Schweizerische Eidgenossenschaft
Confédération suisse
Confederazione Svizzera
Confederaziun svizra

Der Bundesrat

Bern, 12. Dezember 2025

Chancen und Risiken von KI-Systemen in der Cybersicherheit

Bericht des Bundesrates in Erfüllung des Postulates 23.3861 Andrey vom 15. Juni 2023

Inhaltsverzeichnis

1 Einleitung

1.1 Allgemeines

Künstliche Intelligenz (KI) ist ein Thema, welches die Menschheit seit Jahrhunderten fasziniert. Schon in der Antike entwarfen Denker Ideen von Robotern und intelligent agierenden Maschinen. Auch in der Alchemie des Mittelalters war die Schöpfung von künstlichem menschenähnlichem Leben eines der zentralen Themen.[1] Während solche Überlegungen zur KI lange Jahre vor allem unter Philosophen und Mystikern geführt wurden, hat die heutige Diskussion einen viel direkteren Bezug zum gelebten Alltag. Am Beginn dieser Entwicklung steht die wegweisende Veröffentlichung des britischen Mathematikers Alan Turing im Jahr 1950 «Computing Machinery and Intelligence». Darin stellte Turing die zentrale Frage, ob Maschinen denken können und ab wann maschinelle Intelligenz vorliegt.[2] Seither wird intensiv zu KI geforscht und es wurden massive Fortschritte erzielt. Auch wenn umstritten bleibt, ob die heute verfügbaren Anwendungen tatsächlich als «Intelligenz» beschrieben werden können, ist unbestritten, dass moderne KI-Systeme zahlreiche neue Möglichkeiten eröffnen und menschliche Fähigkeiten schon in vielen Bereichen ergänzen und übertreffen. Dabei ist davon auszugehen, dass wir erst am Anfang der Entwicklung dieser neuen Technologien stehen. Die enormen Investitionen in diesen Bereich zeigen, wie hoch die Erwartungen und der Glaube an weiteren Fortschritt sind.[3]

Die rasche Entwicklung weckt aber auch Skepsis. Neben grundsätzlichen Bedenken, ob KI der Menschheit mehr Schaden als Nutzen bringen könnte, bestehen auch sehr konkrete Bedenken über die Auswirkungen der Verwendung von KI-Systemen in unserem Alltag. Ein zentraler Punkt betrifft die Cybersicherheit. Die fortschreitende Digitalisierung hat zu sehr hohen Effizienzgewinnen geführt und ist ein zentraler Treiber des wirtschaftlichen und gesellschaftlichen Fortschritts. Zugleich sind wegen der Digitalisierung auch neue Verwundbarkeiten entstanden. Schon heute stehen wir vor der Situation, dass viele der für die Gesellschaft zentralen Dienstleistungen, wie die Energie- oder medizinische Versorgung[4] sehr stark von funktionierenden IT-Systeme abhängig sind. Wegen der hohen Komplexität dieser Systeme ist es aber auch schwierig, diese vor Risiken zu schützen und Schwachstellen rechtzeitig zu erkennen. Dies macht die Cybersicherheit bereits heute zu einer anspruchsvollen Aufgabe.

Mit den heute verfügbaren Anwendungen der KI wird die Komplexität und die Abhängigkeit von IT-Systemen nochmals erhöht. Es ist für den Menschen zunehmend schwierig nachzuvollziehen, wie KI-Systeme ihre Ergebnisse ermitteln und es ist derzeit noch unklar, welches volle Potenzial KI-Anwendungen für verschiedene Einsatzbereiche entfalten können. Die neuen technologischen Möglichkeiten scheinen daher die Cyberbedrohungen nochmals zu verstärken. Es darf aber nicht ausser Acht gelassen werden, dass diese Möglichkeiten auch genutzt werden können, um Komplexität besser zu verstehen und Risiken frühzeitig zu erkennen. Damit kann KI direkt dazu beitragen, die Cybersicherheit zu erhöhen.

Da sich die Technologie in kurzer Zeit weiterentwickelt und verändert, ist offen, ob KI langfristig eher zur Stärkung der Cybersicherheit oder zur Verschärfung von Bedrohungen beiträgt. Die Entwicklung und der Einsatz von KI können sowohl positive als auch negative Auswirkungen auf die Sicherheit digitaler Systeme und Daten haben. Es ist wichtig sich dieser Fragestellung aktiv zu widmen und eine Einschätzung auf Basis des aktuellen Wissens vorzunehmen. Der vorliegende Bericht dient diesem Zweck. Er soll zudem aufzeigen, welche Massnahmen durch die Verwaltung und Partnerorganisationen der Nationalen Cyberstrategie (NCS) bereits heute umgesetzt werden, und wo allfälliger Handlungsbedarf besteht.

[1] Seowerk. (o. J.). *Die Evolution der Künstlichen Intelligenz: Eine Reise von antiken Visionen bis zur Gegenwart*. SEOWerk. Abgerufen am 5. Oktober 2025, von https://www.seowerk.de/news/die-evolution-der-kuenstlichen-intelligenz-eine-reise-von-antiken-visionen-bis-zur-gegenwart
[2] A. M. Turing (1950) Computing Machinery and Intelligence. Mind, volume 59: 433-460.
[3] Struta, I. (2025, 22. Januar). *GenAI funding hits record in 2024 boosted by infrastructure interest*. S&P Global Market Intelligence, von https://www.spglobal.com/market-intelligence/en/news-insights/articles/2025/1/genai-funding-hits-record-in-2024-boosted-by-infrastructure-interest-87132257
[4] Bundesamt für Bevölkerungsschutz (BABS). (2024). *Schutz kritischer Infrastrukturen*. Schutz kritischer Infrastrukturen

1.2 Prüfauftrag

Ausschlaggebend für diesen Bericht ist das am 15. Juni 2023 überwiesene Postulat 23.3861 Andrey «Chancen und Risiken von KI-Systemen in der Cybersicherheit». Dieses beauftragt den Bundesrat, die folgenden Fragestellungen zu prüfen:

23.3861 Postulat - Chancen und Risiken von KI-Systemen in der Cybersicherheit, Gerhard Andrey

> «Der Bundesrat wird beauftragt, in einem Bericht darzulegen, inwiefern das Aufkommen von Machine Learning, Deep Learning oder Large Language Models - landläufig als Künstliche Intelligenz (KI) bezeichnet - generell Cybersicherheit beeinflusst, welche Chancen sich daraus ergeben und wie mit den Risiken umzugehen ist. Der Bericht soll ebenfalls aufzeigen, inwiefern die Nationale Cyberstrategie NCS den sich rasch verändernden KI-Entwicklungen gerecht wird und welche Schlüsse aus einer solchen Analyse zu ziehen sind.»

Der Bundesrat hat am 16. August 2023 die Annahme des Postulats beantragt. Diesem Antrag ist der Nationalrat nachgekommen. Er hat am 29. September 2023 das Postulat angenommen. Der vorliegende Bericht zeigt auf, inwiefern das Aufkommen von KI die Cybersicherheit der Schweiz beeinflusst und welche Chancen und Risiken sich daraus ergeben. Es werden hierbei zunächst die technischen Grundlagen der verschiedenen KI-Technologien erläutert und deren Funktionsweise in der Cybersicherheit verständlich dargestellt. Darauf aufbauend werden die konkreten Anwendungen von KI analysiert – sowohl als Bedrohung durch Cyberkriminelle als auch als Chance für die Cyberabwehr. Anschliessend wird untersucht, inwiefern die bestehende Nationale Cyberstrategie und die daraus abgeleiteten Massnahmen den sich rasch verändernden KI-Entwicklungen gerecht werden und welche regulatorischen Aspekte zu berücksichtigen sind. Abschliessend fasst der Bericht die Ergebnisse zusammen und weist aus, ob und wo Handlungsbedarf besteht, um das Thema KI in der Cybersicherheit adäquat zu berücksichtigen.

2 Entstehungsgeschichte und Definition von KI

Der Begriff Künstliche Intelligenz (KI) wurde 1955 vom Informatiker John McCarthy geprägt und ursprünglich als «Wissenschaft und Technik intelligenter Maschinen» definiert.[5] Dabei handelt es sich um eine wissenschaftliche Disziplin mit dem Ziel, künstliche Systeme zu entwickeln, die Aufgaben übernehmen können, für deren Bewältigung menschliches rationales Denken oder Intelligenz erforderlich ist.[6] Heute wird unter KI meist nicht mehr eine wissenschaftliche Disziplin verstanden, sondern die Systeme und Anwendungen, welche durch die Forschungs- und Entwicklungsarbeiten in diesem Bereich entstanden sind. Weil sehr viele unterschiedliche Systeme und Anwendungen als KI bezeichnet werden, muss zunächst definiert werden, wie der Begriff im vorliegenden Bericht verwendet wird. Zum besseren Verständnis der verwendeten Terminologie wird kurz begriffsgeschichtlich eingeordnet, welche Entwicklungen zu den heute gebräuchlichen Begriffen geführt haben.

In diesem Bericht wird der Begriff «Künstliche Intelligenz» gemäss der Definition des Kompetenznetzwerks für Künstliche Intelligenz der Bundesverwaltung (CNAI) und analog der Strategie zum Einsatz von KI-Systemen in der Bundesverwaltung[7] verwendet. Das CNAI definiert KI als «*als «einen Computer*

[5] Bonfanti, M., Kohler, K. Künstliche Intelligenz für die Cybersicherheit. CSS Analysen zur Sicherheitspolitik 265 (2020), https://doi.org/10.3929/ethz-b-000417506

[6] CNAI (2024). *Terminologiepapier künstliche Intelligenz*, Version 2.1. https://cnai.swiss

[7] Für den Einsatz von KI-Systemen in der Bundesverwaltung verfügt die Bundesverwaltung seit März 2025 über eine sog. Teilstrategie: KI-Systeme sollen in der Bundesverwaltung verantwortungsvoll eingesetzt werden und sie sollen helfen, die Verwaltungsprozesse effizienter auszugestalten. Um dies zu erreichen, definiert die Strategie die drei Handlungsfelder «Kompetenzen aufbauen», «Vertrauen verdienen» und «Effizienz» steigern. Bundeskanzlei BK. (2025, 11. April). *SB021 – Strategie Einsatz von KI-Systemen in der Bundesverwaltung*. https://www.bk.admin.ch/bk/de/home/digitale-transformation-ikt-lenkung/vorgaben/sb021-strategie-einsatz-von-ki-systemen-in-der-bundesverwaltung.html

so bauen oder programmieren, um Dinge zu tun, die normalerweise menschliche oder biologische Fähigkeiten («Intelligenz») erfordern», z.B. visuelle Wahrnehmung (Bilderkennung), Spracherkennung, Sprachübersetzung, visuelle Übersetzung und Spiele spielen (mit konkreten Regeln). Bei KI geht es um «intelligente» Maschinen («smart machines»), die Aufgaben ausführen können, die normalerweise von Menschen ausgeführt werden («lernende Maschinen»; «learning machines»), d. h. Maschinen «intelligent» machen.».[8] Seit Mitte des 20. Jahrhunderts wurden Anwendungen der KI entwickelt, die dieser Definition entsprechen.[9] Ein frühes Beispiel für die Entwicklung der KI war das Programm ELIZA, das 1966 von Joseph Weizenbaum am MIT entwickelt wurde.[10] Das Programm konnte durch einfache Regeln einen therapeutischen Dialog simulieren und dadurch eine überraschende emotionale Wirkung erzielen. ELIZA gilt wegen seiner psychologischen Wirkung als ein Meilenstein in der frühen Geschichte der KI. Ein weiterer Meilenstein war die Einführung der logikbasierten Programmierung mit Sprachen wie Prolog, die komplexe Wissenssysteme abbildeten und die Grundlage für Expertensysteme der 1980er und 1990er Jahre bildeten und die KI-Forschung prägten.[11]

Parallel wurde vermehrt an «Machine Learning» und «Deep Learning» geforscht, woraus sich die heute verfügbaren Modelle der generativen KI entwickelt haben. Als generative KI werden Modelle bezeichnet, die mithilfe grosser Datenmengen aus der physischen und digitalen Welt trainiert werden, um eigenständig neue, multimodale Inhalte wie Texte, Bilder, Töne, Videos, Simulationen oder Programmcode zu erzeugen. Eine zentrale Untergruppe der generativen KI sind die grossen Sprachmodelle (Large Language Models, LLMs).[12] LLMs bilden die Grundlage vieler gängiger KI-Systeme im Bereich der natürlichen Sprache. Sie sind in der Lage, natürliche Sprache zu verarbeiten, und eigenständig zu erzeugen. Sie können verschiedene Aufgaben übernehmen, wie Übersetzungen, Textverständnis oder die Erstellung neuer Texte.

