KI in der IT-Sicherheit

Die IT-Defense 2026 befasste sich mit dem Einsatz, aber auch der Abwehr von künstlicher Intelligenz im IT-Sicherheitsbereich.

Die IT-Sicherheitskonferenz „IT-Defense 2026“ zeigte, dass künstliche Intelligenz (KI) zu einem zentralen Faktor im Cyber-Raum geworden ist – sowohl auf Seiten der Angreifer als auch der Verteidiger. Die Veranstaltung Anfang Februar 2026 in Würzburg brachte Fachleute aus Wirtschaft, Behörden und Forschung zusammen, um Bedrohungsszenarien, technologische Entwicklungen und regulatorische Rahmenbedingungen zu diskutieren.

Die Bedrohungslage im Cyber-Raum hat sich weiter verschärft. Laut Stefan Strobel, Geschäftsführer von *Cirosec,* habe die von China und Russland ausgehende Cyber-Bedrohung (Datendiebstahl, Industriespionage, Sabotage) gegenüber 2024 jeweils um 46 Prozent zugenommen und damit die größte Bedrohung dargestellt.

Der weltweit entstandene Schaden durch Cyber-Kriminalität wird auf rund 300 Milliarden Euro geschätzt. Allein in Deutschland wurden zuletzt knapp 1.000 Ransomware-Fälle registriert, während die globalen Lösegeldzahlungen die Marke von 800 Millionen US-Dollar überschreiten. Auffällig ist dabei die zunehmende Professionalisierung der Täterstrukturen. Cyber-Kriminalität wird heute vielfach arbeitsteilig organisiert und als Dienstleistung angeboten („Crime-as-a-Service“). Laut Strobel gehe es bei den verschiedensten Angriffsarten um Geld, Erpressung sowie staatliche und ideologische Interessen. Hacktivistische DDos-Angriffe auf Ziele in Deutschland würden sich fortsetzen und KI zunehmend auch für cyberkriminelle Aktivitäten genutzt.

Im Darknet und für Ransomware-Zahlungen wird *Bitcoin* genutzt. Im Februar 2025 wurde die Kryptobörse *Bybit* gehackt. Dabei wurden etwa 1,5 Milliarden US-Dollar in *Ethereum* gestohlen (in 100 USD-Scheinen würden diese 15.000 kg wiegen). Das war der größte Krypto-Hack bisher. Die Täter werden der *Lazarus-Gruppe* zugerechnet, hinter der Nordkorea vermutet wird. Einmal kompromittierte Accounts werden häufig weiterverkauft; sogenannte Affiliates übernehmen deren weitere kriminelle Nutzung.

Lockpicking: Entsperren mechanischer Schlösser ohne Originalschlüssel

Bei Betrugsformen wie *Crypto-Scams, Romance-Scams* oder *Pig-Butchering* („Anfüttern und dann schlachten“) verfolgen Täter einen systematischen Ansatz zum Aufbau emotionaler Bindungen und tragfähiger Vertrauensverhältnisse. Der Erstkontakt erfolgt dabei häufig unter dem Anschein von Zufälligkeit, etwa durch eine angeblich falsch gewählte Telefonnummer, und dient als Einstieg in eine längerfristig angelegte Manipulationsstrategie. Derartige Angriffe werden häufig professionell organisiert und gehen mutmaßlich von sogenannten *Scam-Centern* aus, die insbesondere in Teilen Asiens lokalisiert werden. Berichte deuten darauf hin, dass dort teils unter Zwang arbeitende Personen in groß angelegten Einrichtungen eingesetzt werden, um massenhaft betrügerische Kontakte herzustellen.

Ein prominenter Fall betraf den Betreiber eines solchen Zentrums in Kambodscha, Chen Zhi, der am 7. Jänner 2026 festgenommen und an China ausgeliefert wurde. Im Zuge der Ermittlungen sollen Vermögenswerte in Höhe von rund 14 Milliarden US-Dollar sichergestellt worden sein. Inzwischen ist eine geografische Ausweitung entsprechender Strukturen zu beobachten: Neben Asien entstehen zunehmend *Cyber-Scam-Center* auch im Mittleren Osten, in Zentralamerika sowie in Westafrika.

Parallel dazu zeigen Strafverfolgungsmaßnahmen die zunehmende Professionalisierung der Täterstrukturen. So wurde im Oktober 2025 im Rahmen der *Operation Simcartel* unter Leitung von Europol ein hochentwickeltes Cybercrime-as-a-Service-Netzwerk zerschlagen. Die Gruppierung betrieb rund 1.200 SIM-Boxen, die wiederum tausende SIM-Karten verwalteten und damit eine systematische Verschleierung von Anruferidentitäten ermöglichten. Im Zuge koordinierter Durchsuchungsmaßnahmen wurden am 10. Oktober 2025 insgesamt sieben Verdächtige in Lettland und Estland festgenommen.

