Institut für Internet-Sicherheit if(is) entwickeltes Tool kann dabei wertvolle Unterstützung bieten.

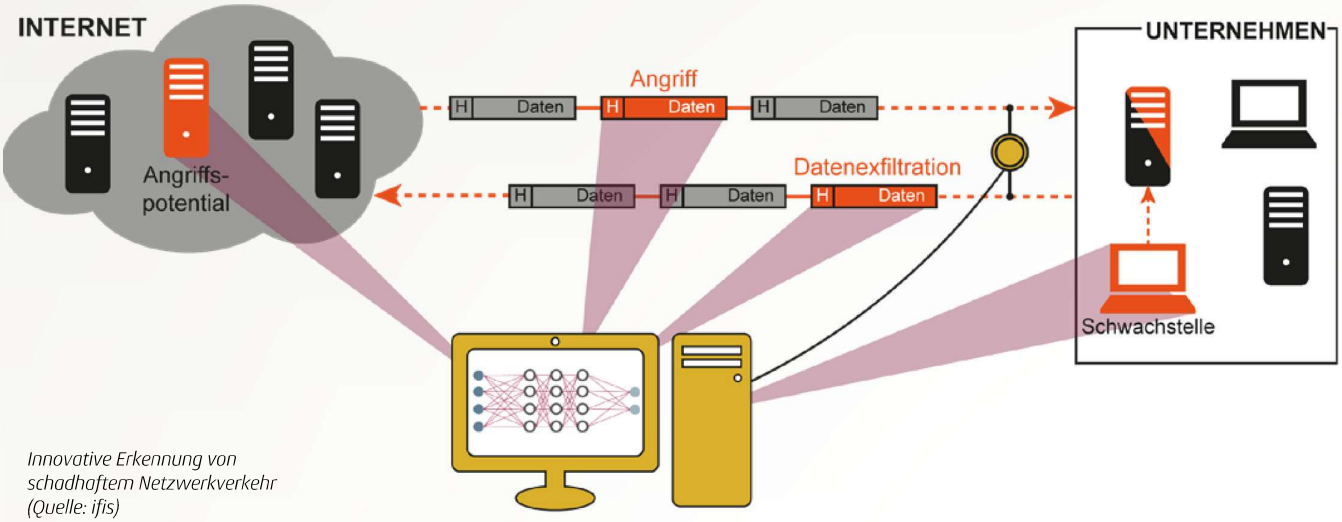
Bild: ©Xtock - stock.adobe.com

Auch wenn die Vorteile beim Einsatz von KI unbestritten sind, so gibt es doch substanzielle Gründe für eine dedizierte Urteilsfindung bezüglich der anwendungsbezogenen Nutzung. Ein wesentlicher ist, dass dem Gros der Menschen im Regelfall die Kenntnisse zur Beurteilung von Technologien fehlt und sie somit erwarten, dass die Unternehmen – konkret die Mitgestalter der digitalen Evolution – ihrer Verantwortung für die Folgenabschätzung sowie den ethisch vertretbaren Einsatz von KI gerecht werden. Diese Verantwortlichkeit resultiert im Wesentlichen auch daraus, dass im Speziellen der Einsatz von KI nicht nur einzelnen Personengruppen Nachteile bringen, sondern ganzen Gesellschaften massiv Schaden zufügen kann.

Denn während einer Person hierdurch eventuell ein Verlust bezüglich der digitalen Sicherheit, Privatsphäre, Reputation oder auch die Verletzung des Grundsatzes der Gleichbehandlung droht, potenzieren sich die möglichen Folgen für die Gesellschaft um ein Vielfaches. Denkbar wäre etwa die Schwächung der Wirtschaftskraft eines Landes, denn bei Unternehmen steigt aufgrund der zunehmenden Komplexität sowie dem Einsatz von KI die Verwundbarkeit und somit das Risiko zur Zielscheibe von Angriffen zu werden. Des Weiteren sind, aufgrund dieser Faktoren, auch negative Effekte in Bezug auf die nationale Sicherheit oder die politische Stabilität als mögliche Konsequenzen vorstellbar.