LLMs werden mit grossen, vielfältigen Datenmengen trainiert und erzeugen Texte, die oft kaum von menschlich verfassten Inhalten zu unterscheiden sind. Obwohl sie primär auf Textdaten basieren, können sie zunehmend auch akustische und visuelle Eingaben verarbeiten, indem diese zuverlässig in Text umgewandelt werden. Der Durchbruch kam vor allem in den letzten fünf Jahren mit der rasanten Entwicklung von LLMs.[13] Bekannte Beispiele für LLMs sind die GPT-Modellreihen[14] (wie in «ChatGPT» von der Firma OpenAI und Microsofts «Copilot»), Googles «Gemini» (früher «Bard»), Metas «LLaMA»-Modelle sowie die «Grok»-Modelle von X und Anthropics «Claude»-Modelle.[15] Die Modelle sind heute für die breite Öffentlichkeit zugänglich. Durch die öffentliche Nutzung konnte die Feinabstimmung der Gesprächsführung rasch weiterentwickelt werden und die Modelle haben sich innerhalb kurzer Zeit erheblich verbessert.[16] Durch die rasche Verbreitung der Anwendungen von LLMs hat sich die öffentliche und fachliche Diskussion stark auf diese Form der KI konzentriert. Im Mittelpunkt dieses Berichts stehen

[8] CNAI (2024), s. 7
[9] Teich, I. Meilensteine der Entwicklung Künstlicher Intelligenz. *Informatik Spektrum* 43, 276–284 (2020). https://doi.org/10.1007/s00287-020-01280-5
[10] Weizenbaum, Joseph. "ELIZA—a computer program for the study of natural language communication between man and machine." *Communications of the ACM* 9.1 (1966): 36-45.
[11] Teich, I. (2020), s.277
[12] CNAI (2024), s. 8
[13] In dem Buch „Large Language Models in Cybersecurity: Threats, Exposure and Mitigation“ von Andrei Kucharavy et al. (2024) wird ein Deep Neural Language Model (DNLM) als ein spezialisiertes Deep Learning-Modell beschrieben, das darauf trainiert ist, natürliche Sprache zu verstehen und zu generieren. Diese Modelle basieren auf neuronalen Netzwerken und werden mit maschinellem Lernen entwickelt, um semantische Zusammenhänge und Kontextinformationen in Texten zu erfassen. Sie verwenden Techniken wie Token-Embedding und selbstüberwachtes Lernen, um die Bedeutung von Wörtern und Sätzen zu erlernen und anzuwenden. DNLMs sind die Grundlage für die Entwicklung von Large Language Models (LLMs), die durch ihre Skalierung und Leistungsfähigkeit eine bedeutende Rolle in der Weiterentwicklung von KI-Agenten spielen.
[14] GPT, kurz für „Generative Pre-trained Transformer“, bezeichnet eine Gruppe von Textgeneratoren, die auf tiefen neuronalen Netzwerken basieren. Im November 2022 erschien ChatGPT-3, ein großes Sprachmodell (LLM), das mithilfe von über 175 Milliarden Textfragmenten trainiert wurde. Das Modell lernt dabei, wie wahrscheinlich bestimmte Wortfolgen sind, und erzeugt Texte, indem es diese Wahrscheinlichkeiten nutzt. Die von ChatGPT produzierten Texte beruhen also nicht auf menschlicher Kreativität oder echtem Verständnis, sondern auf statistischen Mustern, die das System aus den Trainingsdaten gelernt hat. Eckert, C. ChatGPT, generiere mir einen Hackerangriff – Wie Künstliche Intelligenz Cybersicherheit neu definiert (2023), Fraunhofer-Institut für angewandte und Integrierte Sicherheit AISEC, s. 2ff
[15] Kompetenznetzwerk für künstliche Intelligenz (CNAI). (2024, 26. April). *Merkblatt zur Sensibilisierung betreffend grosse KI-Sprachmodelle in der Bundesverwaltung (Version 1.0)*. Eidgenössisches Departement des Innern (EDI). Abgerufen am 5. Oktober 2025, von https://cnai.swiss/dienstleistungen-weitere-dienstleistungen-merkblaetter-zu-ki/
[16] Kucharavy A. and al, Fundamentals of Generative Large Language Models and Perspectives in Cyber-Defence (2023), Cyber Defence Campus, s.1, armasuisse https://doi.org/10.48550/arXiv.2303.12132;
National Cyber Security Centre (NCSC). (n.d.). *The near-term impact of AI on the cyber threat*. Abgerufen am 5. Oktober 2025, von https://www.ncsc.gov.uk/report/the-near-term-impact-of-ai-on-the-cyber-threat ;
Reuters. (2025, 18. März). Europol warns of AI-driven crime threats. Reuters. Abgerufen am 5. Oktober 2025, von https://www.reuters.com/world/europe/europol-warns-ai-driven-crime-threats-2025-03-18/

generative KI-Systeme, die auf der Verarbeitung natürlicher Sprache beruhen. Im folgenden Abschnitt wird definiert, welche Auswirkungen diese Modelle auf die Cybersicherheit haben.

3 Anwendungen von KI in der Cybersicherheit

Automatisierungen und die Analyse von grossen Datenmengen spielen seit den frühen 1980er eine zentrale Rolle in der Cybersicherheit.[17] Im Wettstreit zwischen Angreifenden und Verteidigenden ist es von entscheidender Bedeutung, dass Prozesse zur Suche von Schwachstellen und exponierten Systemen möglichst effizient durchgeführt und neue Angriffsvektoren respektive neue Gegenmassnahmen so rasch wie möglich entwickelt werden können. Sowohl für die Angreifenden als auch für die Verteidigenden ist es darum entscheidend, stets auf die neusten technischen Entwicklungen zurückgreifen zu können. Wie KI in der Cybersicherheit verwendet werden kann und wie sie die Bedrohungslage beeinflusst, wurde deshalb lange vor den ersten öffentlich zugänglichen KI-Modellen der 2010er Jahre diskutiert. Seither hat das Thema an Relevanz gewonnen und wird in der Fachliteratur ausführlich besprochen. Verschiedene akademische Publikationen[18], Fachpublikationen von Behörden[19] und Beiträge in verschiedenen Medien haben das Thema aufgenommen. In diesem Kapitel werden die wichtigsten aktuellen Erkenntnisse zu den Auswirkungen der KI auf die Cybersicherheit ausgeführt.

3.1 Bedrohungen durch KI

In diesem Unterkapitel wird aufgezeigt, wie KI die Cyberbedrohungslage beeinflusst. Dabei stehen drei zentrale Aspekte im Fokus: Erstens wird untersucht, inwiefern KI bestehende Bedrohungen, insbesondere durch «Social Engineering», verstärken kann; es wird aufgezeigt ob und wann in die IT-Systeme von Organisationen eingebaute KI zu neuen Schwachstellen in der Cybersicherheit dieser Organisationen führt und schliesslich wird untersucht, ob KI-Modelle dazu führen, dass vollständig automatisierte Angriffe durchgeführt werden können.

3.1.1 KI als Hilfsmittel für Social Engineering

Unter «Social Engineering» versteht man die Nutzung von Psychologie und Manipulation für Cyberangriffe, um an wertvolle Informationen oder Geld zu kommen.[20] Dabei hat die Verfügbarkeit von leistungsfähigen KI-Tools die Einstiegshürden für Cyberangriffe erheblich gesenkt. Insbesondere generative KI-Technologien ermöglichen die Erstellung täuschend echter Inhalte, die für Betrug und Manipulation eingesetzt werden.

Social Engineering liegt vielen Cyberangriffen zu Grunde. Typische Beispiele sind: «Phishing, «Vishing, Smishing»[21], Betrügerische Supportanrufe[22] und CEO-Betrug[23]. Dabei ist das Ziel der Täter, die Opfer

[17] Anderson, J. P. (1980, 26. Februar). *Computer Security Threat Monitoring and Surveillance* (Bericht Nr. 79F296400). James P. Anderson Co. (CERIAS)
[18] Einen guten Überblick über den Stand der Forschung zur Auswirkung von KI auf die Cybersicherheit gibt die Die Open-Access-Studie «Large Language Models in Cybersecurity – Threats Exposure and Mitigation» (2024) des Cyber Defence Campus der armasuisse. Large Language Models in Cybersecurity
[19] Bundesamt für Sicherheit in der Informationstechnik (BSI). (2024, 23. April). *Einfluss von KI auf die Cyberbedrohungslandschaft* (Version 1.0). https://www.bsi.bund.de/SharedDocs/Downloads/DE/BSI/KI/Einfluss_KI_auf_Cyberbedrohungslage.html /
[20] Quelle: Social Engineering (BACS)
[21] Phishing: Betrüger versuchen, durch gefälschte E-Mails an vertrauliche Daten wie Zugangsdaten oder Kreditkartennummern zu gelangen. Die E-Mails enthalten Links zu täuschend echten Nachahmer-Webseiten, auf denen Nutzer ihre Daten unbewusst preisgeben. Vishing: Betrüger rufen an und fälschen dabei oft die angezeigte Telefonnummer, um als vertrauenswürdige Personen (z. B. Bankmitarbeiter oder Polizisten) aufzutreten. Sie üben Druck aus und fordern vertrauliche Informationen, die später missbraucht werden. Smishing: Über SMS geben sich Betrüger als seriöse Absender wie Paketdienste aus und locken Nutzer mit gefälschten Benachrichtigungen auf manipulierte Webseiten, um persönliche Daten oder Kreditkartendetails zu stehlen. Quelle: Phishing, Vishing, Smishing (BACS)
[22] Betrügerische Supportanrufe geben vor, von einer IT-Firma (meist Microsoft) zu sein und behaupten, der Computer sei infiziert. Sie drängen die Angerufenen, eine Software zu installieren, die den Angreifern Zugriff auf das System ermöglicht. Ziel ist es, meist eine Softwarelizenz oder kostenpflichtige Dienstleistung zu verkaufen und dabei Kreditkartendaten zu erlangen. Quelle: Betrügerische Supportanrufe (BACS)
[23] CEO-Betrug: Bei dieser Masche geben sich Betrüger als Chef oder Präsident aus und fordern dringend Zahlungen an, während die echte Führungskraft für Rückfragen nicht erreichbar ist. Sie sammeln vorab Informationen über das Zielunternehmen oder die Organisation, um einen glaubwürdigen, massgeschneiderten Angriff per E-Mail an Finanzverantwortliche durchzuführen und so Zahlungen zu erwirken. Quelle: CEO-Betrug (BACS)

durch gezielte Beeinflussung dazu zu bringen, bestimmte Handlungen auszuführen, etwa vertrauliche Informationen preiszugeben, Produkte zu kaufen oder finanzielle Mittel freizugeben. Die Täter beobachten das persönliche Umfeld ihrer Zielpersonen, geben sich als jemand anderes aus oder machen sich grundlegende menschliche Eigenschaften wie Hilfsbereitschaft, Vertrauen, Neugier, Angst oder Respekt gegenüber Autoritäten zunutze, um an Informationen zu gelangen oder die Zielperson zu bestimmten Handlungen zu verleiten.