Beim *Business E-Mail Compromise (BEC)* werden Mitarbeitende eines Unternehmens durch vermeintlich vertrauenswürdige Kommunikationspartner – etwa Führungskräfte im Rahmen sogenannter CEO-Fraud-Szenarien – zur Durchführung betrügerischer Geldtransfers veranlasst. Im Unterschied zu klassischen Cyber-Angriffen kommt dabei in der Regel keine Malware zum Einsatz; vielmehr basiert der Angriff auf gezieltem *Social Engineering*.

Durch den Einsatz generativer KI und *Deepfakes* hat diese Betrugsform

Foto: Kuurt Hickisch

eine neue Qualität erreicht. Bereits kurze Audiomitschnitte von etwa zehn Minuten können ausreichen, um täuschend echte synthetische Stimmen für betrügerische Anrufe zu erzeugen. Als Quellenmaterial dienen unter anderem aufgezeichnete Präsentationen, vermeintliche Interviews, Mitschnitte von Team-Meetings oder auch öffentlich zugängliche Analysten-Calls. Zudem wird vor Aktivitäten nordkoreanischer IT-Arbeitskräfte gewarnt, die unter falschen Identitäten in Unternehmen eingeschleust werden oder remote für diese tätig sind. In solchen Fällen besteht das Risiko, dass diese als „trojanische Pferde" fungieren und interne Systeme kompromittieren.

Ein weiteres Beispiel für die sicherheitspolitische Dimension von IT-Risiken zeigte sich im Kontext des NATO-Beitritts Finnlands am 4. April 2023: Unternehmen mit verbliebenen Geschäftsaktivitäten in Russland sahen sich Berichten zufolge gezwungen, innerhalb kürzester Zeit ihre lokalen IT-Systeme unbrauchbar zu machen, um einer drohenden Verstaatlichung zuvorzukommen. Der IT-Experte Sami Laiho schilderte in diesem Zusammenhang technische Vorgehensweisen, mit denen Datenbestände gezielt unlesbar gemacht wurden.

KI-Verordnung. Einen Überblick über die vielfach als „Regelungs-Tsunami" bezeichnete Digitalstrategie der Europäischen Union gab Rechtsanwalt Joerg Heidrich. Ziel dieser regulatorischen Initiativen ist insbesondere die Stärkung der Sicherheit von Daten und Netzwerken sowie die Schaffung eines einheitlichen digitalen Binnenmarkts.

Der Sicherung von Daten und IT-Infrastrukturen dienen unter anderem der bereits in Kraft getretene *Data Governance Act,* die bis Oktober 2026 in nationales Recht umzusetzende *NIS-2-Richtlinie* sowie der *Data Act,* der seit September 2025 teilweise Anwendung findet. Auf die Regulierung digitaler Märkte und Dienste zielen der *Digital Services Act (DSA)* und der *Digital Markets Act (DMA)* ab, die beide bereits in Kraft getreten sind und insbesondere große Plattformanbieter adressieren.

Der *AI-Act* (im Folgenden: KI-Verordnung) verfolgt einen risikobasierten Regulierungsansatz für KI-Systeme. Seine Bestimmungen werden schrittweise bis 2027 wirksam. Auf diesen

Referenten bei der IT-Defense 2026: Jörg Heidrich, Candid Wüest, Stefan Strobel und Jos Wetzels

Rechtsrahmen ging der Vortragende im weiteren Verlauf vertiefend ein. Ziel der *Verordnung (EU) 2024/1689 vom 13. Juni 2024 (KI-Verordnung)* ist es sicherzustellen, dass KI-Systeme sicher, transparent und ethisch ausgestaltet sind und dabei die Grundrechte gewahrt bleiben. Der Regelungsansatz ist primär schutzorientiert und dient insbesondere dem Schutz von Bürgerinnen und Bürgern vor potenziellen Risiken durch KI, weniger hingegen deren Förderung. Die Verordnung adressiert sowohl Anbieter *(Provider)* als auch Betreiber *(Deployer)* von KI-Systemen. Dabei ist zu beachten, dass substanzielle Änderungen an der Zweckbestimmung oder Nutzung eines Systems dazu führen können, dass ein Betreiber regulatorisch als Anbieter qualifiziert wird – mit entsprechend erweiterten Pflichten.