ANWENDUNG VON KI ZUR STEIGERUNG DER IT-SICHERHEIT

Um auf dieses Gefahrenpotenzial künftig besser vorbereitet zu sein und diesem auch mit geeigneten Maßnahmen etwas entgegensetzen zu können, ist der Einsatz von KI definitiv zu prü-



Innovative Erkennung von schadhaftem Netzwerkverkehr (Quelle: ifis)

fen. Hier gibt es bereits diverse Anwendungsszenarien, welche die Vielfalt in diesem Bereich demonstrieren. So zählt beispielsweise das Detektieren von Angriffen über ein Netzwerk oder über Endgeräte zu den wichtigsten Punkten in Hinblick auf die Reduzierung von Risiken. Das automatisierte Erkennen von sicherheitsrelevanten Ereignissen – etwa zur Priorisierung – ist extrem hilfreich, um Cybersicherheitsexperten von zeitaufwendigen Analysearbeiten zu entlasten.

Insgesamt trägt die (Teil-)Autonomie im Hinblick auf Reaktionsfähigkeit von Cybersicherheitssystemen maßgeblich zur Erhöhung der Resilienz bei. Zum Beispiel könnten, unmittelbar nach Erkennung eines Angriffs, die Firewall-Regeln sofort so geändert werden, dass Angriffsflächen reduziert werden, während gleichzeitig die relevanten Prozesse im Unternehmen aufrechterhalten bleiben.

Da sich die Malware zunehmend dynamisch verändert, kann deren Bekämpfung zukünftig nur durch den Einsatz von KI optimiert werden. Da ständig neuere Varianten erscheinen, sind Analyse und Aktualisierungen der Signatur-Datenbanken kaum noch effizient zu bewältigen. KI-basierte Detektoren können genutzt werden, um verdächtige Aktivitäten in Echtzeit zu erkennen. Anomalie-Erkennung oder Predictive Malware-Analysis sind Verfahren, die durch den Einsatz von KI deutlich verbessert werden können.[1]

Weitere wichtige Bereiche für den KI-Einsatz in der IT-Sicherheit sind unter anderem: Identifizierung von Spam-Mails oder Fake News, sichere Softwareentwicklung, IT-Forensik und Threat Intelligence.

KI UND SICHERHEIT – WIRKLICH EINE FRAGE DER ETHIK?

Da durch den Einsatz von KI nicht nur neue Schwachstellen entstehen können, sondern auch neue Angriffsmöglichkeiten auf Staat und Unternehmen realisierbar sind, erscheint es somit als eine logische Schlussfolgerung, diese kriminellen Attacken auf gleicher Ebene abzuwehren – also mithilfe von KI den Schutz von IT-Systemen und IT-Infrastrukturen zu optimieren. Obwohl KI im Kontext der IT-Sicherheit allein aufgrund der Analysemöglichkeiten als absolut probates Mittel zur Abwehr scheint, muss auch hier – wie bei jedem Einsatz von KI – die Frage bezüglich der Angemessenheit evaluiert werden. Die Notwendigkeit einer Evaluation sowie etwaige weitere Forschungsfragen, die in diesem Kontext generell zu untersuchen sind, lassen sich exemplarisch anhand von zwei Szenarien illustrieren:

1. Privatheit vs. Allgemeinwohl: Problem Schutz von Daten

Basierend auf der Annahme, dass mittels KI aufgrund der Analysefähigkeiten eine Optimierung der IT-Schutzmaßnahmen realisierbar ist, lässt sich hypothetisch ableiten, je mehr Daten mit sicherheitsrelevanten Informationen zur Verfügung stehen, desto schneller und präziser kann ein Angriff detektiert werden. Dies könnte zu dem Schluss führen, dass für eine bestmögliche Kennung in bestimmten Fällen die Einbeziehung personenbezogener Daten von Mitarbeitern unverzichtbar ist, da ein Großteil der Angriffe mithilfe von Social Engineering und Malware über die Endgeräte der Mitarbeiter indirekt ausgeführt werden. Eine Notwendigkeit für deren Hinzunahme könnte sich zum Beispiel für den Fall ergeben, dass ein relevanter Internetprovider angegriffen wird, da dies eine präzise Ursachenforschung für die optimale Reaktion zur schnellstmöglichen Wiederherstellung dieser kritischen Infrastruktur verlangt. Der Bedarf ist theoretisch nachvollziehbar – andererseits werden dadurch die Individualrechte außer Kraft gesetzt, da die Nutzung von personenbezogenen Daten durch die DS-GVO exakt limitiert ist. Im konkreten Fall würde dies bedeuten, dass durch eine Analyse der Daten parallel das Verhalten der Mitarbeiter ausgewertet und somit deren Privatsphäre verletzt wird.