Durch den Einsatz von KI ergeben sich im Bereich Social Engineering neue und vielfältige Möglichkeiten, solche Angriffe mit vergleichsweise geringem Aufwand und wenig Fachwissen durchzuführen.[24] Studien zeigen, dass gezielte Phishing-Nachrichten durch LLMs wie ChatGPT in kurzer Zeit und ohne aufwändige Recherche zu den Zielpersonen umgesetzt werden können, vorausgesetzt, die in den LLMs eingebauten ethische Schutzmechanismen werden umgangen. Dafür existieren bereits zahlreiche Methoden die KI dazu zu bringen, Phishing-Nachrichten zu generieren.[25] Damit können zahlreiche individualisierte, sprachlich überzeugende Phishing-E-Mails generiert werden, die herkömmliche Erkennungssysteme umgehen. Diese Nachrichten können in verschiedenen Sprachen und gezielt für bestimmte Branchen, Länder oder bestimmte Personen formuliert werden. Besonders relevant für die Schweiz: Einige Modelle sind bereits mehrsprachig, so zeigt ChatGPT beispielsweise die Fähigkeit, Schweizer Dialekte zu erzeugen.[26]

Ein weiteres Beispiel ist die zunehmende Gefährdung durch CEO-Betrug mittels «Deep Fakes». Mithilfe von KI werden täuschend echte Bilder, Videos oder Audiodateien erstellt, mit denen Kriminelle die Stimme oder das Gesicht von Führungspersonen imitieren können. So werden Mitarbeitende durch gefälschte Sprachnachrichten oder Videos dazu verleitet, betrügerische Überweisungen vorzunehmen.[27] Zu den gängigsten Techniken von Deep Fake gehören das «Face Swapping», bei dem ein Gesicht in einem Video durch ein anderes ersetzt wird sowie das «Facial Reenactment», bei der die Mimik und Lippenbewegungen einer Person auf eine andere Person übertragen werden. Ein Beispiel dafür sind gefälschte Werbeanzeigen mit öffentlichen Personen, in denen vermeintlich Bundespräsidentin Karin Keller-Sutter Werbung für betrügerische Investitionsplattformen macht.[28]

Insgesamt ist festzustellen, dass die Verwendung von KI für Social Engineering bereits verbreitet ist. Es muss davon ausgegangen werden, dass der Trend zu mehr und qualitativ besseren Social Engineering Angriffen weitergehen wird. Zudem werden die verwendeten Modelle durch die häufige Anwendung stets besser, so dass es immer schwieriger wird, reale von gefälschten Inhalten zu unterscheiden. Die qualitative und quantitative Entwicklung im Bereich des Social Engineerings ist ein unmittelbarer Effekt von KI und zum heutigen Zeitpunkt jener Faktor, bei welchem KI die Cyberbedrohungslage am stärksten direkt beeinflusst.

3.1.2 KI als Instrument für Cyberangriffe

Der Nutzen von KI für Cyberangriffe beschränkt sich jedoch keineswegs nur auf Social Engineering. KI kann auch dazu verwendet werden, Schwachstellen zu suchen, Angriffsvektoren zu modifizieren oder neu zu entwickeln und die Effektivität von Cyberangriffen durch eine gezielte Vor- und Nachbereitung zu erhöhen.

Bei der Suche nach Schwachstellen kann die Fähigkeit von KI genutzt werden, in sehr grossen Datenmengen innert kürzester Zeit Muster zu erkennen und diese auszuwerten. Deshalb kann KI verwendet werden, um Sicherheitslücken im Quellcode in Software zu analysieren oder flächendeckend nach

[24] Kraut, Dipl-Ing Hubert, and FH-Prof Dipl-Ing Alexander Mense. "Einfluss und Auswirkungen von Artificial Intelligence auf Social Engineering." (2024), s.1
[25] Kraut (2024), s.9
[26] Kucharavy A. and al, Fundamentals of Generative Large Language Models and Perspectives in Cyber-Defence (2023), Cyber Defence Campus armasuisse https://doi.org/10.48550/arXiv.2303.12132
[27] Siehe: Woche 14: Online-Meeting mit Deep-Fake-Chef: CEO-Betrug 2.0
[28] Siehe: Woche 35: Betrug mit bekannten Persönlichkeiten – Die Schattenseiten der künstlichen Intelligenz

Schwachstellen zu suchen.[29] Solche Recherchen wurden zwar bisher auch automatisiert durchgeführt, die KI verringert jedoch den Aufwand und das benötigte technische Vorwissen dafür beträchtlich. Eine Studie zum Einsatz generativer KI bei der Suche von Software-Schwachstellen zeigt, dass mit Hilfe von KI auch sogenannte «Zero-Day-Angriffe»[30] automatisiert durchgeführt werden können. Allerdings ergab die Untersuchung auch, dass solche Angriffe bislang nur dann erfolgreich waren, wenn die KI auf öffentlich zugängliche Informationen über Schwachstellen zugreifen konnte.[31]

Wie stark KI die Entwicklung von neuen Angriffsvektoren beeinflussen wird, ist noch nicht vollständig absehbar. KI trägt sicher dazu bei, dass schädliche Software («Malware»)[32] mit deutlich weniger Fachwissen entwickelt werden kann. Wie bei der Entwicklung von Methoden für Social Engineering haben verschiedene Studien gezeigt, dass Schutzmechanismen von KI-Systemen, welche verhindern sollten, dass KI zur Entwicklung von Malware genutzt wird, umgangen werden können.[33]

Der Hauptnutzen von KI für die Angreifenden besteht darin, dass sie die Entwicklung der Angriffsmethodik effizienter macht, wodurch Angreifende weniger Zeit in die Erarbeitung investieren müssen. Sie können zudem KI verwenden, um die von ihnen entwickelten Methoden zu testen, zu verbessern und zu modifizieren. Studien zeigen, dass KI bereits heute zur Verbesserung oder zum Verbergen von Schadsoftware eingesetzt wird. Dadurch kann die Schadsoftware durch KI typische Erkennungssysteme umgehen und die Abwehr von Angriffen erheblich erschweren.[34]

Dennoch sind präzise Eingaben («Prompts») erforderlich, um funktionsfähigen Angriff-Code zu generieren. Dies setzt mindestens ein gewisses Mass fachlicher Expertise in den Bereichen Cybersicherheit und Programmierung voraus.[35] Es sind bisher keine Fälle bekannt, bei welcher KI selbständig für ausgewählte Ziele neue Angriffsvektoren erfolgreich entwickelt hat.[36] Da dazu eine grosse Menge an Trainingsdaten nötig wäre und die KI die Gelegenheit haben müsste, den Angriffsvektor zu testen und weiterzuentwickeln, bleiben die Hürden für die Entwicklung solcher Instrumente für Cyberangriffe relativ gross.

Bei der Vor- und Nachbereitung von Cyberangriffen kann KI aber eine grosse Rolle spielen. Neben der Suche nach Schwachstellen kann KI sehr hilfreich sein, um Ziele zu identifizieren. Kriminelle können beispielsweise ohne grossen Aufwand nach Unternehmen suchen, bei denen sie sich hohe Erfolgschancen für einen Erpressungsversuch ausrechnen. Auch bei der Nachbereitung von Angriffen ist KI ein gut geeignetes Hilfsmittel. Wenn Kriminelle bei einem Cyberangriff eine hohe Datenmenge entwendet haben, können sie KI einsetzen, um diese Daten zu analysieren und so beispielsweise durch gezielte Erpressungen einen möglichst hohen Profit aus den Daten zu generieren.

[29] Schneier, B. (2019, 7. Januar). *Machine learning to detect software vulnerabilities*. Schneier on Security. https://www.schneier.com/blog/archives/2019/01/machine_learnin.html
[30] Zero-Day-Exploit (auch Zero-Day-Bedrohung genannt) ist ein Angriff, bei dem böswillige Akteure eine Sicherheitslücke ausnutzen, für die es noch keine Lösung gibt. Man spricht von einer „Zero-Day"-Bedrohung, weil der Entwickler oder das Unternehmen, sobald die Schwachstelle entdeckt wird, „null Tage" Zeit hat, um eine Lösung zu finden. Was ist ein Zero-Day-Exploit? | Zero-Day-Bedrohungen | Cloudflare
[31] Müller, A. (2023). Einfluss künstlicher Intelligenz auf die IT Security von Unternehmen – Chancen und Risiken. In Smarte Unternehmen, digitale Gesellschat. In: Wie KI & Co. das Leben und Arbeiten der Zukunt gestalten könn(t)en IT Security im Zeitalter der Digitalisierung (S. 40). https://d-nb.info/137132221X/34
[32] Malware, oder «schädliche Software», ist ein Überbegriff, der jedes bösartige Programm oder jeden Code beschreibt, der für Systeme schädlich ist.Feindlich, aufdringlich und absichtlich zerstörerisch - Malware versucht, Computer, Computersysteme, Netzwerke, Tablets und mobile Geräte zu infiltrieren, zu beschädigen oder zu deaktivieren, indem sie oft die Kontrolle über Teile der Geräteoperationen übernimmt. Ähnlich wie die menschliche Grippe stört sie die normale Funktion. Was ist Malware? Definition, Arten und Schutz von Malware.
[33] Niu, Z., Ren, H., Gao, X., Hua, G., & Jin, R. (2024). Jailbreaking attack against multimodal large language model. S.1 *arXiv preprint arXiv:2402.02309*.
[34] Fritsch, L., Jaber, A., & Yazidi, A. (2022, May). An overview of artificial intelligence used in malware. In *Symposium of the Norwegian AI Society* (pp. 41-51). Cham: Springer International Publishing.
[35] Pa Pa, Y. M., Tanizaki, S., Kou, T., Van Eeten, M., Yoshioka, K., & Matsumoto, T. (2023). An Attacker's Dream? Exploring the Capabilities of ChatGPT for Developing Malware. In Proceedings of CSET 2023 - 16th Cyber Security Experimentation and Test Workshop (pp. 10-18). (ACM International Conference Proceeding Series). ACM. https://doi.org/10.1145/3607505.3607513
[36] Europol, Trend Micro Research, & United Nations Interregional Crime and Justice Research Institute. (2020, 19. November). *Malicious uses and abuses of artificial intelligence*. Europol. https://www.europol.europa.eu/publications-documents/malicious-uses-and-abuses-of-artificial-intelligence unicri.org+2unov.tind.io+2

KI kann bei der Entwicklung und Ausführung von Cyberangriffen ein wichtiges Hilfsmittel sein, dessen Bedeutung weit über die Generierung von Social Engineering-Angriffen hinausgeht. KI macht die Entwicklung von Angriffsvektoren deutlich einfacher und schneller. Es ist deshalb absehbar, dass die Bedeutung von KI für Cyberangriffe weiter an Bedeutung gewinnen wird.

3.1.3 Angriffe auf KI

Nachdem die beiden vorherigen Kapitel aufgezeigt haben, wie KI für Angriffe genutzt werden kann, geht dieses Kapitel der Frage nach, ob die Nutzung von KI in Organisationen zu neuen Verwundbarkeiten führt. Diese Nutzung hat in den letzten Jahren zugenommen und es ist davon auszugehen, dass KI-Anwendungen zu einem immer wichtigeren Instrument im Arbeitsalltag in vielen Organisationen werden.[37] KI wird vermehrt in bestehende IT-Prozesse integriert. Dabei stellen sich grundsätzliche Fragen bezüglich der Informationssicherheit und des Datenschutzes. Bei der Einführung von KI-Anwendungen sollte jede Organisation prüfen, wie sie sicherstellen kann, dass keine sensiblen Informationen (z.B. Geschäftsgeheimnisse) offenbart werden und dass die Anforderungen des Datenschutzes eingehalten werden können.

