Fazitbericht 18. Security Talk: Generative AI – Leistungsstarkes Produktionswerkzeug oder fundamentales Sicherheitsrisiko?

Referat Katharina Fulterer, Eraneos Switzerland AG

Chancen, Herausforderungen und Einsatzmöglichkeiten für künstliche Intelligenz (KI)

Was sind die tatsächlichen Auswirkungen von KI für den Mensch? Es gibt Aussagen, wie die von Bill Gates, der sagt, KI sei unser neues Mobiltelefon. Oder Sundar Pichai von Google vergleicht sie mit Innovationen wie Feuer und Elektrizität. Es gibt jedoch auch Stimmen wie die von Jeff Bezos, der davon spricht, dass wir erst am Anfang dieser Entwicklung stehen. Wo sind nun mögliche Einsatzgebiete, wo können wir KI als Unternehmen, als öffentliche Institution wirklich sinnstiftend nutzen. Ein sicher grosser Evergreen ist das Thema Routenoptimierung. Unsere Navigations-Apps, unsere Fitness-Tracker, helfen uns in verschiedenen Bereichen unseres Lebens. Auch grosse Unternehmen wie die Schweizerische Post setzen KI ein, um bei der Zustellung von Paketen die Routen zu optimieren. Dabei werden verschiedene Daten einbezogen, wie z.B. Wetterdaten, das Verkehrsaufkommen oder das Paketvolumen, um einen effizienten Einsatz ihrer Ressourcen zu gewährleisten. Auch in der Betrugsbekämpfung leistet KI einen wertvollen Beitrag und wird von Banken eingesetzt, um täglich Milliarden von Transaktionen zu überprüfen.

Generative KI als neuer Player

Es gibt mittlerweile viele Beispiele, z.B. bei der Bildgenerierung, wo **KI tatsächlich an Leistungen von Menschen herankommt oder diese sogar übertrifft.** Im Kontext von Generative AI lohnt es sich, nochmals zu differenzieren, wovon wir genau sprechen. Schlussendlich geht es um das Thema intelligente Maschinen oder intelligente Computerprogramme - künstliche Intelligenz. Seit den 1950er Jahren forschen verschiedene Experten an diesem Thema. Sei es im Bereich des Machine Learning oder in den vergangenen Jahren auch immer stärker zu neuronalen Netzwerken, die in den Bereich des Deep Learning fallen.

Wenn wir das Deep Learning ein wenig näher betrachten, dann sehen wir zwei Bereiche, die heute Entwicklungspotenzial zeigen. Das ist zum einen der Bereich der Predictive AI, die klassische AI. Dort geht es darum, auf Basis von historischen Daten Muster zu erkennen und zukünftige Ereignisse vorherzusagen. Beispiele dafür sind Wettervorhersagen oder die Routenoptimierung. Demgegenüber sehen wir seit der Veröffentlichung von ChatGPT Modelle, die

in den Bereich der generativen künstlichen Intelligenz fallen. Bei diesen geht es darum, auf Basis von Trainingsdaten komplett neue Outputs zu generieren. Diese können sowohl in textlicher Form, als Stimme, in Bildern, und vielen weiteren Modalitäten erzeugt werden. Generative **künstliche Intelligenz ist ein Medium, das uns erlaubt, in den Dialog mit künstlicher Intelligenz zu treten**. Sie ist quasi eine dialogfähige Schnittstelle, mit der wir Menschen im täglichen Leben konversieren können.

Daneben gibt es Anwendungen, die in Kombination von Predictive und Generative AI, heute um einiges besser funktionieren als vor wenigen Jahren. Dazu gehört z.B. auch das Übersetzen von Sprachen, Testing, Korrektur und Generierung von Software-Code, und auch die automatisierte Erstellung von Dokumenten. Und dann gibt es Bereiche, wo Gen-AI Outputs ermöglicht, die vorher nicht möglich waren. Z.B. eine personalisierte Kundeninteraktion oder das Thema Deepfakes.

Risiken beim Einsatz von KI

Neben einer Vielzahl an Chancen, birgt der Einsatz von KI genauso auch viele Risiken, die adressiert werden müssen. Diese können in drei Bereiche unterteilt werden. Es gibt Risiken, die uns als Menschen betreffen, z.B. wenn eine KI-Ergebnisse produziert, die einen gewissen Bias beinhalten oder persönliche Daten preisgeben. Es gibt Risiken, die Unternehmen betreffen, z.B. wenn Daten wiederverwendet werden, die gesetzlichen Auflagen unterliegen oder wenn Mitarbeiter*innen unzufrieden sind, weil sie gewisse Aufgaben an eine KI abgeben müssen. Weiter gibt es Risiken, welche die Gesellschaft betreffen, bspw. der Einfluss von künstlicher Intelligenz auf die Versorgungssicherheit der Schweiz. Schlussendlich geht es darum, vertrauenswürdige KI-Lösungen zu entwickeln und diese auch in die Anwendung zu bringen.

Anwendungen der Generative AI in Schweizer Unternehmen

Wo werden heute Anwendungen mit Generative AI von Schweizer Unternehmen im Schweizer Markt eingesetzt? Im öffentlichen Diskurs ist es nicht immer transparent dargestellt, wie anspruchsvoll der Einsatz von KI im Unternehmenskontext tatsächlich ist. Laut der OECD setzten 2023 lediglich 8% der Unternehmen im OECD-Raum KI ein. Dies liegt an der Komplexität der Implementierung. Oft scheitern Unternehmen nach einem erfolgreichen PoC (Proof of Concept), daran, das Projekt wirklich zu skalieren, es in ein produktives Setting zu überführen. In der realen Geschäftswelt kommen komplett neue Einflussfaktoren dazu und die neuen KI-Lösungen müssen an bestehende Systeme angebunden werden. Oft scheitern solche Projekte auch daran, dass gewisse Sicherheitsthemen, gewisse Governance-Themen in einem ersten schlanken PoC nicht ausreichend beleuchtet und berücksichtigt wurden. Deshalb versuchen wir als Unternehmen diese Themen ausreichend zu beleuchten und uns Gedanken zu machen, wie wir von Anfang an sichere und vertrauenswürdige Systeme konzipieren können. Ein Beispiel für eine solche Lösung ist die Aufnahme und Erfassung von relevanten Informationen.