Die KI-Verordnung unterscheidet vier Risikokategorien von KI-Systemen. An deren oberster Stufe stehen Anwendungen mit unannehmbarem Risiko, deren Einsatz grundsätzlich verboten ist (Art. 5 KI-VO). Hierzu zählen insbesondere: der Einsatz manipulativer Techniken zur Verhaltensbeeinflussung, die Ausnutzung von Vulnerabilitäten, etwa aufgrund von Alter, Behinderung oder sozialer Lage, *Social Scoring* durch staatliche oder private Akteure, bestimmte biometrische Anwendungen wie Systeme zur Vorhersage von Straftaten oder zur Emotionserkennung, etwa im Beschäftigungskontext.

Hochrisikosysteme bilden die zweite Stufe der KI-Verordnung. Hierbei handelt es sich um Anwendungen, die ein erhebliches Risiko für die Gesundheit, Sicherheit oder die Grundrechte natürlicher Personen darstellen, jedoch aufgrund ihres sozioökonomischen Nutzens nicht verboten werden. Entsprechend unterliegen sie strengen regulatorischen Anforderungen.

Für Hochrisikosysteme gelten umfassende Pflichten. Dazu zählen insbesondere die Einführung eines Risiko- und Qualitätsmanagementsystems sowie hohe Anforderungen an Datenqualität und -integrität. Die Konformität mit der KI-Verordnung ist durch eine CE-Kennzeichnung nachzuweisen. Anbieter sind verpflichtet, ihre Systeme vor dem Inverkehrbringen einer Konformitätsbewertung zu unterziehen. Verstöße können mit Bußgeldern von bis zu 15 Millionen Euro oder 3 % des weltweiten Jahresumsatzes sanktioniert werden.

Innerhalb der Hochrisikosysteme wird weiter differenziert: Einerseits betrifft dies KI als Sicherheitskomponente in regulierten Produkten, etwa in medizinischen Geräten, Aufzügen, Fahrzeugen oder Luftfahrtsystemen. Andererseits umfasst es eigenständige KI-Systeme mit potenziellen Auswirkungen auf Grundrechte. Diese sind in Anhang III der KI-Verordnung aufgelistet und betreffen unter anderem biometrische Identifikation und Kategorisierung, kritische Infrastrukturen, Bildung und Ausbildung, Beschäftigung und Personalmanagement, den Zugang zu wesentlichen Dienstleistungen, Strafverfolgung, Migration und Grenzkontrolle sowie Justiz und demokratische Prozesse.

Die spezifischen Anforderungen an Hochrisiko-KI-Systeme sind in Abschnitt 2 der Verordnung geregelt, während Abschnitt 3 die Pflichten von Anbietern, Betreibern und weiteren Beteiligten festlegt. Ergänzend ist gemäß Art. 27 eine Grundrechte-Folgenabschätzung durchzuführen.

Besondere Relevanz entfaltet dies im Personalwesen: Anwendungen zur Bewerberauswahl, zur Bewertung von Leistung oder zur Entscheidung über Beförderungen und Kündigungen fallen regelmäßig in den Hochrisikobereich. Entsprechend sind hier strenge Anforderungen an Transparenz, Nachvollziehbarkeit und Risikomanagement zu erfüllen.

Fotos: Kurt Hickisch

Der risikobasierte Ansatz der Verordnung führt insgesamt zu hohen technischen und organisatorischen Anforderungen, insbesondere im Bereich der Cybersicherheit. Hierzu zählen spezialisierte Sicherheitsarchitekturen sowie kontinuierliche Überwachungs- und Kontrollmechanismen.

Die dritte Stufe umfasst KI-Systeme mit begrenztem Risiko. Diese sind weder als unzulässig noch als hochriskant eingestuft, unterliegen jedoch Transparenzpflichten. Insbesondere müssen sie als KI-Systeme erkennbar sein. Beispiele hierfür sind *Chatbots*, *Deepfakes*, Systeme zur Emotionserkennung oder zur Analyse von Kundenstimmungen. Die technischen Anforderungen beschränken sich im Wesentlichen auf grundlegende Maßnahmen zur Sicherstellung der Systemintegrität sowie auf Transparenzfunktionen.

Die vierte Stufe betrifft KI-Systeme mit minimalem Risiko, etwa in Form von KI-gestützten Spielen, Spamfiltern oder kreativer Software. Diese unterliegen keinen spezifischen regulatorischen Anforderungen und sind frei nutzbar.