Für die Cybersicherheit ist entscheidend, dass die Verwendung von KI in einer Organisation zu neuen Verwundbarkeiten gegenüber Cyberangriffen führen kann. Da viele Organisationen erst am Anfang der Integration von KI in ihre IT stehen, gibt es wenige Erkenntnisse darüber, welche Sicherheitsherausforderungen sich stellen.[38] Es lassen sich aber mindestens drei zentrale Herausforderungen feststellen: die Schwierigkeit das Verhalten von KI-Anwendungen zu überwachen, die Gefahr von manipulierter KI und die Abhängigkeit von Herstellern von KI-Anwendungen.

Beim Einsatz von KI stellt sich zunächst die Herausforderung des Sicherheitsmonitorings. Aufgrund der teilautonomen Funktionsweise von KI-Anwendungen lässt sich nur schwer überprüfen, ob die KI wirklich nur innerhalb ihres vorgesehenen Einsatzbereichs agiert. Fehlfunktionen oder gar Manipulationen der KI-Anwendungen sind entsprechend schwer erkennbar. Es sind zahlreiche Angriffsvektoren bekannt, deren Ziel es ist, ein KI-System ausser Betrieb zu setzen oder dessen Leistung einzuschränken, die Integrität eines Modells zu verletzen, um die Ergebnisse zu beeinflussen, die Vertraulichkeit zu gefährden oder Funktionen zu missbrauchen.[39]

Die Gefahr von manipulierter KI besteht vor allem durch Malware oder das Hinzufügen von Hintertüren, sogenannte «Backdoors», in den verfügbaren Modellen. Weil diese stark verbreitet sind, ist es für Cyberangreifer attraktiv Malware und Backdoors in solche Modelle einzuschleusen. Erste Fälle solcher Backdoors in KI sind bereits bekannt.[40] Eine weitere Möglichkeit der Manipulation besteht darin, das Verhalten der KI mittels gefälschter Trainingsdaten zu beeinflussen. Auch wenn diese Art der Manipulation, sogenanntes «Data Poisoning», mit viel Aufwand verbunden ist, kann sie eine Option für Angreifenden sein, da sie kaum nachgewiesen werden kann.

Schliesslich muss berücksichtigt werden, dass der Einsatz von KI auch neue Abhängigkeiten schafft. Der Markt wird derzeit von mehreren Anbietern geprägt, die in einem intensiven Wettbewerb um Marktanteile stehen. Kunden haben die Wahl zwischen unterschiedlichen Modellen, müssen jedoch unter Berücksichtigung wirtschaftlicher Interessen sowie geopolitischer Risiken sorgfältig prüfen, welcher Anbieter für welche Anwendung geeignet ist. Dies gilt unter dem Gesichtspunkt der digitalen Souveränität insbesondere auch für den Staat.

---

[37] Siehe dazu Einblick in die Nutzung von Künstlicher Intelligenz in Schweizer Unternehmen oder Columbus Consulting - Studie zum Einsatz von Daten und KI in der Schweiz 2024.pdf

[38] In einer Studie des Nationalen Testinstituts für Cybersicherheit (NTC) über die Integration von KI im Stromökosystem, haben aber Expertinnen und Experten aufgezeigt, welche Sicherheitsherausforde-rungen sich dabei stellen können. Sommer, D. M., Huber, F., Schöni, P., & Reischuk, R. M. Analyse potenzieller Cybersicherheitsrisiken beim Einsatz von KI im Zukunftsmodell des Schweizer Stromökosystems 20250225-analyse-ki-im-stromoekosystem.pdf

[39] Schweizerische Akademie der Technischen Wissenschaften (SATW). (n.d.). *(Adversarial) Artificial Intelligence*. Abgerufen am 5. Oktober 2025, von https://www.satw.ch/de/news/adversarial-artificial-intelligence

[40] Cohen, D. (2025). *Data scientists targeted by malicious Hugging Face ML models with silent backdoor*. Abgerufen am 5. Oktober 2025, von https://jfrog.com/blog/data-scientists-targeted-by-malicious-hugging-face-ml-models-with-silent-backdoor/

Bei der Analyse des Einflusses von KI auf die Cyberbedrohungslage ist es daher wichtig, nicht nur zu betrachten, wie KI für Angriffe genutzt werden kann, sondern auch, welche neuen Verwundbarkeiten durch den Einsatz von KI entstehen. Da KI-Systeme manipuliert werden können und potenzielle Hintertüren enthalten, müssen in Organisationen auch bei der Nutzung von KI konsequent die Prinzipien der Cybersicherheit beachtet werden.

3.2 Chancen durch KI

Es wurde bereits aufgezeigt, dass KI in verschiedener Hinsicht die Cyberbedrohungslage verschärft. KI kann aber auch eine Chance für die Cybersicherheit sein. Sie kann herkömmliche Cybersicherheitsmassnahmen ergänzen und neue Verteidigungsstrategien ermöglichen. Laut der Forschung und Expertenmeinung könnte generative KI künftig helfen, Schadsoftware, Zero-Day-Schwachstellen, Anomalien im Verhalten von IT-Systemen und Datenlecks frühzeitig zu erkennen. KI ersetzt dabei keine klassischen Sicherheitslösungen nicht, sondern beschleunigt die Erkennung und automatisiert Routineaufgaben. Aktuell dient sie vor allem als Unterstützung für Entwicklerinnen und Entwickler sowie für Cybersicherheitsanalysten und nicht als Ersatz.[41]

Analog zum Kapitel über die Cyberbedrohungen durch KI soll im Folgenden aufgezeigt werden, wie KI für die Cybersicherheit genutzt werden kann.

3.2.1 Verhinderung von Social Engineering

In Kapitel 3.1.1. wurde aufgezeigt, dass KI verwendet wird, um Social Engineering Angriffe durchzuführen. Umgekehrt kann aber KI auch verwendet werden, um Social Engineering zu erkennen. Bei der Erkennung von gefälschten Inhalten wurden grosse Fortschritte erzielt. Wie weit diese gehen, zeigt das Beispiel eines Tools, welches das Blinzelverhalten von Menschen auf Videos analysiert und so feststellen kann, ob dieses durch KI generiert wurde.[42] Auch für Texte und Bilder existieren Algorithmen, welche angeben können, wie hoch die Wahrscheinlichkeit ist, dass die Inhalte durch KI erstellt worden sind. Solche Anwendungen können für die Identifikation von Social Engineering genutzt werden, wenn sie mit weiteren Faktoren (z.B. Absender, Herkunft, etc.) kombiniert werden, welche heute bereits zur Entdeckung von Social Engineering Versuchen verwendet werden.

Während KI ein wesentlicher Treiber für die Zunahme der Qualität und Quantität von Social Engineering Angriffen ist, kann zugleich festgestellt werden, dass sie auch Teil der Lösung ist. KI kann sehr gut eingesetzt werden, um durch KI erzeugte Inhalte zu erkennen, welche für Cyberangriffe eingesetzt werden. Problematisch bleibt dabei, dass die Angreifenden üblicherweise einen Schritt voraus sind, indem sie neue Methoden finden, ihre Inhalte echt erscheinen zu lassen. Die Cyberabwehr muss sich dann jeweils reaktiv wieder auf die neuen Vorgehensweisen ausrichten. Studien zeigen, dass KI-gestützte Social Engineering-Angriffe deutlich erfolgreicher sind als herkömmliche Methoden, da viele Betroffene Schwierigkeiten haben, die Echtheit von Inhalten zu beurteilen. Dadurch ist der Vorteil von KI für Angreifende im Bereich Social Engineering aktuell grösser als der Nutzen für die Abwehr.[43]

Neben der Früherkennung haben KI-Anwendungen auch grosses Potential bei der Schulung von Mitarbeitenden. Bereits heute sind viele KI-basierte «Phishing Awareness Trainings» online verfügbar. Bei Bedarf können Unternehmen mit diesen Hilfsmitteln ohne viel Aufwand ihre Mitarbeitenden über die

[41] Cloudflare. (o. D.). *theNET | Erkennung von KI-Schwachstellen: ein zweischneidiges Schwert*. Cloudflare. Abgerufen am 5. Oktober 2025 von https://www.cloudflare.com/de-de/the-net/ai-vulnerabilities/
[42] Demao Xiong, Zhan Wen, Cheng Zhang, Peng Xiao, Meiqin Wu,(2025) "Deepfake Detection Based on LSTM Networks with Facial Geometric Features", *4th International Conference on Electronics, Integrated Circuits and Communication Technology (EICCT)*, pp.627-630, 2025.
[43] Resilience. (2025, 16. September). *KI-Phishing erreicht 54 Prozent Erfolgsquote*. ad-hoc-news.de. https://www.ad-hoc-news.de/boerse/news/ueberblick/ki-phishing-erreicht-54-prozent-erfolgsquote/68194315

Erkennung von Phishing Mails und andere Social Engineering Angriffen aufklären. Viele Unternehmen nutzen solche Angebote für ihre internen Schulungen.[44]

3.2.2 Entdeckung von Schwachstellen

Die Möglichkeiten von KI bei der Suche nach Schwachstellen nutzen nicht nur Cyberangreifer, sondern auch jenen, welche die Cybersicherheit verbessern wollen. Dies beginnt bereits bei der Entwicklung von neuer Software. Die Entwicklerplattform GitHub integriert beispielsweise generative KI in ihrer Code-Scan-Funktion, die Sicherheitslücken im Code aufspürt und Warnmeldungen anzeigt. Dieser Ansatz kann zwar zu falsch-positiven Ergebnissen führen, aber in Kombination mit einer manuellen Analyse die Identifizierung von Sicherheitslücken beschleunigen.[45]

KI ist auch für Sicherheitsforschende und Pentesterinnen und Pentester sehr gut geeignet, um Schwachstellen zu identifizieren. Sie nutzen dabei dieselben Möglichkeiten wie Angreifende und profitieren genauso wie diese davon, dass KI die Suche und Analyse von Schwachstellen massiv vereinfacht. Nach wie vor ist aber ein fundiertes Fachwissen nötig, damit die KI richtig eingesetzt und die gefundenen Schwachstellen auf ihre Relevanz für die Cybersicherheit beurteilt werden können. KI ist deshalb ein Instrument zur Unterstützung von Fachleuten, wird diese aber in absehbarer Zeit nicht ersetzen können.

KI kann die Sicherheitsforschenden zudem bei der Bewertung und der Kategorisierung von Schwachstellen unterstützen. So kann KI eingesetzt werden, um den Aufwand für die nötige Dokumentation und Bewertung von Schwachstellen zu reduzieren. Auch im Bundesamt für Cybersicherheit (BACS) wird KI für diese Zwecke bereits eingesetzt.

Es lässt sich schlussfolgern, dass KI bei der Entdeckung von Schwachstellen sowohl für Angreifer als auch für Sicherheitsforschende ein wichtiges Hilfsmittel sein kann. Da die Suche nach Schwachstellen hochgradig automatisiert durchgeführt wird, führt der Einsatz von KI nicht zu einer grundlegend neuen Ausgangslage. Die Suche nach bekannten bestehenden Schwachstellen kann aber effizienter durchgeführt und der Aufwand für die Dokumentation reduziert werden.