Generative KI im Bereich des Knowledge Managements

Generative AI hat sehr breite Anwendungsfelder, aber dasjenige, das im Moment im Schweizer Markt am aktivsten von vielen Unternehmen verfolgt wird, ist im Kontext des Knowledge Managements. Wie gelingt es uns, verfügbare Daten unseren Mitarbeitenden, vielleicht in einem zweiten Schritt auch unseren Kunden, möglichst effizient und sinnstiftend zur Verfügung zu stellen? Dahinter steht die Fähigkeit, Dokumente zu analysieren, zu synthetisieren und daraus gute Zusammenfassungen zu erstellen. In der Polizeiarbeit gibt es sehr ähnliche Use Cases, insb. wenn es darum geht, Vorfälle aufzunehmen. Dies ist ein relativ mühsamer Prozess, den wir aktuell in einem PoC zu automatisieren versuchen, indem eine KI gesprochene Sprache entsprechend aufnimmt, ein Protokoll erstellt und die wichtigen Punkte nochmal zusammenzufasst und in die vorhandenen Systeme überträgt.

Generative KI bei Chat- und Voicebots

Ein weiterer Anwendungsfall von KI sind Chat- und Voicebots. In der heutigen Version sind diese für Kund*innen oft ein echtes Übel. Generative KI ist mittlerweile in der Lage, die gesprochene Sprache, und zwar nicht nur das Hochdeutsch, sondern auch die verschiedenen Schweizer Dialekte, zu verstehen, die Absicht der Kunden zu erkennen, und daraus die entsprechenden Aktionen abzuleiten.

Generative KI im militärischen Kontext

Weiter sehen wir im militärischen Kontext viele Einsatzmöglichkeiten für generative KI. Das könnte z.B. die Simulation von Trainingsszenarien sein. Heute werden Newsartikel, Social-Media-Einträge, Twitter-News in mühsamer Handarbeit analysiert aufbereitet, um Simulationen und Trainings möglichst realitätsnah wirken zu lassen. Genau hier kann Generative AI mit seiner kreativen Kraft einen entsprechenden Beitrag leisten.

Referat Stefan Preuss, Mobiliar

Generative KI zu Marketingzwecken

Generative KI ist für Marketingzwecke das Einfachste. Die Technologie funktioniert hervorragend. Man kann damit so viel machen. KI ist überall. Aktuell läuft die nächste Welle der Text-to-Video oder Text-to-Image-Generation durch. Es gibt keine Guardrails Content-Policy, die vorgeben, was erlaubt ist und was nicht. So hat man auch die Möglichkeit,

problematische Videos zu generieren. Möglicherweise kennt jemand das Tool Flux von Black Forest, eine deutsche Firma. Das ist einer der schnellsten Bildgeneratoren aktuell auf dem Markt. Wenn man Flux sagt, dann bekommt man einen weissen männlichen Doktor in meinem Alter. Demnach steckt ein Bias in diesem Tool, weil es diese Aufgabe gar nicht interpretieren kann. Wenn man das hingegen mit einer Google-Suche vergleicht und dieselbe Frage stellt, war das Resultat vor fünf Jahren noch gleich, heute aber etwas gemischter. Diese Tools haben demnach ihre Grenzen.

Gescheiterte KI-Anwendungen

«The best AI fails 2020-2024»: Wir haben ca. 2020 angefangen, grössere AI Fails zu sammeln. Ein schönes Beispiel ist ein Schachroboter, der in der Lage ist, Schach zu spielen und simultan mehrere Bretter zu bedienen. Dieser war jedoch nicht darauf trainiert, dass der kleine Junge, gegen den er gespielt hat, einfach mal ins Spielfeld reingreift. Der Roboterarm hätte dann stoppen und rausgehen sollen. Er hat die Situation jedoch nicht erkannt, hat zugegriffen und dem Jungen den Finger gebrochen. Ein etwas aktuellerer Fall von McDonalds handelt von einem relativ grossen Projekt, das eingestellt werden musste. Man wollte generative KI in einem Drive-Through einsetzen, um den Bestellvorgang zu automatisieren. Das führte dazu, dass 260 Eistees bestellt und verschiedene Parallelbestellungen von anderen Standorten mit aufgenommen wurden. Wir müssen aus diesen Beispielen lernen, wie Fehler zustande kommen und vermieden werden können.

Vertrauen in KI als essenzielle Komponente

Warum sollten wir KI vertrauen? Nach der Wall of Shame hat man erstmal Zweifel an dieser Technologie. Wir nutzen Technologie in unserem täglichen Leben tausendmal und vertrauen ihr dementsprechend auch tausendmal. Wir fahren mit der Bahn, wir vertrauen dem Zugführer und der SBB, dass der Zug ankommt. Wir vertrauen dem Piloten, der unser Flugzeug fliegt, dass er dieses heil landen wird. Am Ende **geht es auch bei der KI um das Thema Vertrauen**. Dazu gab es einen sehr spannenden Artikel von Bruce Schneier, einem Cyber-Security-Spezialisten. Dieser hat vier Dimensionen definiert, wann wir der Technologie vertrauen können. Zum einen, wenn die Technologie an sich Sicherheitsmechanismen impliziert und ausgereift ist. Wenn es innerhalb der anbietenden Firma Werte gibt, die nicht nur in der Gewinnmaximierung liegen, sondern auch darin, ein Produkt bereitzustellen, das funktioniert. Oder auch das Bewusstsein für die eigene Reputation, was mit diesem Produkt verbunden ist. Und schlussendlich eine sinnvolle Regulierung, die auch ein Bestrafungssystem für Fehlverhalten und Versagen mitberücksichtigt.

Es stellt sich die Frage, vertraue ich der Technologie und würde ich diese nutzen? Schneier hat gesagt, bei der Secure Technology und bei der Regulation sind sogenannte Social Trust Merkmale relevant und bei Reputation und Value sind es Interpersonal Trust Merkmale. Und wenn wir über generative KI sprechen, dann sehen wir sehr viel Menschliches in dieser Technologie. Die KI antwortet auf eine Art, wie wir auch sprechen. Und deshalb legen wir verstärkt Wert auf den Interpersonal Trust und vernachlässigen den Social Trust etwas zu sehr.