Daraus resultiert folgendes Dilemma: Die Ethik fordert, dass keine grundlegenden Rechte, wie die Privatheit, zugunsten eines höheren Ziels völlig aufgegeben werden dürfen. Dagegen steht die Prämisse des Strebens nach dem (maximierten) Gesamtnutzen für die Gesellschaft. Aus dieser resultiert zwangsläufig die Fragestellung, wann es angeraten oder sogar unabdingbar ist, die Rechte des Individuums zugunsten des Wohles für die Gemeinschaft aufzuheben. Eine mögliche Annäherung zur Auflösung dieses Dilemmas könnte sein, hierfür Grenzen zu definieren, indem Kriterien dafür festgelegt werden, wann das Individualrecht nachrangig zu behandeln ist. Hierfür bedarf es unter anderem der Forschung bezüglich der Wertvorstellungen, auch um sicherzustellen, dass die Wertmaßstäbe zu Beurteilung mit den höheren Zielen einer Gesellschaft konform sind.[2]

2. Strike Back: Problem der Unvollständigkeit der Daten

Dem Konstrukt des „Strike Back" liegt die Hypothese zugrunde, dass ein Angriff durch einen Gegenangriff beendet werden kann. Unter Einsatz von KI wäre es theoretisch möglich, eine vollkommen neutrale Einschätzung zu ermitteln, was unternommen werden müsste, um den Angreifer zum Aufgeben zu motivieren.

Die ethische Frage in diesem Kontext dreht sich um den Begriff der Neutralität und ob es überhaupt möglich ist, die Vollständigkeit der Daten zu erreichen, um eine ausgewogene und verantwortliche Entscheidung durch KI-Systeme errechnen zu lassen. Davon ausgehend, dass ein Strike Back automatisiert erfolgt, könnte möglicherweise ein Schaden auf einer höheren Ebene angerichtet werden, der moralisch nicht vertretbar wäre – wie etwa ein Gegenschlag, der gleichzeitig auch die Stromversorgung von allen Krankenhäusern in einer Region lahmlegen würde. Fragestellungen wie diese zeigen den eklatanten Forschungsbedarf dahingehend auf, inwieweit es leistbar ist, vertrauenswürdige KI-Systeme zu entwickeln, die den Schutz der Zivilgesellschaft gewährleisten können.[2]

EIN KONKRETES DILEMMA IN BEZUG AUF KI UND ETHIK

Konkrete Anwendungsszenarien belegen, dass das Dilemma bei der Entwicklung von KI-basierten Anwendungen darin besteht, dass während dieses Prozesses der Fokus vorrangig auf der bestmöglichen Funktionalität des Produkts beziehungsweise Services im Hinblick auf deren bestmöglicher Wirkweise liegt. Relevante Kriterien, welche die Rechte der Nutzer beeinträchtigen könnten, haben für Entwickler und Programmierer zumeist keine hohe Priorität. Vielfach erscheint eine KI-basierte Anwendung auch auf den ersten Blick unter diesen Aspekten völlig unproblematisch. Da bestimmte Implikationen erst beim Einsatz offensichtlich werden, sollte eine allgemeine Überprüfung bezüglich der intendierten Nutzung stets obligatorisch sein. Denn nur durch die Analyse im Hinblick auf die ethischen Prinzipien einer KI-basierten Anwendung lassen sich Erkenntnisse ermitteln, ob ein Produkt beziehungsweise Service mit den Werten einer gerechten Gesellschaft wie Autonomie oder Solidarität konform ist. Eine Überprüfung der KI-basierten Anwendung gemäß diesen Grundsätzen ist im Weiteren prinzipiell auch bei dem speziellen Einsatzszenario angezeigt, denn abhängig von diesem können die Auswirkungen jeweils differierend sein.