3.2.3 Früherkennung und Abwehr von Cyberangriffen

Das Potential von KI bei der Früherkennung und Abwehr von Cyberangriffen ist gross. Wenn KI auf Informationen über aktuelle Bedrohungen Zugriff hat, kann sie eingesetzt werden, um Abwehrsysteme automatisch an neue Angriffsmuster anzupassen. Verhaltensanalysen und Anomalieerkennung ermöglichen es KI-Systemen zudem, den Netzwerkverkehr kontinuierlich zu überwachen und verdächtige Aktivitäten in Echtzeit zu identifizieren.
KI kann auch bei der Triage von Warnmeldungen eine entscheidende Rolle spielen, indem sie hilft, aus einer Vielzahl von Alarmen diejenigen mit der höchsten Priorität herauszufiltern. Bei grossen Unternehmen oder zentralen «Security Operations Centers (SOCs)» stellt diese Priorisierung oft eine Herausforderung dar. KI kann Alarme automatisch klassifizieren und mit einem Risikowert-Score versehen, der ihre Kritikalität bewertet und priorisiert. Dadurch können Analystinnen und Analysten sich gezielt auf priorisierte Bedrohungen konzentrieren, während weniger gravierende oder eindeutig falsch-positive Meldungen nachrangig behandelt werden. KI kann Analystinnen und Analysten bei monotonen Aufgaben wie der Alarm-Triage, der Routine-Reaktion[46] und der Berichterstattung deutlich entlasten. Dadurch lässt sich das Risiko von «Alert Fatigue» reduzieren, einem Phänomen, bei dem wichtige Signale übersehen werden, weil zu viele irrelevante Alarme verarbeitet werden müssen.[47] Ähnlich wie bei der

[44] Swiss Cyber Experts. (2025). *Fragenkatalog: Künstliche Intelligenz (KI) in der Abwehr und Durchführung von Cyberangriffen – Technische Aspekte und Trends*.
[45] GitHub. (o. D.). *About code scanning*. In GitHub Docs. Abgerufen am 04.10.25, von https://docs.github.com/en/code-security/code-scanning/introduction-to-code-scanning/about-code-scanning
[46] Routine Reaktion: DataGuard (o. J.). *What is incident handling checklist?* https://www.dataguard.com/blog/what-is-incident-handling-checklist/
[47] Hossmann, T. (2025). *Sinnvoller Einsatz von KI in der Cybersecurity*. IT-Matchmaker News. Abgerufen am 5. Oktober 2025, von https://news.it-matchmaker.com/wie-kuenstliche-intelligenz-sinnvoll-einzusetzen-ist%E2%80%8B/

Schwachstellenerkennung kann KI zudem auch bei der Früherkennung und Frühwarnung sehr gut für die Dokumentation entdeckter Vorfälle eingesetzt werden, was zu einer weiteren Entlastung der Analyseteams beiträgt.

Mit diesen Anwendungen ist das Potential von KI bei der Früherkennung und Abwehr nichterschöpft. SOAR-Plattformen (Security Orchestration, Automation and Response) werden zunehmend verwendet, um mittels vordefinierter Reaktionspläne für bekannte Bedrohungen automatisierte Prozesse zur Vorfallbewältigung auszulösen.[48] Der Grad der Automatisierung kann dabei von den verantwortlichen Teams selbst gewählt werden. So kann KI gezielt dort eingesetzt werden, wo dies für Analyseteams geeignet ist und wo sie sich in ausreichendem Mass auf die KI verlassen können.
Wie bei anderen Anwendungen liegen die grossen Vorteile von KI-Lösungen in der hohen Skalierbarkeit, der Fähigkeit hohe Datenmengen zu verarbeiten und in der Verfügbarkeit. Bei Sicherheitsteams, welche Tag und Nacht zur Verfügung stehen müssen, können Hilfsmittel, welche ohne Pause verfügbar sind und keine Limiten bezüglich der zu überwachenden Systeme haben, sehr wertvoll sein. In einer Umfrage unter den Mitgliedern der «Swiss Cyber Experts» geben nahezu alle befragten Unternehmen an, KI in der Cyberabwehr in einer unterstützenden Rolle zu verwenden.[49]

Durch die Umsetzung solcher Lösungen kann dem Mangel an Ressourcen im Bereich der Cybersicherheit teilweise entgegengewirkt werden. Zu beachten ist aber auch im Bereich der Früherkennung und Abwehr, dass KI vor allem unterstützend wirken kann. Es wird weiterhin erfahrenes menschliches Fachpersonal bei der Vorfallbewältigung brauchen, um sicherzustellen, dass Cyberangriffe rechtzeitig mit Unterstützung aller verfügbaren technischen Hilfsmittel erkannt und abgewehrt werden. Beim Einsatz autonomer oder teilautonomer KI-Systeme sollte die Verantwortung stets bei den Mitarbeitenden verbleiben. Daher ist sicherzustellen, dass Entscheidungen weiterhin von ihnen getroffen werden.

3.3 Analyse

Die Analyse der Auswirkung von KI auf Cyberbedrohungen und auf die Mittel der Cybersicherheit zeigt, dass relevante Entwicklungen bereits stattgefunden haben oder zu erwarten sind. KI-Anwendungen werden insbesondere im Bereich des Social Engineering weiterhin spürbare Veränderungen bringen. Es muss davon ausgegangen werden, dass die Qualität und Quantität dieser Angriffe weiter zunehmen werden. Zwar wurde gezeigt, dass KI auch für die Erkennung solcher Angriffe verwendet werden kann, im Wettstreit zwischen Angriff und Abwehr bestimmen aber häufig die Angreifende den Fortschritt der Entwicklung. Es ist daher fraglich, ob die verfügbaren Schutzmittel immer rechtzeitig bereitstehen und von den Anwendern auch tatsächlich genutzt werden.
Bei der Analyse von Schwachstellen und bei der Entwicklung respektive der Früherkennung und Abwehr von Cyberangriffen wird KI bereits sowohl von Angreifern wie auch von Sicherheitsteams eingesetzt. Die Entwicklung steht noch am Anfang und die tatsächlichen Auswirkungen lassen sich derzeit nicht abschliessend beurteilen. Klar ist, dass KI die Schwelle für Cyberangriffe weiter senkt, so dass solche mit weniger vertieften Kenntnissen und in kürzerer Zeit entwickelt werden können. Umgekehrt bietet KI viele Möglichkeiten, um die heute oft stark beanspruchten Sicherheitsteams zu entlasten, da KI Aufgaben wie die Früherkennung von Bedrohungen oder die Erkennung von Schwachstellen deutlich beschleunigen kann.

Obwohl diese Entwicklungen für die Cybersicherheit von grosser Bedeutung sind, muss auch klar festgehalten werden, dass KI die Cybersicherheit bisher nicht grundlegend verändert hat. Bereits bekannte Angriffs- und Abwehrmuster werden durch KI einfacher und breiter umsetzbar. Es haben sich durch den Einsatz von KI aber keine völlig neuen «Tactics, Techniques and Procedures» (TTP)[50] im Bereich der

[48] Usman, U. (2025, 14. August). *Ganzheitlicher Schutz vor KI-gesteuerten Angriffen*. Nomios Deutschland. https://www.nomios.de/nachrichten/defending-against-machine-led-attacks/
[49] Swiss Cyber Experts. (2025). *Fragenkatalog: Künstliche Intelligenz (KI) in der Abwehr und Durchführung von Cyberangriffen – Technische Aspekte und Trends*.
[50] Tactics, Techniques, and Procedures beschreiben das Verhalten eines Bedrohungsakteurs. Um sich vor Angriffen zu schützen und Cyberrisiken zu minimieren, ist es wichtig, die von den Angreifern eingesetzten TTP zu kennen und zu verstehen. Vgl: MITRE Corporation. (2023). *Tactics, Techniques, and Procedures (TTPs)*. MITRE ATT&CK®. https://attack.mitre.org/resources/faq/

Cyberbedrohungen herausgebildet. Wichtig ist auch festzuhalten, dass in naher Zukunft nicht zu erwarten ist, dass vollständig automatisierte Cyberangriffe möglich sind. KI kann zwar Teilprozesse bei der Vorbereitung und Ausführung von Cyberangriffen automatisieren, letztendlich ist aber menschliches Handeln und Entscheiden für solche Angriffe nach wie vor notwendig. Erst wenn KI tatsächlich fähig sein sollte, selbständig neue Angriffsvektoren zu entwickeln und einzusetzen, müsste von einer KI-Revolution in der Cybersicherheit gesprochen werden. Auf dem Stand des heutigen Wissens ist eine solche Entwicklung in den nächsten Jahren 3 – 5 Jahren nicht zu erwarten.[51]

4 KI in der Nationalen Cyberstrategie

Die Schweiz verfügt seit 2012 über eine Nationale Cyberstrategie (NCS). Diese wurde zuletzt im Jahr 2023 aktualisiert[52]. Das Postulat 23.3861 enthält explizit den Auftrag zu prüfen, ob die Strategie den durch KI hervorgerufenen Entwicklungen gerecht wird. Diesem Prüfauftrag wird in diesem Kapitel nachgegangen. Zunächst wird analysiert, wie die NCS KI thematisch abhandelt. Anschliessend wird für jedes strategische Ziel der NCS aufgezeigt, welche KI-bezogenen Vorhaben umgesetzt werden, wo dabei die Schwerpunkte liegen und wo allfälliger Handlungsbedarf besteht.

4.1 Wie die NCS das Thema KI adressiert

Die NCS legt auf strategischer Ebene fest, mit welchen Massnahmen die Schweiz welchen Cyberbedrohungen entgegenwirken will. Dabei ist die Strategie umfassend angelegt. Es sollen grundsätzlich alle relevanten Cyberbedrohungen berücksichtigt werden. Die NCS, die 2023 vom Bundesrat verabschiedet wurde, ist die erste Ausführung der Strategie mit offenem Zeithorizont. Sieverfolgt das Ziel, die Cybersicherheit durch langfristig ausgerichtete Massnahmen nachhaltig zu stärken. Damit dies gelingen kann, muss die NCS dynamische Entwicklungen berücksichtigen. Dazu gehören geopolitische und technologische Veränderungen, welche wiederum durch verschiedene Faktoren vorangetrieben werden und sich unterschiedlich auf die Cyberbedrohungslage auswirken.

KI gehört unbestritten zu den technologischen Entwicklungen, welche einen Einfluss auf Cyberbedrohungen haben. Die vorgängigen Kapitel haben diese Auswirkungen aufgezeigt. Die Strategie nennt KI entsprechend als einen zentralen technologischen Einflussfaktor auf die Cyberbedrohungslage. Beschrieben wird der Einfluss wie folgt: «*Durch teilweises oder gänzlich autonomes maschinelles Lernen sind Anwendungen der KI fähig, sehr komplexe Analysen in kurzer Zeit durchzuführen. Diese Möglichkeiten können genutzt werden, um Systeme besser zu schützen, umgekehrt können sie aber auch dazu verwendet werden, Angriffe effektiver und effizienter durchzuführen. Weil sich viele Organisationen bei Entscheidungen zunehmend auf die Analysen von KI-Anwendungen verlassen, sind auch Angriffe auf solche Anwendungen ein relevantes Bedrohungsszenario. Zudem können KI-Anwendungen auch ohne Fremdeinwirkung ein Sicherheitsrisiko darstellen, wenn eine fehlerhafte Anwendung eine Störung oder ein Datenleck verursacht*».[53]

In der NCS sind die in diesem Bericht analysierten Auswirkungen der KI auf die Cybersicherheit grundsätzlich beschrieben. Die NCS verfolgt aber bei der Beschreibung der strategischen Ziele und der Herleitung von Massnahmen einen konsequent technologieneutralen Einsatz. Sie versteht sich als konzeptioneller Rahmen, der offen und flexibel ist, um neue technologische Entwicklungen dynamisch zu integrieren. Entsprechend adressieren weder die strategischen Ziele noch die Massnahmen direkt die

[51] Axios. (2025, Januar 7). *Goldilock warns of AI-powered "agentic malware" by 2027*. https://www.axios.com/2025/01/07/goldilock-agentic-malware-2027-doomsday

[52] Der Bundesrat und die Kantone legen die neue Nationale Cyberstrategie fest. https://www.ncsc.admin.ch/dam/ncsc/de/dokumente/strategie/cyberstrategie-ncs/Nationale-Cyberstrategie-NCS-2023-04-13-DE.pdf.download.pdf/Nationale-Cyberstrategie-NCS-2023-04-13-DE.pdf

[53] NCS, S. 7.