Autonomes Fahren als Beispiel

Ich nehme das autonome Fahren als veranschaulichendes Beispiel, welches jedoch kein spezifisches Gen-AI Thema ist. Die Automobilhersteller sagen alle, die Technologie sei bereit und funktioniere. Ich als kritischer Betrachter würde sagen, wenn das autonome Fahren bereit ist, dann müsste auch die Zahl der LKW-Fahrer*innen sinken, was gemäss den Zahlen in den USA und in Europa nicht der Fall ist. Das steht also völlig im Widerspruch zur Technologie und zum Technologieversprechen. Warum? Die Technologie ist vorhanden. Wir haben in modernen Fahrzeugen Radar, GPS, 4G, Sensoren, die von 2 cm bis 250 m jeden möglichen Wert messen und jedes Fahrszenario 10 Sekunden im Voraus berechnen. Die modernen Fahrzeuge heutzutage sind Rechenzentren auf Rädern. Die Technologie ist also bereit. Ist es eine Frage der Regulierung? In der Schweiz ist mittlerweile zumindest ein Entwurf für die Regulierung des autonomen Fahrens vorhanden. In Deutschland ist das autonome Fahren auf Level 5 möglich, bis 60 km/h auf der Autobahn (!).-Weder die Technologie, die Regulation, noch die Nutzenden sind das Problem. Wenn wir nach San Francisco schauen, läuft momentan das grösste Experiment in autonomem Fahren. Es sind ungefähr 1600 Fahrzeuge im Einsatz. Schaut man sich die Fahrberichte an, so ist es rein statistisch die sicherste Art des Fahrens. Seit dem Uber-Unfall vor sechs Jahren passieren keine tödlichen Unfälle mehr. Die Technologie ist ausgereift, jedoch bleiben Herausforderungen bestehen.

Schauen wir uns an, was alles schief gehen kann. Möglichkeit eins, das Telefonnetz fällt aus. Nummer zwei, die Polizei hält ein autonomes Fahrzeug an, das Fahrzeug bremst kurz, die Polizei steigt aus, läuft hin, das Fahrzeug gibt Gas und fährt weiter, um 200 Meter später wieder anzuhalten. Nummer drei, jugendliche Gangs machen sich einen Spass daraus, die Sensoren mit Verkehrspylonen abzudecken und erzeugen so einen künstlichen Verkehrsstau. Nummer vier, ein menschlich gefahrenes Fahrzeug fährt eine Frau an, die Frau wird unter ein autonomes Fahrzeug geschleudert, das autonome Fahrzeug ist auf diese Situation nicht Zuletzt hat sich jemand den Spass gemacht, ein T-Shirt mit einem Stoppzeichen anzuziehen und hat sich an den Strassenrand gestellt. Das autonome Fahrzeug hielt an, weil es ein Stoppzeichen erkannt hatte.

KI funktioniert zu 95-99 Prozent, nie zu 100 Prozent

Es gibt Situationen, die man nicht vorhersehen kann. Ich habe es das «Fly-and-Cow-Dilemma» genannt, weil niemand damit rechnet, dass eine Kuh durch die Gegend fliegt und das Fahrzeug dieser ausweichen muss. KI wird also zu 95 oder zu 99 Prozent funktionieren, aber sie wird eben nie zu 100 Prozent funktionieren. Deshalb müssen wir in der Lage sein, auch dieses eine Prozent so abzufedern, dass keine Schäden eintreten und insbesondere keine Personen zu Schaden kommen.

Zusammenfassend schreibt die Realität ihre eigenen Gesetze, auch im Zusammenhang mit KI. Selbst wenn wir alle beschriebenen Dimensionen, so gut wie möglich umsetzen und so gut wie möglich zu beherrschen versuchen. Wir haben das Gefühl, wir müssten heute alles mit KI machen, **aber nicht jedes Problem lässt sich mit KI lösen**. Es lässt sich vielleicht anders

lösen. Dennoch ist es wichtig, zu überlegen, wann KI und wann Human Power oder alternative Lösungen geeignet sind.

Referat Dr. Thomas Rothacher, armasuisse

Technologieentwicklung im letzten Jahrhundert

In den letzten 100 Jahren sieht man eine exponentielle Zunahme des Technologietransfers in unserem Umfeld. Früher war das Militär noch ein Technologietreiber. Heute ist es hingegen die zivile Welt, welche die Technologie antreibt. Die finanziellen Verhältnisse haben sich geändert. Weiter hat der Technologiewandel viel mehr Einfluss auf unsere Kultur als wir bemerken. Das iPhone, welches seit 14 Jahren im Gebrauch ist, hat die Welt disruptiv verändert. Wissen hat durch Google einen anderen Wert bekommen und künstliche Intelligenz wird uns diese Tatsache noch deutlicher vor Augen führen. Ein weiteres Phänomen ist die Durchdringung dieser Technologie, welche immer schneller und breitflächiger geschieht. Bei Netflix hat es noch dreieinhalb Jahre gedauert, bis sie eine Million Abonnent*innen / Nutzer*innen erreicht hatten. ChatGPT benötigte dagegen gerade einmal fünf Tage, bis dieser Meilenstein erreicht wurde. Eine weitere Grafik zeigt, wie lange es gedauert hat, bis ein Algorithmus eine gewisse Tätigkeit besser ausführen konnte als der durchschnittliche Mensch. Bei der Handschrifterkennung hat dies über zehn Jahre gedauert. Wenn wir ins Jahr 2018 gehen, hat es bei der Spracherkennung und dem -verständnis nur noch zwei Jahre gedauert. **Es gab also eine enorme Beschleunigung dieser Fähigkeitsentwicklungen.**

Einfluss der generativen KI auf die Cybersicherheit

Wenn wir auf armasuisse zu sprechen kommen, haben wir einen grossen Vorteil gegenüber privatwirtschaftlichen Unternehmen. Wenn diese ein Produkt verkaufen wollen, müssen alle Eventualitäten ausgeschlossen werden können. Sie haben vorhin gehört, es könnte schwierig werden, das letzte Prozent bis zur vollständigen Kontrolle der KI zu erreichen. Das ist bei W&T (Abteilung Wissenschaft und Technik der armasuisse) etwas einfacher. W&T beschäftigt die Frage, wie wir neue Gadgets zur Truppe oder in die Anwendung bringen können und dies noch im Versuchs- oder im Teststadium.