Cybercrime und Bedrohungslage. Cyber-Vorfälle stellen derzeit eines der größten Geschäftsrisiken dar. In Deutschland belief sich die Schadenssumme im Jahr 2025 auf rund 47 Milliarden Euro (2024: 36 Milliarden Euro). Vertreter von Strafverfolgungsbehörden betonten die zunehmende Professionalisierung und Internationalisierung der Täterstrukturen. Ein prägnantes Beispiel für die Auswirkungen von Cyber-Angriffen ist ein international tätiger Hersteller von Waagen, bei dem es nach einem Angriff über 1.000 Tage dauerte, die IT-Infrastruktur weltweit vollständig zu erneuern. Dies verdeutlicht die zentrale Bedeutung von Cyber-Sicherheit: Sie schafft zwar keinen unmittelbaren Mehrwert im Vertrieb, ist jedoch Voraussetzung für wirtschaftliche Handlungsfähigkeit.

Die aktuelle Bedrohungslage ist insbesondere durch Datendiebstahl und „Crime-as-a-Service"-Modelle geprägt. Täter agieren global vernetzt und arbeitsteilig. Exemplarisch wurde die Ransomware „GandCrab" genannt, die 2019 unter anderem die Württembergischen Staatstheater sowie zahlreiche Unternehmen in Deutschland traf. Nach mehrjährigen Ermittlungen konnte ein beteiligter Täter identifiziert,

Verbrecherjagd im Cyber-Raum: Oberstaatsanwalt M. Heim, EKHK Daniel Lorch

ausgeliefert und im Januar 2026 zu einer Freiheitsstrafe verurteilt werden. Auch gegen Mitglieder der Nachfolgegruppierung „REvil" wurden international Strafmaßnahmen ergriffen. Parallel dazu werden Strafverfolgungsmaßnahmen intensiviert: Internationale Kooperationen führten etwa zur Zerschlagung der Hackergruppe „Hive", selbst unter den erschwerten Bedingungen des Ukraine-Kriegs.

Den ersten dokumentierten Einsatz von KI als Schadsoftware setzte Cyber-Sicherheitsexperte Candid Wüest mit *LameHug* im Juli 2025 an. *LameHug* kann dank der KI-Integration seine Aktionen an die Umgebung des Opfers anpassen („adaptive Malware"). Für Forschungszwecke wurde *Yutani Loop* entwickelt, als Beispiel für eine agentengestützte KI-Malware *(agentic AI-Malware)*, die eigenständig Ziele wählt und ihre Vorgangsweise anpassen kann. Angriffe werden durch KI automatisiert und beschleunigt. Dessen ungeachtet sah Wüest in KI-gestützter Schadsoftware lediglich eine Erweiterung bestehender Bedrohungen, der mit herkömmlichen Mitteln begegnet werden kann – sofern diese ordnungsgemäß eingesetzt werden.

Chris Wysopal referierte über den Einsatz von KI beim Erstellen von Codes und das dadurch erfolgte Wiederaufleben bereits überwunden geglaubter Schwachstellen. Über eine vom Unternehmen *Cirosec* für Analysewecke selbst entwickelte Software *(Allpacka)* berichtete Leon Schmidt dieses Unternehmens.

IT Technik. Elektronische Brieftaschen *(Digital Wallets)* sind als Zahlungsmittel bereits weit verbreitet – aber wo und wie sicher sind die dahinter stehenden persönlichen Daten hinterlegt? Über das beim deutschen Bundesministerium für Digitales und Staatsmodernisierung (BMDS) laufenden Projekt, das auch physische Berechtigungsnachweise wie Personalausweis, Führerschein, Gesundheitskarte, einschließt, berichtete Torsten Lodderstedt. Anfang 2027 soll die Nutzung dieser Dokumente auf dem Smartphone möglich sein.

Über die Sicherheit des im BOS-Bereich genutzten digitalen Bündelfunks (Funk, Telefonie, Datendienste) TETRA berichtete Jos Wetzels, *Midnight Blue*. Über Versuche, Chips über elektromagnetische Strahlung zu hacken, berichtete Thomas Roth. Bei einem Ausstellungsstand im Foyer des Vortragssaals konnten sich die Besucher im Entsperren mechanischer Schlösser *(Lockpicking)* üben und, bei Zeitmessung, dies wettkampfmäßig betreiben.

IT-Defense. An der Veranstaltung haben etwa 300 Fachleute aus dem IT-Bereich, von Unternehmen, Militär, Polizei und Verwaltung, teilgenommen. Die IT-Defense 2027 wird vom 27. bis 29. Jänner 2027 in Frankfurt am Main stattfinden. *Kurt Hickisch*

Foto: Kurt Hickisch