Chancen der KI und ethische Risiken müssen abgewogen werden (Quelle: ifis)

EIN VIELVERSPRECHENDER UNTERSTÜTZUNGSANSATZ – DAS ETHIK.KI.TOOL

Damit Entwicklungsunternehmen in der Lage sind, ihr eigenes Handeln bezüglich eines werteorientierten und vertrauenswürdigen Umgangs mit KI-Anwendungen zu reflektieren, wurde im Institut für Internet-Sicherheit if(is) im Rahmen eines interdisziplinären Forschungsprojekts das informationstechnische Ethik.KI.Tool konzipiert und realisiert, um Unternehmen bei diesem Prozess zu unterstützen. „Entwicklungsunternehmen" meint hier Unternehmen, die mit ihrer Expertise ausschließlich auf die Entwicklung von Anwendungen, Dienstleistungen oder Produkten fokussiert sind und diese Lösungen ihren Kunden anbieten, oder Lösungen gemäß der Spezifikation von Kunden konzipieren und umsetzen. Für sie präsentiert das Ethik.KI.Tool ein Ergebnis für die relevanten ethischen Fragestellungen – ausreichend profunde Basisinformationen vorausgesetzt.

Ethik.KI.Tool (Quelle: ifis)

Die Nutzung des Ethik.KI.Tool läuft in vier grundlegenden Phasen:

1. Phase: Theoretische und praktische Auseinandersetzung mit der Ethik

In dieser Phase werden die ethischen Werte, die allgemein bei KI-basierten Anwendungen eine Rolle spielen, im spezifischen Kontext untersucht. Diese sind – verkürzt dargestellt:

Fairness und Gerechtigkeit

Im Hinblick auf die KI-basierte Anwendung bedeutet dies, dass alle Nutzer angemessen und transparent behandelt werden.

Gleichheit

Im Kontext der KI besteht eine Gefahr der Diskriminierung, beispielsweise bei individualisierten Preisen für Frauen und Männer, vor allem dann, wenn für den Einzelnen nicht erklärbar oder nachvollziehbar ist, aus welchem Grund ein höherer Preis für ein Produkt/eine Leistung bezahlt werden muss.

Solidarität

Im Kontext der KI laufen beispielsweise Versicherer zunehmend Gefahr, gewisse Krankheiten bestimmten Verhaltensweisen zuzuschreiben und daraufhin Personen gesondert zu bewerten und einzustufen. Dies widerspricht dem Solidaritätsgedanken, der für eine Gesellschaft von hoher Relevanz ist.

Toleranz

Für die Gesellschaft ist Toleranz ein wichtiges Faktum und wird unter anderem diskutiert in Zusammenhang mit kulturellen Unterschieden sowie prinzipiell anderen Wertesystemen. Somit ist die Toleranz auch ein wichtiger Wert im Hinblick auf den Wunsch von Nutzern bezüglich der Kontrolle der eigenen digitalen Identität, der unterschiedlich stark ausgeprägt sein kann.

Freiheit

Insbesondere im Kontext der KI ist das ein relevanter Wert, da dies auch die Autonomie des Einzelnen im Gegensatz zur Kontrolle durch ein System beinhaltet.

Kontextuelle Integrität

Mit der kontextuellen Integrität wird berücksichtigt, dass der Einzelne die Freiheit hat, in verschiedenen Lebensbereichen unterschiedliche

Daten preiszugeben. Denn wenn die Verwendung dieser Daten nicht mehr der ursprünglichen Absicht entspricht, wird die kontextuelle Integrität verletzt – zum Beispiel, wenn sehr vertrauliche Informationen aus dem Freundeskreis in individualisierte Preise einfließen.

2. Phase: Arbeiten mit dem Ethik.KI.Tool

In dieser Phase bearbeiten die verantwortlichen Mitarbeiter mit dem Ethik.KI.Tool nacheinander die folgenden Kategorien:

1.) Hier werden vorab **Grundsatzfragen** geklärt, beispielsweise ob die vorliegende KI-basierte Anwendung personenbezogene Daten nutzt oder ob mit dieser eventuell Werte der Gesellschaft verletzt werden.

2.) Die Kategorie **ethische Reflexion** zum Einsatz von KI-basierten Anwendungen dient zur Selbsteinschätzung gegen welche ethischen Werte potenziell verstoßen wird.

3.) Die Kategorie **Operationalisierung genereller ethischer Prinzipien** behandelt die Frage der Kontrolle und Verantwortlichkeiten bezüglich Datengewinnung und -nutzung.

4.) Die Kategorie **Datengewinnung und -nutzung** für die betrachtete KI-basierte Anwendung überprüft die hier relevanten Kriterien wie Aktualität und Vollständigkeit.