Wir haben uns gefragt, wo wir diese Technologie einsetzen sollen. Es gibt in der Armee gewisse Wettbewerbe, wo Gut gegen Böse spielt. Wir sind als Schweiz an einem internationalen

Wettbewerb angetreten und 2017 im hinteren Drittel gelandet. Daraufhin haben wir Daten von mehreren Jahren gesammelt und damit Algorithmen programmiert und trainiert. Beim Wettbewerb Block Shields 19 sind wir dann dank der Hilfe dieser Algorithmen im vordersten Drittel gelandet. Diese Form der KI ist bereits seit 2020 in den Systemen der Armee operativ im Einsatz.

Lernfähigkeit der KI

Es gibt heute bereits War Games, die mittels künstlicher Intelligenz gespielt werden können. Wir können bspw. zwei künstliche Intelligenzen gegeneinander spielen lassen. Es geht hier darum, Truppen mit verschiedenen Technologien auszurüsten und sie dann gegeneinander zu positionieren. Das Erstaunliche daran ist, dass der Lösungsraum, den eine künstliche Intelligenz vorschlägt, ein anderer ist als der, den wir wählen würden. Obwohl die künstliche Intelligenz ein Bias haben kann, ist der Bias des Menschen im Normalfall noch ausgeprägter. Wir haben die Technologie im neuen taktischen Simulator der Luftwaffe angewendet. Die künstliche Intelligenz berechnet anhand einer Vielzahl an Daten, welches der optimale Weg für den Gegner wäre, um unsere Stellungen anzugreifen. Daraus lernen wir wiederum, was dies für unsere Systeme bedeutet - wie sie beschaffen sein müssen und wie wir sie am besten anordnen.

Durch KI gesteuerte Drohnen

Drohnen sind eines der grossen Themen, das von der künstlichen Intelligenz befeuert oder beschleunigt worden ist. Ein ukrainischer Offizier, der Leiter dieser Drohneneinsätze ist, hat uns berichtet, in der Ukraine bräuchten sie im Moment etwa alle drei Monate eine Million Drohnen. Weiter würden sie 95% der geschützten Fahrzeuge mit diesen Drohnen zerstören, weil sie nicht über ausreichend Artillerie-Munition verfügen. Die Drohnen werden in die Luft gesetzt und dann gilt «Fire and Forget». Sie sind also nicht mehr unter menschlicher Kontrolle, sondern suchen sich ihre Ziele mittels künstlicher Intelligenz selbst aus und werden auch stetig weiterentwickelt. Die Ukraine ist das erste Land, welches eine unbemannte Einheit als eigene Gattung eingeführt hat. Das ist ein Thema, das wir heute in einigen Bereichen der Armee bereits diskutieren.

Auch die Schweiz ist im Bereich KI und Drohnen eine führende Nation. Unsere Hochschulen werden als Teil der grössten Talentschmiede der Welt angesehen. Wir sind im globalen Index der Innovation auch seit einigen Jahren die Nummer eins. Armasuisse hat sich deshalb dafür eingesetzt, dass in den letzten zwei Monaten die Task Force Drohnen gegründet wurde. Dies mit der Absicht, die einzelnen Player in der Schweiz zu vernetzen und einen Beitrag für unsere Sicherheit zu leisten.

«Nicht die intelligenteste Spezies wird überleben, sondern jene, die sich am besten anpassen kann.»

Mein Fazit: die Geschwindigkeit der angesprochenen Entwicklungen nimmt exponentiell zu, besonders im Bereich der Sicherheit. Und wir sehen heute Revolutionen in der Kriegsführung, die über allfällige Disruptionen hinausgehen. Weiter rüsten die Nationen um uns herum

alle enorm auf. Das ist jetzt nicht mein Votum als Leiter von W&T, sondern als stellvertretender Rüstungschef. Die Schweizer*innen befinden sich in einer Bubble und nehmen nicht wahr, was um uns herum passiert. Wir müssen uns jedoch an unserem Umfeld orientieren. Denn, es wird nicht die intelligenteste Spezies sein, die am Schluss überleben wird, sondern diejenige, die sich am besten anpassen kann.

Paneldiskussion

Neben den beiden Referenten, Stefan Preuss und Dr. Thomas Rothacher, nahmen am Panel weitere Experten teil: Jennifer Scurrell, Doktorantin am Center for Security Studies der ETH Zürich, Patrick Fontana, Digital & App Innovation Specialist bei Microsoft und Dr. Peter Friedli, Partner bei Eraneos Switzerland. Moderiert wurde das Panel von Fredy Müller, Geschäftsführer des FORUM SICHERHEIT SCHWEIZ.

KI als fundamentale Veränderung

Fredy Müller: Stefan Preuss, ist die Technologie-Revolution, die wir mit künstlicher Intelligenz erleben, ein ähnlich grosser Entwicklungsschritt wie Lesen, Schreiben oder Rechnen?

Stefan Preuss: Meiner Meinung nach, definitiv ja. Wenn man die Entwicklung seit der industriellen Revolution sieht, beginnend von der Dampfmaschine, über die Elektrizität, die Informationstechnologie, und jetzt auch das Thema mit der künstlichen Intelligenz, ist es definitiv revolutionärer Schritt, denn dahinter stecken Daten. Und wir haben diese Daten nun auswertbar zur Verfügung, wir können sie nutzen. Schaut man nicht nur auf die generative KI, sondern gerade auf Bilderkennungsverfahren im medizinischen Bereich, gibt uns das einen unheimlichen Push.

Freddy Müller: Patrick Fontana, Sie sind seit 20 Jahre im Tech- Bereich tätig. Was hat sich innerhalb dieser Zeit bei Microsoft verändert?