Chancen und Risiken von KI-Systemen in der Cybersicherheit

Frage nach den Auswirkungen von KI auf die Cybersicherheit. Solche Fragen sollen vielmehr in der Umsetzung der Massnahmen berücksichtigt werden.

Um zu prüfen, ob dieser Ansatz angesichts der Entwicklung der KI angemessen ist, wird im Folgenden aufgezeigt, welche KI-bezogenen Vorhaben im Rahmen der NCS bereits umgesetzt und ob Lücken festgestellt werden. Dafür werden die Arbeiten entlang der fünf strategischen Ziele pro Unterkapitel untersucht.

4.1.1 Ziel 1: Selbstbefähigung

Die Schweiz baut ihre Stellung als einer der weltweit führenden Wissens-, Bildungs- und Innovationsstandorte auch in der Cybersicherheit aus. Sie nutzt diese Fähigkeiten, um Cyberrisiken über die Lieferketten eigenständig zu beurteilen, technologische Entwicklungen zu antizipieren und agil darauf zu reagieren. Die Bevölkerung ist über Cyberrisiken informiert und gewinnt dadurch Vertrauen in die Nutzung digitaler Dienstleistungen.	
Bezug zu KI	Für den Bildungs-, Forschungs- und Innovationsstandort Schweiz ist es von zentraler Bedeutung, dass Kompetenzen im Umgang mit KI aufgebaut werden. Dies schliesst auch die Beurteilung von Cyberbedrohungen im Zusammenhang mit der Verwendung von KI ein. Zudem ist die Sensibilisierung zur Auswirkung von KI auf die Cybersicherheit wichtig. Wie gezeigt wurde, ist ein starker Einfluss von KI auf die Bedrohung durch Social Engineering feststellbar. Im Rahmen der Umsetzung der NCS muss daher ein Schwerpunkt auf die Information aller Betroffenen (Bevölkerung, Unternehmen, Behörden) zu dieser Bedrohung gesetzt werden.

Stand der Umsetzung (Schwerpunkte)	**Bildung, Forschung und Innovation** Die Schweiz ist im Bereich der Bildung, Forschung und Innovation im internationalen Vergleich sehr gut aufgestellt, insbesondere auch im digitalen Bereich.[54] Schweizer Innovationen setzen stark auf Qualität, Transparenz und Vertrauenswürdigkeit.[55] Im Rahmen der NCS engagieren sich verschiedene Akteure dafür, die vorhandenen Kompetenzen zu nutzen, um Aspekte der Cybersicherheit vertieft zu analysieren und um Cybersicherheit als Bestandteil der Innovation bei KI zu integrieren. Begleitend zur Veröffentlichung der Strategie 2023 hat die Schweizerische Akademie der Technischen Wissenschaften (SATW) beispielsweise in einer Arbeitsgruppe eine Analyse zu Cyberbedrohungen in und durch KI veröffentlicht.[56] Der Cyber-Defence Campus (CYD Campus) der armasuisse W+T befasst sich ebenfalls intensiv mit dem Thema KI und Cybersicherheit. In Zusammenarbeit mit der Hochschule für Technik und Wirtschaft Wallis hat er eine umfangreiche Studie zu den Bedrohungen und Auswirkungen generativer KI veröffentlicht.[57] In einem aktuellen Projekt mit dem AI & Digital Economy Lab der Universität Lausanne wird zudem untersucht, welche Schutzmassnahmen Unternehmen gegenüber KI-basierten Cyberbedrohungen umsetzen. Mit Fellowship- und Praktikumsprogrammen des Campus werden gezielt Expertinnen und Experten im Bereich KI und Cybersicherheit ausgebildet, beispielsweise mit Arbeiten zu datenschutzfreundlichen Sprachmodellen, an der Absicherung maschineller Lernsysteme und an der Entwicklung robuster und interpretierbarer KI-Methoden.[58] **Bedrohungslage und Sensibilisierung** Das BACS informiert kontinuierlich über aktuelle Trends und neue Bedrohungen im Bereich der Cybersicherheit. Eine Grundlage dafür sind freiwillige Meldungen. Auf dieser Basis hat das BACS bereits über den zunehmenden Missbrauch von KI für Betrugsversuche mit öffentlichen Persönlichkeiten berichtet. Darüber hinaus zeigt das BACS auf, wie solche Bedrohungen frühzeitig erkannt werden können, und gibt konkrete Empfehlungen zum Schutz und zum richtigen Umgang mit solchen Vorfällen.[59] Ergänzend zu diesen Warnungen engagiert sich das BACS gemeinsam mit Partnerorganisationen bei Sensibilisierungskampagnen. Der Einsatz von KI bei Cyberangriffen stand im Fokus der Kampagne des BACS zum «European Cyber Security Month». Gemeinsam mit den Partnerorganisationen «Jugend und Medien» sowie «Seniorweb.ch» wurden zielgruppenspezifische Inhalte für Jugendliche, Berufstätige und Seniorinnen und Senioren entwickelt. In einem Kurzvideo wurde veranschaulicht, wie einfach es für Cyberkriminelle ist, mithilfe von KI ein täuschend echtes Deep-Fake-Video zu erstellen.[60]
Handlungsbedarf	Sowohl im Bereich Forschung, Bildung und Innovation als auch bei der Sensibilisierung der Öffentlichkeit bestehen viele Initiativen und Projekte. Insbesondere bei der Forschung sind aber viele Projekte nicht oder nur teilweise in die Umsetzung der NCS integriert. Damit Cybersicherheit bei KI strategisch angegangen werden kann, sollten die Projekte stärker über die Massnahmen 1 und 2 der NCS koordiniert werden.

[54] Die Schweiz belegt 2024 im “World Digital Competitiveness Ranking» des IMD Platz 2. World Digital Competitiveness Ranking 2024 - IMD business school for management and leadership courses
[55] Beispiele dafür sind das Start-up AlpineAI, welches vertrauenswürdige KI-Anwendungen entwickelt (AlpineAI AG. (n.d.). *SwissGPT – KI-Anwendungen für kritische Sektoren*. Abgerufen am 5. Oktober 2025, von https://alpineai.swiss/) und die von der ETHZ und EPFL entwickelte offene KI «Apertus» (Apertus: Ein vollständig offenes, transparentes und mehrsprachiges Sprachmodell | ETH Zürich)
[56] Schweizerische Akademie der Technischen Wissenschaften (SATW). (2020, 21. Oktober). *Cybersecurity Map*. https://www.satw.ch/de/cybersecurity/cybersecurity-map
[57] Studienergebnisse zu Bedrohungen und Auswirkungen von generativer künstlicher Intelligenz auf die Cybersicherheit
[58] Meier (2025) Threats and Opportunities in AI-generated Images for Armed Forces.pdf, CYD armasuisse
[59] Siehe: Woche 35: Betrug mit bekannten Persönlichkeiten – Die Schattenseiten der künstlichen Intelligenz
[60] Einsatz von KI bei Cyberangriffen – Fokusthema des BACS im «European Cyber Security Month»: vbs.admin.ch/de/nsb?id=102634

4.1.2 Ziel 2: Sichere und verfügbare digitale Dienstleistungen und Infrastruktur

Die Schweiz setzt flächendeckend Massnahmen zur Stärkung der Cyberresilienz um. Bund und Kantone schaffen die nötigen Rahmenbedingungen dafür, dass ein hohes Schutzniveau gewährleistet ist, sichere digitale Infrastrukturen, Produkte und Dienstleistungen eingesetzt werden und die Risikobereitschaft bewusst gesteuert wird.	
Bezug zu KI	KI hat einen wichtigen Einfluss darauf, wie Schwachstellen erkannt und behoben werden. Sie wird aber auch von Angreifern genutzt, um Schwachstellen zu finden und auszunutzen. Dieser Aspekt der KI ist relevant für das Ziel von sicheren digitalen Dienstleistungen und Infrastrukturen. Weiter ist von Bedeutung, dass KI-bezogene Bedrohungen über Standards adressiert werden, welche in einem koordinierten Ansatz von Kantonen und Bund als Rahmenbedingungen für den Einsatz von KI definiert werden.
Stand der Umsetzung (Schwerpunkte)	**Regulierung und Standardisierung** Die Frage, ob Cybersicherheit bei KI reguliert werden soll und welche internationalen Standards definiert sind, muss im Gesamtkontext der Frage nach Regulierung von KI betrachtet werden. Dazu hat eine interdepartementale Arbeitsgruppe des Bundes eine Analyse publiziert, welche den Handlungsbedarf und mögliche Regulierungsansätze auflistet. Die Studie listet auch auf, welche Standards bereits heute verfügbar sind.[61] **Erkennung und Behebung von Schwachstellen** In Kapitel 3 wurde dargelegt, dass KI eine grosse Relevanz bei der Suche, der Schliessung aber auch bei der Ausnutzung von Schwachstellen hat. Es findet ein regelmässiger Austausch über den Einsatz von KI zwischen Expertinnen und Experten im Schwachstellenmanagement statt, die direkt an der Umsetzung der NCS beteiligt sind. Daran beteiligt sind unter anderem das BACS, das NTC, der CYD Campus, Hochschulen wie die ETH Zürich sowie IT-Unternehmen. **Koordination zwischen Bund, Kantone und Privatsektor** Für die Nutzung von KI durch Behörden bestehen Koordinations- und Fachgremien. Das Kompetenznetzwerks für künstliche Intelligenz (CNAI)[62] soll den Einsatz von und das Vertrauen in KI und andere neue Technologien innerhalb der Bundesverwaltung und darüber hinaus rasch und nachhaltig fördern. Das Informationssicherheitsgesetz (ISG) legt den Rahmen und die Verfahren für die Sicherheitsmassnahmen im Bund fest und hat für die Kantone dann Gültigkeit, wenn diese auf die Informatikmittel des Bundes zugreifen oder klassifizierte Informationen des Bundes bearbeiten. Das ISG gilt auch für Fragen der Informations- respektive Cybersicherheit bei der Nutzung von KI.
Handlungsbedarf	In Bezug auf die Standardisierung und Regulierung, ebenso wie bei der Koordination zwischen Bund, Kantonen und Privatsektor, setzt die Umsetzung der NCS auf bestehende Prozesse und Gremien, um das Thema Cybersicherheit in der KI anzugehen. Dadurch ist der Bezug zu übergeordneten Fragen (z.B. generelle Regulierung von KI, generelle Koordination in der Cybersicherheit) gewährleistet. Es fehlt jedoch an einer Übersicht zum Stand der Arbeiten. Dies trifft auch auf den Status quo beim Einsatz von KI im Schwachstellenmanagement zu. Handlungsbedarf besteht demnach nicht bei der Schaffung von neuen Prozessen und Initiativen, sondern bei der Erfassung der bestehenden Arbeiten mit Bezug auf die Cybersicherheit bei KI. Dieser Austausch sollte über die Massnahmen 5, 6 und 7 der NCS gestärkt werden.