5.) Die Kategorie **Durchsetzung der Informationellen Selbstbestimmung** bei der KI-basierten Anwendung dient der eigenen Auditierung, ob diese optimal umgesetzt wird, zum Beispiel bezüglich der Bereitstellung von Zugriffsmöglichkeiten auf persönliche Informationen.

6.) Die Kategorie **Rechtliche Aspekte** der Daten für die betrachtete KI-basierte Anwendung fragt die Umsetzung der vorgeschriebenen Regeln ab, etwa in Bezug auf die Möglichkeiten zur Löschung.

7.) Die Kategorie **IT und IT-Sicherheit** für die betrachtete KI-basierte Anwendung handelt die wichtigen Fragen nach den IT-Sicherheitszielen, dem Level der IT-Sicherheitsmaßnahmen sowie die IT-Sicherheit der verwendeten KI-Technologie und -Anwendungen ab.

3. Phase: Ergebnisse des Ethik.KI.Tool

Die Ergebnisse des Ethik.KI.Tool werden für jeden Schritt und als Gesamtheit aufgeführt. Dabei findet eine Klassifizierung von „Grün" bis „Rot" mit den entsprechenden Zwischenstufen statt.

GRÜN: Die KI-basierte Anwendung ist aus ethischer Sicht (völlig) unbedenklich. Die wesentlichen Kriterien und Werte der Gesellschaft werden angemessen berücksichtigt.

ROT: Die KI-basierte Anwendung ist aus ethischer Sicht abzulehnen.

4. Phase: Aktionen zur Optimierung der Ergebnisse

Insbesondere im Hinblick auf die Zwischenstufen ist es für Unternehmen sinnvoll, gemeinsam mit Beratern die Ergebnisse im Rahmen eines Workshops aufzuarbeiten und die weitere Vorgehensweise für die Erreichung einer wertekonformen KI-basierten Anwendung zu entwerfen.

FAZIT

Der Einsatz von KI macht bestimmte Analysen, die als Basis einer komplexen Urteilsfindung dienen, und viele andere Aktivitäten erst jetzt möglich oder führt dazu, dass diese verbessert werden können.

Doch aufgrund der Tatsache, dass die Einführung neuer Technologien stets mit Implikationen auf die Gesellschaft verbunden sind, gilt es, hier regelmäßig sorgfältig die Chancen und Risiken abzuwägen. Dabei ist es definitiv notwendig, auch die Ethik zu berücksichtigen, damit sichergestellt ist, dass Technologie im besten Sinne sowohl zum Wohl des Einzelnen als auch der Gesellschaft konzipiert und eingesetzt wird. Unterstützung bei der Entwicklung und Umsetzung von KI-basierten Systemen gemäß ethischer Kriterien kann zukünftig auch Tool-basiert – beispielsweise mit dem Ethik.KI.Tool angeboten werden.

Insgesamt müssen diese Anforderungen ernst genommen werden, denn die daraus resultierende Vertrauenswürdigkeit ist maßgeblich für die Akzeptanz der KI-Anwendung. Zusätzlich sollte hier durch eine fundierte Aufklärungsarbeit aller beteiligten Parteien bezüglich der Chancen und Risiken der KI-Technologie Transparenz und damit Verständnis geschaffen werden. ■

ULLA COESTER
ist Gründerin und CEO des Unternehmens xethix-empowerment. Als solche berät sie bei Prozessen zur digitalen Ethik sowie bei Digitalisierungsprojekten. Zudem ist sie Lehrbeauftragte für digitale Ethik (Hochschule Fresenius, Köln) und Mitglied der Standardization Evaluation Group 10/IEC: Ethics in Autonomous and Artificial Intelligence Application.

NORBERT POHLMANN
ist Professor für Informationssicherheit und Leiter des Instituts für Internet-Sicherheit – if(is) an der Westfälischen Hochschule in Gelsenkirchen sowie Vorstandsvorsitzender des Bundesverbands IT-Sicherheit – TeleTrusT und im Vorstand des Internetverbandes – eco.

Literatur:

[1] N. Pohlmann: „Cybersicherheit - Das Lehrbuch für Konzepte, Mechanismen, Architekturen und Eigenschaften von Cybersicherheitssystemen in der Digitalisierung", Springer Vieweg Verlag, Wiesbaden 2019

[2] U. Coester, N. Pohlmann: „Wer macht die Spielregeln für die KI?" IT & Production – Zeitschrift für erfolgreiche Produktion, Technik-Dokumentations-Verlag, 2019