Patrick Fontana: Sehr viel. Zum einen hat man den Schritt von einer reinen Verkaufs- und Lizenzorganisation in eine Technologie-Weiterentwicklungs- und -Adaptierungsfirma gemacht. Die künstliche Intelligenz ist dazu der nächste Schritt. Man spricht nicht mehr einfach

davon, ein Produkt herauszubringen, sondern wir stellen die meisten Technologien, als Services zur Verfügung. Ein einzelnes Unternehmen hat oft gar nicht mehr die Ressourcen, alles selbstständig zu machen. Das bringt auch im Unternehmen immer wieder einen Twist, weil man sich genau überlegen muss, was der Use-Case hinter den KI-Anwendungen ist. Alle gescheiterten Implementierungsversuche hatten keinen klar definierten Use-Case.

Freddy Müller: Also es nicht mehr das Ziel, ein Produkt per se auf den Markt zu bringen, sondern eine Dienstleistung oder ein Gesamtsystem?

Patrick Fontana: Zum einen sicher eine Dienstleistung, zum anderen aber auch ein Ökosystem.

Freddy Müller: Peter Friedli, Du bist seit vielen Jahren in der Beratung, im Sicherheitsbereich tätig. Wie hast Du diesen Technologiewandel erlebt?

Dr. Peter Friedli: Ich spüre einen grossen Bedarf nach Beratungsdienstleistungen. Wie Patrick Fontana es gesagt hat, müsste man eigentlich überlegen, welche Probleme wir effektiv haben und wie wir diese lösen können. Ob dann KI oder generative KI die Lösung ist, sei dahingestellt.

Freddy Müller: Jennifer, Du gehörst zur jüngeren Generation und bist mittendrin auch mit deinen Forschungen. Wie erlebst Du diesen Hype der KI und der generativen KI?

Jennifer Scurrell: Ich würde KI nicht als Hype bezeichnen, sondern als Revolution. Vor allem angesichts meines Hintergrunds mit Human-Air-Interaction. Da sieht man schon, was für Auswirkungen KI gerade heute hat. Seien es die Chatbots, die immer smarter werden, aber auch im Bereich der Medizin oder der Psychotherapie können diese Bots viel Gutes bewirken. Ich beschäftige mich seit ungefähr acht Jahren mit Machine Learning, Deep Learning und AI und möchte daher nochmal sagen, diese Technologie ist absolut revolutionär.

Verständnis über die Funktionsweise der KI - keine zwingende Voraussetzung für Vertrauen

Freddy Müller: Wie gut kennen die Leute KI und generativer KI überhaupt?

Jennifer Scurrell: Erstaunlicherweise benutzen wirklich sehr viele Leute ChatGPT, beispielsweise auch meine Mutter. Ich würde nicht sagen, dass sie weiss, wie es funktioniert. Aber muss sie das wirklich wissen? Es ist doch wichtiger, allgemeine Skills im Umgang mit Technologien wie dem Internet, Social Media und KI zu erwerben. Das kritische Denken ist im Umgang mit jeglicher Technologie essenziell wichtig.

Freddy Müller: Stefan Preuss, wie viel muss man aus einer Unternehmensperspektive über die Technologie wissen? Soll man sie einfach anwenden und als User nicht zu viel hinterfragen?

Stefan Preuss: Das ist aus der Perspektive der Versicherung vielleicht nicht der beste Weg, weil man dann sehr stark mit dem Kundenvertrauen spielt. Ich würde einen vorsichtigen und

bewussten Einsatz von solchen Technologien, das Ausprobieren und Vergleichen nahelegen. Es wird augenscheinlich, wie stark eine Technologie z.B. bei der Schlachtfeldentwicklung ein bisher gewohntes Verhalten komplett über den Haufen wirft und Dinge verändert. Aber das passt natürlich nicht in jedes Unternehmensszenario.

Die sechs Prinzipien für den Einsatz von Responsible AI

Freddy Müller: Patrick Fontana, wir kennen die sechs Prinzipien, die es braucht, um KI verantwortungsbewusst zum Einsatz zu bringen. Bei Microsoft ist der Umgang mit diesen Prinzipien sicher ein ständiges Thema.

Patrick Fontana: Die sechs Prinzipien für den Einsatz von Responsible AI sind definitiv ein Thema, aber sie gehen schlussendlich wieder genau in die Thematik rein, die wir vorhin gehört haben. Denn wenn wir eine Technologie einsetzen, müssen wir darauf vertrauen, dass der Output auch stimmt. Mit Blick auf die Prinzipien, muss eine AI transparent, sicher und nachvollziehbar sein, um Biases zu verhindern. Wenn ein Unternehmen dahinterstehen kann, dann schafft das Vertrauen. Es bleibt die Frage, ob das gesetzlich festgehalten werden muss, damit sich jeder daranhält, oder ob eine Selbstdeklaration genügt, welche keine Konsequenzen nach sich zieht. Das ist die Frage hinter den sechs Prinzipien von Responsible AI.

Regulation der KI

Freddy Müller: Reicht eine Selbstdeklaration aus oder brauchen wir Gesetzesbestimmungen für KI, um Missbräuche zu verhindern?

Dr. Thomas Rothacher: In der militärischen Welt ist es mit der Regulierung etwas schwierig. Ich bin jemand, der nicht wirklich an diese Regulierung glaubt. Denn all jene Entwicklungen und Funktionen, die möglich sind, werden auch früher oder später eintreten. Wir müssen uns mehr Gedanken machen, wie wir damit umgehen wollen und nicht etwas regulieren, das sowieso kommen wird. Die Herangehensweise der EU an eine solche Regulation ist die schlechteste Art und Weise.

Jennifer Scurrell: Ich stimme auch aus wissenschaftlicher Sicht zu. Eine sinnvolle Regulation ist in Ordnung, aber ich finde es problematisch, wie die EU die Thematik angeht. Fortschritt und Kreativität werden behindert, was somit auch Europa als potenziellen Leader im KI-Bereich hemmt. Deshalb braucht es eine sinnvolle, aber geringfügige Regulation.

Freddy Müller: Es ist eine Gratwanderung. Stefan Preuss, Du hast es auf einer Folie gezeigt, es braucht Regulierung.