[61] Eidgenössisches Justiz- und Polizeidepartement (EJPD) – Bundesamt für Justiz (BJ). (2024, 31. August). *Rechtliche Basisanalyse im Rahmen der Auslegeordnung zu den Regulierungsansätzen im Bereich künstliche Intelligenz* (BJ-D-F0D63401/159). Bern: Bundesamt für Justiz. Retrieved from https://www.bakom.admin.ch/dam/bakom/en/dokumente/KI/analyse_juritisch.pdf4 Juristischen Basisanalyse_DE

[62] Schweizerische Eidgenossenschaft. (n.d.). *Startseite – CNAI – Kompetenznetzwerk für künstliche Intelligenz*. Abgerufen am 5. Oktober 2025, von https://cnai.swiss/

4.1.3 Ziel 3: Wirksame Erkennung, Verhinderung, Bewältigung und Abwehr von Cyberangriffen

Die Schweiz verfügt in allen Lagen über die nötigen Kapazitäten und Organisationsstrukturen, um Cyberbedrohungen und -vorfälle rasch zu erkennen und deren Schäden zu minimieren. Vorfälle werden auch dann bewältigt, wenn sie über längere Zeit andauern und verschiedene Bereiche gleichzeitig betreffen.	
Bezug zu KI	KI kann zur Früherkennung beitragen, erleichtert andererseits aber auch den Angreifern die Entwicklung neuer Angriffsvektoren. Mit geeigneten Massnahmen soll die Bedrohungslage besser eingeschätzt und robuste Reaktionsmechanismen für Cybervorfälle etabliert werden. Die systematische Analyse solcher Vorfälle trägt zu einer kontinuierlichen Verbesserung der Sicherheitsarchitektur bei.
Stand der Umsetzung (Schwerpunkte)	**Früherkennung und Vorfallmanagement** KI wird von vielen Sicherheitsteams zur Früherkennung und als Hilfsmittel zur Bewältigung von Cybervorfällen eingesetzt. Bei vielen dieser Anwendungen kann noch nicht abschliessend beurteilt werden, welches Potential sie zur effektiven Entlastung von Fachleuten erzielen können. Wichtig ist daher der Austausch zwischen Sicherheitsteams, welche solche Lösungen nutzen. In der Schweiz findet ein solcher Austausch an Fachkonferenzen[63] statt. Zudem forschen verschiedene Hochschulen in Zusammenarbeit mit dem CYD Campus und internationalen Partnern zum Potential und den Grenzen von KI in der Früherkennung und beim Vorfallmanagement.[64] **Cyberdefence** Im Bereich der Cyberdefence legt die «Gesamtkonzeption Cyber (GK Cyber)» [65] fest, welche Fähigkeiten die Schweizer Armee bis Mitte der 2030er-Jahre aufbauen muss, um Bedrohungen im Cyber- und elektromagnetischen Raum wirksam begegnen zu können. Dabei wird der zunehmende Einsatz von KI als strategisch wichtig hervorgehoben. Insbesondere betont die GK Cyber das Potenzial von KI, die Geschwindigkeit und Effizienz bestimmter Operationen wesentlich zu erhöhen, etwa im Bereich der Lagebeurteilung, automatisierten Reaktion oder Entscheidungsunterstützung in hochdynamischen Bedrohungsszenarien.
Handlungsbedarf	Weil bei vielen Anwendungen von KI für die Früherkennung, das Vorfallmanagement und die Cyberdefence nicht abschliessend beurteilt werden kann, welche Möglichkeiten und Grenzen bestehen, ist der Wissensaustausch zu solchen Fragen essenziell. Dieser findet in der Schweiz statt, ist jedoch meist informell organisiert und wird nicht systematisch aufgearbeitet. Massnahme 12 der NCS sollte durch formelle Austauschformate sowie standardisierte Dokumentationsprozesse ergänzt werden, um den Erfahrungsaustausch zur Anwendung von KI in der Cyberabwehr zu stärken und bestehende informelle Dialoge systematisch zu bündeln und nachhaltig nutzbar zu machen. Das Postulat 24.4265 der sicherheitspolitischen Kommission des Ständerates «KI-Strategie für die Sicherheit und Verteidigung der Schweiz» wurde am 12. Dezember 2024 an den Bundesrat überwiesen. Im Bericht zum Postulat wird der Bundesrat aufzeigen, wie KI im Bereich Cyberdefence angewendet wird.

[63] Bsp.: Swiss Cyber Security Days 2025, Global Cyber Conference 2025
[64] Eidgenössisches Departement für Verteidigung, Bevölkerungsschutz und Sport. (2021, 10. Dezember). *Die Zusammenarbeit des CYD Campus mit den Hochschulen, der Wirtschaft und den Kantonen : Bericht des VBS an den Bundesrat*. Bern. https://www.newsd.admin.ch/newsd/message/attachments/69525.pdf
[65] Gesamtkonzeption Cyber Abgerufen am 5. Oktober 2025

4.1.4 Ziel 4: Effektive Bekämpfung und Strafverfolgung der Cyberkriminalität

Die Schweiz baut ihre Fähigkeiten aus, Verursacher von Cyberangriffen zu identifizieren, strafrechtlich im Verbund zu verfolgen und im Rahmen der gesetzlichen Möglichkeiten zu verurteilen.	
Bezug zu KI	Cyberkriminalität entwickelt sich rasant, kennt keine geografischen Grenzen und nutzt digitale Infrastrukturen, um mit minimalem Aufwand maximalen Schaden zu verursachen. KI ist für Cyberkriminelle ein sehr nützliches Hilfsmittel, da sie Cyberangriffe mit weniger Fachwissen und in kürzerer Zeit entwickeln können. Gleichzeitig bietet KI auch Potenzial für die Strafverfolgung: Sie kann zur Entwicklung von Präventionsmassnahmen, zur Aufdeckung von Straftaten und zur Unterstützung bei deren Aufklärung eingesetzt werden.
Stand der Umsetzung (Schwerpunkte)	**KI in der digitalen Forensik** Bei digitalen Straftaten sind Ermittler oft vor der Herausforderung, sehr grosse Datenmengen und komplexe Metadaten auszuwerten. Die Analyse dieser Informationen ist zeit- und ressourcenintensiv und fordert hohe technische Kompetenzen. KI kann bei diesen Arbeiten ein nützliches Hilfsmittel sein, etwa zur automatisierten Mustererkennung oder zur Priorisierung relevanter Spuren. Damit solche Instrumente wirksam und verantwortungsvoll eingesetzt werden können, ist eine gezielte Aus- und Weiterbildung der Ermittlungsfachleute entscheidend. **Klärung von Möglichkeiten für «Predictive Policing»** Die Verwendung von KI ist sinnvoll, wenn aus grossen Fallzahlen rasch Schlüsse gezogen werden sollen. Solche Ansätze bestehen in verschiedenen Bereichen der Strafverfolgung und werden als «Predictive Policing» bezeichnet. Da in der Cyberkriminalität hohe Fallzahlen bestehen, können solche Methoden neue Erkenntnisse bringen. Erste Erkenntnisse können durch KI-basierte Auswertungen von gemeldeten Cybervorfällen gewonnen werden. Voraussetzung dafür ist, dass dabei die datenschutzrechtlichen Anforderungen eingehalten werden. Das BACS arbeitet daran, die zur Verfügung stehenden Informationen systematischer auszuwerten und testet dabei auch, welche KI-Anwendungen hilfreich sein können.
Handlungsbedarf	KI kann ein nützliches Hilfsmittel bei der Strafverfolgung von Cyberkriminalität sein. Es fehlen heute aber noch Hilfsmittel und Ausbildungen, welche es den zuständigen Behörden erleichtern KI einzusetzen. Die Massnahmen 12 und 14 der NCS sollten deshalb mit folgenden Projekten ergänzt werden: Analyse der rechtlichen Rahmenbedingungen für den Einsatz von KI im Bereich der Strafverfolgung, Entwicklung einheitlicher Schulungsprogramme für Ermittlungsfachleute, die sowohl technische Anwendungen als auch ethische und rechtliche Aspekte abdecken; die Bereitstellung leistungsfähiger, sicherer Analyseplattformen für grosse Datenmengen; sowie die Definition klarer Leitlinien zu Datenschutz, Nachvollziehbarkeit und verantwortungsvollem Einsatz von KI - dies soll in Abstimmung mit den Gesetzgebungsarbeiten des Bundesamtes für Justiz im Rahmen der Ratifizierung des Übereinkommens des Europarates über KI stattfinden. Die Projekte werden durch das BACS in enger Zusammenarbeit mit den betroffenen Stellen begleitet.

4.1.5 Ziel 5: Führende Rolle in der internationalen Zusammenarbeit

Die Schweiz setzt sich auf operativer und strategischer Ebene für einen offenen, freien und sicheren Cyberraum und für die umfassende Anerkennung, Einhaltung und Durchsetzung des Völkerrechts im digitalen Raum ein. Das internationale Genf wird als führender Standort für Debatten zur Cybersicherheit genutzt. Die Schweiz kann bei Differenzen mit Bezug zu Cyberoperationen als Vermittlerin auftreten.	
Bezug zu KI	Auf internationaler Ebene wird die rasante Entwicklung der KI und ihr Einfluss auf die Cybersicherheit zunehmend erkannt.[66] Zugleich hat sich die Cybersicherheit angesichts der wachsenden Cyberbedrohungslage und geopolitischen Spannungen zu einem aussen- und sicherheitspolitischen Schlüsselthema entwickelt. Die diplomatische Präsenz der Schweiz soll aktiv genutzt werden, um multilaterale Cyberdiskussionen mitzugestalten und Rahmenbedingungen für den verantwortungsvollen Umgang mit neuen Technologien zu fördern.
Stand der Umsetzung (Schwerpunkte)	**KI-Konvention des Europarats** Auf internationaler Ebene wurden als Reaktion auf die rasanten Entwicklungen und Herausforderungen KI-Regelwerke wie die KI-Konvention des Europarats entwickelt. Die Schweiz hatte den Vorsitz im zuständigen Verhandlungsgremium (CAI – Committee on Artificial Intelligence) und konnte dadurch massgeblich zu den Verhandlungen beitragen.[67] Die Schweiz hat die Konvention unterzeichnet. Eine Vernehmlassung zur Umsetzung der Konvention soll bis Ende 2026 vorgeschlagen werden. **KI-Gipfeltreffen in der Schweiz** Der Bundesrat hat sich an seiner Sitzung vom 25. Juni 2025 dafür ausgesprochen, dass 2027 in Genf ein Gipfeltreffen zur künstlichen Intelligenz (KI) durchgeführt werden soll. Er hat das Eidgenössische Departement für Umwelt, Verkehr, Energie und Kommunikation (UVEK) und das Eidgenössische Departement für auswärtige Angelegenheiten (EDA) damit beauftragt, die nötigen Vorbereitungsarbeiten an die Hand zu nehmen.[68] Diese internationale Konferenz, bringt Staaten, Unternehmen, Forschende und Zivilgesellschaft zusammen, um konkrete Massnahmen zur Förderung von KI im Dienst des Gemeinwohls, zur Innovation und für vertrauenswürdige KI-Governance zu vereinbaren. **Verantwortungsvolle KI-Systeme** Das EDA hat gemeinsam mit den Eidgenössischen Technischen Hochschulen in der Schweiz und internationalen Partnern die Initiative ICAIN (International Computation on AI Network) lanciert. Ziel dieser Initiative ist die Förderung transparenter, sicherer und verantwortungsvoller KI-Systeme. Im Zentrum steht der Aufbau allgemein zugänglicher KI-Infrastrukturen, welche die internationale Zusammenarbeit und ethische Standards stärken. Die Schweiz positioniert sich damit als aktive Mitgestalterin einer wertebasierten globalen KI-Governance.[69]
Handlungsbedarf	Mit der geplanten Ausrichtung eines KI-Gipfels 2027 in Genf bietet sich die Gelegenheit, die Rolle der Schweiz als neutrale Akteurin in der globalen digitalen Gouvernanz zu stärken. Die Schweiz hat bereits ein gut etabliertes Netzwerk an internationalen Kontakten im Bereich Cybersicherheit. Diese können verstärkt auch für den Themenbereich KI genutzt werden. Generell gilt es, die internationalen Kontakte und Prozesse in der Cybersicherheit und im Bereich KI gut aufeinander abzustimmen, so dass Synergien genutzt werden können.