Stefan Preuss: Mir schlagen diesbezüglich zwei Herzen in der Brust. Zum einen bin ich der Meinung, es braucht zwingend Regulation für eine disruptive Technologie wie KI. Wir haben es mit dem Internet und Social Media verpasst zu regulieren und heute versucht man mühsam, die Büchse der Pandora wieder zu schliessen. Insofern war ich froh, dass es den EU AI Act gab. Es braucht jedoch noch einige Jahre, bis man die Objekte, die man regeln möchte, definiert hat und versteht, wovon man überhaupt spricht. Insofern bin ich da sehr zwiegespalten. Aus meiner Perspektive bin ich natürlich froh, wenn ich eine Sollvorgabe habe, an der ich mich orientieren kann. Aber es braucht noch viel Zeit in der Umsetzung.

Freddy Müller: Patrick Fontana, bräuchte es eine globale oder lokale Regulierung? Es wurde gesagt, dass man im Bereich Soziale Medien wahrscheinlich einiges verpasst hat. Microsoft wird in diesem Zusammenhang oft angeprangert.

Patrick Fontana: Für uns wäre es natürlich deutlich einfacher, wenn alles irgendwo global geregelt würde. Dann könnten wir unsere ganzen Compliance-Prozesse mit dieser Regulierung abstimmen und hätten dementsprechend weniger Aufwand. Fakt ist, dass zurzeit jeder einzelne Staat und jede Organisation versucht, solche Regulierungen einzuführen. Wir müssen uns jeweils diesen Regulierungen gegenüber beweisen und diese einhalten. Das machen wir gerne, denn es fördert das Vertrauen. Wir sind ein wenig zwiegespalten: Regulierung ja, aber nur so, dass Innovation und Entwicklung nicht gehemmt werden.

Durch (generative) KI erreichte gesellschaftliche und politische Veränderungen

Freddy Müller: Als nächstes sprechen wir über die gesellschaftlichen und politischen Veränderungen, welche durch KI oder generative KI generiert werden. Das ist vor allem das Forschungsgebiet von Jennifer Scurrell.

Jennifer Scurrell: Ich schaue mir vor allem an, wie Chatbots in den sozialen Medien die politische Meinung beeinflussen können. Das Thema war 2016 erstmals gross in den Medien präsent, als der Verdacht bestand, dass russische Bots die US-Wahlen beeinflusst hatten. In diesem Fall waren sich die Wissenschaftler*innen uneinig. Einige Studien haben gezeigt, dass die Bots aufgrund ihrer grossen Anzahl tatsächlich die Ergebnisse beeinflusst hatten. Andere Studien belegen jedoch, dass diese Bots nicht klug genug waren, um die politische Meinung effektiv beeinflussen zu können. Mit ChatGPT haben wir eine völlig andere Situation, weil die Bots immer besser und schlauer werden. Leute können Bots teilweise nicht mehr von Menschen unterscheiden. In den letzten Monaten wurden viele Studien zu diesem Thema veröffentlicht. Es konnte gezeigt werden, dass diese Bots andere Leute mindestens genauso gut wie Menschen mit Propaganda beeinflussen können. Hier haben wir bereits ein grosses Problem.

Umgang mit Bots

Freddy Müller: Patrick Fontana, wie geht ein Unternehmen wie Microsoft mit dieser Bot-Thematik um?

Patrick Fontana: Wir als Microsoft bieten ja nicht einen Bot an, der alles kann. Wir bieten vielmehr klare Referenzarchitekturen an und unterstützen die Kund*innen direkt. Bei der Verwendung von OpenAI haben wir ein Zwei-Zonen-Sicherheitssystem. Wir haben eine maschinelle Rastererkennung für Fragen, die nicht gestellt werden sollen, wie beispielsweise zum Thema Suizid oder Fragen mit manipulativem Charakter. Aus datenschutzrechtlichen Gründen kann diese menschliche Komponente auch herausgenommen werden. In diesem Fall müssen wir uns als Unternehmen jedoch absichern, denn wir sind in der Verantwortung, dass gewisse Fragen nicht beantwortet werden dürfen. Ein Beispiel hierzu ist ein Proof of Concept, den wir mit der Armee durchgeführt haben.

KI und die junge Generation

Freddy Müller: Peter Friedli, wie geht ein Beratungsunternehmen mit dieser Herausforderung um?

Dr. Peter Friedli: Wir sind ja fast alle Digital Natives, oder sind es zumindest geworden. Meine 9-jährige Tochter kommt jedoch in eine Welt hinein, in der alles Digitale bereits normal ist. Wie bringen wir diese Generation dazu, damit umgehen zu können? Es ist gewissermassen ein Train-the-Trainer-Prinzip, wie man es aus der Armee kennt. Sind wir als Trainer bereit, das auch weiterzugeben? Sind die Lehrkräfte bereit, das weiterzugeben? Ich denke, das sind auch Fragen, die sich auf die Unternehmen übertragen lassen.

Jennifer Scurrell: Viele Menschen, seien es Schüler*innen oder Erwachsene, sind **betroffen,** aber **nicht darauf vorbereitet.** Wir werden von den sozialen Medien täglich beeinflusst und mit Informationen bombardiert. Als ChatGPT herauskam, wollte man es an den Schulen und Universitäten verbieten. Das ist jedoch der falsche Ansatz. Wir müssen den Kindern von klein auf beibringen, wie man mit diesen Technologien umgeht. In einem Fach «Technologie» könnte das einmal pro Woche thematisiert werden. Es würde darum gehen, spielerisch auszuprobieren, welche Fragen welche Antworten generieren, welche Daten verwendet werden und wie diese kombiniert werden?

Freddy Müller: Sollte die Gesellschaft, primär in Unternehmen, trainiert und bewusst gemacht werden, was uns hier bevorsteht?

Stefan Preuss: Es ist fast unmöglich, das aus der Elternperspektive hinzubekommen. Wenn man sieht, was über die sozialen Medien an die Kinder herangetragen wird und dass KI vieles noch verstärkt, ist es ein sehr schwieriger Kampf. Auf der anderen Seite ist es in der Technologiefolgenabschätzung spannend. Denn wir sind das erste Mal mit einer Technologie konfrontiert, die uns zwingt, über unsere ethischen Grundhaltungen nachzudenken. Wollen wir Human-in-the-Loop für die Waffensysteme beibehalten? Oder was bedeutet es, Social Scoring Systeme in der Gesellschaft zu integrieren, nicht um sie anzuwenden in der Schweiz, sondern um sie zu verkaufen?

Freddy Müller: Thomas Rothacher, was können wir tun, um uns von Deepfakes nicht täuschen zu lassen, sei es in der Zivilgesellschaft oder in der militärischen Welt?

Dr. Thomas Rothacher: Ich verstehe die Technologie hinter KI nicht wirklich, aber was ich gelernt habe, ist gewisse Plausibilitätsüberlegungen zu machen. Es hilft bei Deepfake-Videos, einmal zu überlegen, wie realistisch ist das Gezeigte. Aus meiner Sicht sollte dieses kritische Denken und kritische Diskutieren gezielter gefördert werden.

In den Unternehmen braucht es eine Adaptation im Umgang mit KI

Freddy Müller: Patrick Fontana, ich vermute, dass Microsoft aufgrund seiner Expertise sehr gefragt ist an den Schulen etc.. Habt Ihr Gelegenheit, dort aufzutreten?

Patrick Fontana: Wir stellen uns aktiv zur Verfügung. Die Nachfrage ist jedoch v.a. auf Grundschulniveau sehr schwach, ebenso im Bereich der Basislehre oder der achten und neunten Klasse. Die Hochschulen hingegen beschäftigen sich aktiv mit der Thematik. Auch wenn wir in den Unternehmen unterwegs sind, sagen wir immer, dass es ein Adaptionsprogramm braucht. Dabei besteht der erste Schritt darin, zu lernen, mit dieser Technologie umzugehen und die gelieferten Informationen anständig zu validieren.

Freddy Müller: Gibt es einen Bedarf, dass wir mit dem Internet, Social Media und KI professioneller umgehen sollten?

Dr. Peter Friedli: Aus meiner Sicht sollte es dort, wo es noch keinen Bedarf gibt, zwingend einen geben. Für Unternehmen, die aus der Ideation-Phase raus sind, schreiben wir oft KI-Strategien. Dabei geht es nicht nur um den technischen Teil. Es geht vielmehr um das Operative, um Prozesse und Überlegungen zum Datenabfluss. Ein typischer Use Case: ich muss eine Antwort auf ein Mail schreiben. Dazu kopiere ich den Inhalt, rufe ChatGPT auf und lasse mir Antworten erstellen. Bei diesem Vorgang fliesst eine grosse Menge an Daten ab. Dagegen hilft keine Firewall. Zum Schluss folgt eine Art Portfolio-Management und eine Priorisierung der KI-Aktivität. Die Technik ist also nur ein kleiner Teil dieser Prozesse.

Erfolgreiche Anwendung der KI in Unternehmen und Behörden

Freddy Müller: Kommen wir zum dritten Themenblock, zur Anwendung von KI in Unternehmen und bei Behörden. Stefan Preuss, Du hast anschaulich dargestellt, was beim Einsatz von KI schiefgehen kann. Was sind die drei wichtigsten Elemente, damit der Einsatz von KI gelingt?

Stefan Preuss: KI basiert, wie bereits erwähnt, auf Unternehmensdaten. Nur damit kann für den unternehmensspezifischen Kontext auch ein Mehrwert generiert werden. Die **Informationen müssen in geeigneter Qualität und Menge vorliegen**. Das ist für die meisten Unternehmen wahrscheinlich die grösste Herausforderung. Ausserdem sollte man innerhalb eines **Strategie-Szenarios, möglichst viele kleinere Spielszenarien zulassen**, bevor man den ganz grossen Wurf macht. Denn auch wir bei die Mobiliar können uns keine 99 Prozent Genauigkeit leisten, wir brauchen 100 Prozent. Bevor man diesen Schritt macht, sind also noch sehr viele Erfahrungswerte nötig.

Freddy Müller: Wir möchten an diese Stelle das Publikum fragen: Nutzen Sie KI, haben Sie bereits Erfahrungen? Wer hat ein Chatbot im Einsatz?

Georg Kaufmann: Ich bin Stabschef im Personalbereich der Armee. Wir sind gerade dabei, einen Chatbot einzuführen. Zurzeit haben wir eine Hotline, die jährlich 30.000-40.000 Anfragen beantwortet. 60 Prozent davon per Telefon und 40 Prozent per E-Mail. Der Chef der Armee möchte jetzt ein neues Projekt, den First-Level-Support, integrieren. Dieses würde einen zusätzlichen Personalbedarf 1,6 Millionen bedeuten, was nicht infrage kommt. Daher sind wir beim Chatbot gelandet und jetzt im MVP. Zurzeit führen wir Truppenversuche durch. Danach werde ich das Projekt der Armeeführung präsentieren. Diese wird dann entscheiden, ob der Chatbot zum Einsatz kommt, denn auch dafür sind finanzielle und personelle Ressourcen notwendig.

Patrick Fontana: Das Kommando Ausbildung hat schon länger ein Projekt am Laufen. Mit KI treten wir häufig über den Support Kanal in Kontakt. In diesem Bereich hat ChatGPT die Komponente des Sprach-Textverständnis reingebracht, was uns erlaubt, aktiv mit unseren Daten zu kommunizieren. Der generativen KI fehlt im Moment eine relativ wichtige Komponente, jenes des Überblicks über den ganzen Case sowie verschiedene Varianten zu finden. Hier sind wir im Moment limitiert. Das wäre der nächste Entwicklungsschritt. Heute stabilisiert man die ganze Phase der KI. Will man den nächsten Schritt überhaupt gehen? Zuerst müssen diesbezüglich ethische Fragen geklärt werden, bevor die KI quasi selbstdenkend agiert, die Situation überblickt und Varianten findet.

Freddy Müller: Wir sprechen von AGI, Artificial General Intelligence. Was ist Deine Einschätzung, wann kommt das?

Patrick Fontana: Wie immer plant Microsoft bei solchen Roadmaps maximal zwei Jahre in die Zukunft und schaut dann weiter. Bill Gates hat in seinem letzten Podcast, The Next Big Thing, gesagt, er gehe von 10-15 Jahren aus. Ich würde mich dieser Meinung anschliessen.