[66] Arctic Wolf 2025 Trends Report Reveals AI is Now the Leading Cybersecurity Concern for Security and IT Leaders - Arctic Wolf
Arctic Wolf. (2025). *2025 trends report reveals AI is now the leading cybersecurity concern for security and IT leaders*. Abgerufen am 5. Oktober 2025, von https://www.arcticwolf.com/resources/2025-trends-report-ai-cybersecurity/
[67] "Botschaft zur Umsetzung der Strategie Künstliche Intelligenz des Bundes" Schweizerische Eidgenossenschaft. (2025). *Botschaft zur Umsetzung der Strategie Künstliche Intelligenz des Bundes*. Abgerufen am 5. Oktober 2025, von news.admin.ch/de/nsb?id=104646
[68] Schweizerische Eidgenossenschaft. (2025, 25. Juni). *Die Schweiz möchte einen Gipfel zur künstlichen Intelligenz im Jahr 2027 durchführen*. Eidgenössisches Departement für auswärtige Angelegenheiten (EDA). Abgerufen am 5. Oktober 2025, von https://www.eda.admin.ch/countries/ireland/de/home/aktuell/news.html/content/eda/de/meta/news/2025/6/25/qewY8BHWPhcMQEYV12fkr
[69] International Computation and AI Network (ICAIN). (n.d.). *ICAIN – International Computation and AI Network*. Abgerufen am 5. Oktober 2025, von https://icain.ch/

4.2 Bewertung und Handlungsbedarf

Die Analyse der Inhalte und des Stands der Umsetzung der NCS in den fünf strategischen Zielen hat gezeigt, dass die NCS in den Beschreibungen der Ziele technologieneutral bleibt und nicht auf die konkreten Herausforderungen durch KI eingeht. Für alle strategischen Ziele werden Projekte mit direktem KI-Bezug umgesetzt.

Der Ansatz der NCS, die Massnahmen nicht auf einzelne Technologien auszurichten, bewährt sich demnach. Die NCS folgt einem prozessbasierten Ansatz, der es erlaubt, Technologietrends frühzeitig zu erkennen, systematisch zu bewerten und in die übergeordneten strategischen Ziele zu überführen. KI ist ein Beispiel für eine Technologie, die sich von einem Trend zu einem universell eingesetzten Werkzeug entwickelt. Solche Entwicklungen sollten nicht isoliert betrachtet werden, sondern als Teil eines kontinuierlichen Prozesses, der flexibel genug ist, um neue Technologien, sobald sie eine gewisse Reife und Relevanz erreicht haben, strukturell in bestehende Sicherheitsstrategien zu integrieren. Dies stärkt die Zukunftsfähigkeit und Reaktionsgeschwindigkeit der NCS gegenüber technologischen Veränderungen und neuen Bedrohungslagen.

Dennoch wurde bei der Analyse der Umsetzungsarbeiten in den strategischen Zielen auch Handlungsbedarf offensichtlich. Der prozessorientierte und technologieneutrale Ansatz der NCS hat zur Folge, dass der Einfluss der KI auf die Cybersicherheit nicht systematisch dokumentiert und bewertet wird. Dieser ist in einzelnen Projekten zur Erreichung der verschiedenen strategischen Ziele zwar adressiert, es fehlt aber eine Übersicht und eine umfassende Analyse. Weil die KI als Querschnittsthema innerhalb der Cybersicherheit aber ein hohes Gewicht gewonnen hat, wäre eine solche übergeordnete Übersicht hilfreich, um die verschiedenen NCS Vorhaben mit Bezug zur KI besser aufeinander abzustimmen.

Als Handlungsbedarf wird deshalb festgehalten, dass NCS Massnahmen mit KI spezifische Projekte ergänzt werden und die Dokumentation und Bewertung der laufenden Arbeiten zu KI im Rahmen der NCS verbessert werden muss. Die bundesinterne Umsetzung wird durch das BACS koordiniert. Im Jahresbericht der NCS soll künftig in einem separaten Kapitel über die relevanten Arbeiten zu KI in der Cybersicherheit berichtet werden. Durch die Berichterstattung kann der Steuerungsausschuss der NCS (StA NCS) beurteilen, wo in Bezug auf KI weiterer Handlungsbedarf bei der laufenden Umsetzung besteht.

Zudem überprüft der StA NCS die NCS mindestens alle fünf Jahre, wirkt an ihrer Weiterentwicklung mit und erarbeitet Anpassungsvorschläge zuhanden des Bundesrates.[70] Erkennt der StA NCS einen Bedarf für gezielte Massnahmen im Bereich KI, kann er diese dem Bundesrat unterbreiten. Der Bundesrat wird somit bis spätestens Ende 2027 im Rahmen der geplanten Überprüfung der NCS beurteilen, ob die Integration von KI in die bestehenden Ziele und Massnahmen weiterhin zweckmässig ist oder ob KI als eigenständiges Themenfeld in der NCS verankert werden soll.

5 Schlussfolgerungen

Im Bericht wurde aufgezeigt, dass KI sowohl Chancen als auch Risiken für die Cybersicherheit der Schweiz mit sich bringt. KI wirkt dabei als Katalysator bestehender Trends in der Cybersicherheit, ohne jedoch die grundlegenden Prinzipien massgeblich zu verändern. Sie beschleunigt bestehende Entwicklungen, vergrössert deren Ausmass und erleichtert sowohl die Durchführung von Cyberangriffe als auch deren Abwehr. Die schnelle Verbreitung von KI-Systemen eröffnet vielfältige Anwendungsmöglichkeiten für Wirtschaft, Verwaltung und Gesellschaft, insbesondere im Bereich der Automatisierung und datenbasierten Entscheidungsunterstützung. Gleichzeitig bringt dieser Fortschritt neue sicherheitsrelevante Herausforderungen mit sich, da böswillige Akteure KI zunehmend für gezielte Angriffe nutzen, etwa zur

[70] Schweizerische Eidgenossenschaft. (n.d.). *Steuerungsausschuss der Nationalen Cyberstrategie (NCS)*. Abgerufen am 5. Oktober 2025, von https://www.ncsc.admin.ch/ncsc/de/home/strategie/steuerungsausschuss-ncs.html

Automatisierung von Phishing-Kampagnen, zur Erstellung von Deep Fakes oder zur Umgehung klassischer Sicherheitsmechanismen.

Insgesamt ist zu erwarten, dass die Bedeutung der KI für die Cybersicherheit zunehmen wird. Auch wenn Automatisierungsprozesse in der Cybersicherheit schon seit vielen Jahren wichtig sind, hat die öffentliche Verfügbarkeit von KI-Modellen zu einem Sprung in der Anwendung von automatisierten Datenanalysen geführt. Diese Entwicklung hat erst begonnen und die Möglichkeiten sind noch längst nicht ausgeschöpft. Auch wenn von KI keine Revolution bisheriger Grundsätze der Cybersicherheit zu erwarten ist, so ist doch damit zu rechnen, dass die Dynamik, die Komplexität und damit auch die Unsicherheit im Cyberraum nochmals zunehmen werden.

Die Frage, ob diesen Herausforderungen auf strategischer Ebene genügend Beachtung geschenkt wird, ist berechtigt. Die Analyse der Berücksichtigung von KI in der NCS hat gezeigt, dass die NCS eine solide Grundlage bietet, um den Auswirkungen der KI auf die Cybersicherheit gerecht zu werden. Dennoch zeigt die Analyse, dass auf Projektebene weiterer Handlungs- und Koordinationsbedarf besteht. Um das Thema auf strategischer Ebene adäquat verfolgen zu können, müssen die Informationen aus den verschiedenen Projekten der Cybersicherheit mit KI-Bezug konsequenter dokumentiert und zusammengeführt werden. Auf dieser Basis ist es dann möglich, strategische Handlungsoptionen zu erarbeiten. So wird frühzeitig deutlich, welche Auswirkungen der Einsatz von KI auf die Cybersicherheit haben soll, statt die Strategie später reaktiv an technologische Entwicklungen anpassen zu müssen.

Glossar

Cyberangriff	Cybervorfall, der absichtlich ausgelöst wurde.[71]
Cyberkriminalität	Cyberkriminalität umfasst die Gesamtheit aller strafbaren Handlungen und Unterlassungen im Cyberraum. Dazu gehören Straftaten, die sich gegen das Internet, informationstechnische Systeme oder deren Daten richten, ebenso wie Delikte, die mithilfe digitaler Technologien begangen werden. Durch die fortschreitende Digitalisierung verlagern sich immer mehr klassische Vergehen in den digitalen Raum, wodurch neue Herausforderungen für die Strafverfolgungsbehörden entstehen.
Cybersicherheit	Anzustrebender Zustand, bei dem die Datenbearbeitung, insbesondere der Datenaustausch zwischen Personen und Organisationen, über Informations- und Kommunikationsinfrastrukturen wie beabsichtigt funktioniert.[72]
Cybervorfall	Ereignis bei der Nutzung von Informatikmitteln, das dazu führt, dass die Vertraulichkeit, Verfügbarkeit oder Integrität von Informationen oder die Nachvollziehbarkeit ihrer Bearbeitung beeinträchtigt ist.[73]
Deep Learning	Eine spezielle Form des Machine Learning, die mit künstlichen neuronalen Netzen arbeitet. Diese Netze sind dem menschlichen Gehirn nachempfunden und bestehen aus vielen miteinander verbundenen Schichten ("deep" = tief). Deep Learning kann besonders gut komplexe Muster erkennen, zum Beispiel in Bildern, Sprache oder Texten. Anwendungen finden sich in der Gesichtserkennung, Sprachassistenten oder beim autonomen Fahren.[74]
Generative KI	Als generative KI werden Modelle bezeichnet, die mithilfe grosser Datenmengen aus der realen und digitalen Welt trainiert werden, um eigenständig neue, multimodale Inhalte wie Texte, Bilder, Töne, Videos, Simulationen oder Programmcode zu erzeugen. Eine zentrale Untergruppe der generativen KI sind die grossen Sprachmodelle (Large Language Models, LLMs). Diese Technologien bilden die Grundlage vieler gängiger KI-Systeme im Bereich der natürlichen Sprache.[75]
KI (Künstliche Intelligenz)	Ein Sammelbegriff für Computertechnologien, die Aufgaben erledigen können, für die normalerweise menschliche Intelligenz erforderlich ist. Dazu gehören das Erkennen von Mustern, das Treffen von Entscheidungen oder das Lernen aus Erfahrungen. KI umfasst verschiedene Technologien wie Machine Learning,

[71] NCS, S.37
[72] NCS, S. 37
[73] NCS, S. 37
[74] CNAI (2024), s. 7
[75] CNAI (2024), s. 8

	Spracherkennung oder Bildanalyse. Wichtig: Alle heutigen KI-Systeme sind auf spezielle Aufgaben spezialisiert – eine KI mit menschenähnlicher Allgemein-Intelligenz existiert nicht.[76]
Machine Learning (Maschinelles Lernen)	Eine Methode, bei der Computer aus Daten lernen, ohne dass jede Regel explizit programmiert werden muss. Das System analysiert grosse Datenmengen, erkennt Muster und kann dann Vorhersagen für neue, unbekannte Daten treffen. Beispiel: Ein E-Mail-Programm lernt aus Tausenden von E-Mails, Spam von normalen Nachrichten zu unterscheiden. Machine Learning ist die Grundlage für viele moderne KI-Anwendungen und wird in der Cybersicherheit sowohl für Angriffe als auch zur Verteidigung eingesetzt.[77]
Zero-Day Bedrohung	Zero-Day-Exploit (auch Zero-Day-Bedrohung genannt) ist ein Angriff, bei dem böswillige Akteure eine Sicherheitslücke ausnutzen, für die es noch keine Lösung gibt. Man spricht von einer „Zero-Day“-Bedrohung, weil der Entwickler oder das Unternehmen, sobald die Schwachstelle entdeckt wird, „null Tage“ Zeit hat, um eine Lösung zu finden.[78]

[76] CNAI (2024), s. 7
[77] CNAI (2024), s. 8
[78] Was ist ein Zero-Day-Exploit? | Zero-Day-Bedrohungen | Cloudflare