KI im militärischen Kontext

Freddy Müller: Wir kommen zum vierten Themenblock, zum militärischen Kontext. Thomas Rothacher, Du hast gesagt, die Schweiz kann in der Top-Liga mitspielen. Was müssen wir tun, um unsere Position hinter den USA und China halten können?

Dr. Thomas Rothacher: In der Forschung und in Bereichen der Innovation sind wir auf Platz 3. Bei der Umsetzung hapert es allerdings ein wenig. Wie schaffen wir es also gerade unter erschwerten Bedingungen, wenn die Grenzen im Konfliktfall beispielsweise nicht mehr so offen sind, wichtige Produkte in der Schweiz herzustellen? Das ist, meiner Meinung nach, die grössere Herausforderung.

Dr. Peter Friedli: Lustigerweise wurden gerade heute zwei Studien von armasuisse publiziert, wie der Technologiestandort Schweiz gefördert werden sollte. Ein wesentlicher Punkt darin war Transparenz zu schaffen. Ich denke, da gibt es seitens armasuisse oder Swiss Innovation Forces die Bestrebungen, das Ökosystem der Startups in der Schweiz näher an die Rüstungsindustrie heranzubringen und dort die Widerstände zu brechen. V.a. der Nutzen für die Armee und Verteidigung soll deutlich werden.

Patrick Fontana: Als ranghöchster Offizier bei Microsoft Schweiz darf ich dazu auch Stellung nehmen. Das Defense Ecosystem funktioniert. Die Frage ist immer, wie regional muss eine Aktion erfolgen. Vertraue ich beispielsweise einem amerikanischen Unternehmen, wenn ich in einem Konflikt bin oder muss alles innerhalb der Landesgrenzen geschehen? Wir stehen aktiv im Austausch mit den amerikanischen Kollegen bei DARPA (Defense Advanced Research Projects Agency), setzen aber auch entsprechende Use Cases hier in der Schweiz um und unterstützen dadurch wiederum das hiesige Ökosystem.

Freddy Müller: Wie können wir im Bereich Polizei, Justiz und Nachrichtendienste, KI und generative KI optimal nutzen?

Dr. Thomas Rothacher: Dazu gibt es bereits einige Beispiele. Wir setzen KI beispielsweise im Bereich der Stabsübungen ein. Wie füttere ich dabei das System mit den eigenen Daten? Die Szenarienentwicklung lassen wir mittlerweile durch KI durchführen, ebenso die Datenanalyse bei fusionierten Lagebildern. Bei der Text- und Bildanalyse kommt heute ebenfalls KI zum Einsatz. Die ersten Analysen werden heute schon oft durch KI gemacht und die KI hilft auf diese Weise den Verantwortlichen, Entscheidungen zu treffen. Der Mensch ist also noch immer im Loop drin.

Kritisches Hinterfragen ersetzt das Wissen

Nathalie Gratzer (Publikumsfrage): Besteht die Gefahr, dass wir durch künstliche Intelligenz verblöden? Wie es vorhin erwähnt wurde, erweitert KI meine Kompetenzen zu gewissen Teilen und liefert mir schnell Lösungen, ohne dass ich selbst nachdenken, logisch überlegen und

mein Gehirn anstrengen muss. Hätte ChatGPT bereits zu meiner Schulzeit existiert, hätte ich wahrscheinlich die wenigsten Aufgaben selbst erledigt und auch nur die Hälfte davon gelernt.

Thomas Rothacher: Das Verhältnis hat sich vollkommen geändert und wir müssen heute mit gewissen Dingen anders umgehen. Ich habe in der Schule Wissen gepaukt. Das hat für mich einen Leistungswert. Heute ist es nicht mehr schwierig, an Wissen zu gelangen. Die Schwierigkeit ist, dieses Wissen zu bewerten und plausibilisieren? Ich habe keine Ahnung, wie ich meinen Kindern beibringen soll, wie sie Wahrheiten von Unwahrheiten unterscheiden können. Mit dem Thema umzugehen, ist eine Herausforderung.

Stefan Preuss: Ich sehe den Punkt absolut, die Fähigkeit, Dinge kritisch zu hinterfragen, gewinnt an Wichtigkeit. Wenn wir ein medizinisches Assistenzsystem betrachten, ist ein erfahrener Arzt in der Lage, zu beurteilen, ob dessen Aussagen sinnvoll und korrekt sind. Das kritische Hinterfragen ist eine wichtige Kompetenz von uns Menschen. Die Frage ist, können wir diese Kompetenz erhalten?

Bei mir sieht es im Alltag so aus. Ich habe drei verschiedene Chatbots offen und stelle allen drei Bots die gleiche Frage. Anschliessend vergleiche ich die Antworten und arbeite mit derjenigen weiter, die mir am plausibelsten erscheint. Ich versuche die Antworten kritisch zu hinterfragen und verlange immer auch die Quellen und Links, damit ich die Informationen verifizieren kann. Das ist eine Methodenfrage, eine Methodenkompetenz.

Freddy Müller: Jennifer Scurrell, wie stark verlässt Du Dich auf KI und generative KI?

Jennifer Scurrell: Ich wende sie in meiner Forschung nicht an und nehme eine mögliche Antwort, die mir ChatGPT auf meine Forschungsfrage gibt, auch nicht als Wahrheit an. Um nicht zu verblöden, ist es wichtig, den Umgang mit solchen Informationen zu lernen und ausgespucktes Wissen zu hinterfragen. Weiter müssen wir lernen, die richtigen Fragen zu stellen. Wenn uns das gelingt, hat das Arbeiten mit KI grosses Potenzial.

Patrick Fontana: Wir machen mittlerweile die Erfahrung, dass wir Kurse brauchen, um die entsprechenden Informationen aktiv zu bewerten. Ein Chatbot und ein KI-Assistent sind hilfreich, um Aussagen zu treffen und auch einige wenige Erkenntnisse zu erlangen. Man muss sich jedoch auch der Konsequenz dahinter bewusst sein. Das ist eine Kompetenz, die wir in Zukunft ausbauen müssen.

Schlusswort: Lisa Kontratieva

Zusammenfassend haben wir gesehen, dass KI und insbesondere generative KI sehr viel Potenzial bietet. Es gibt unendlich viele Anwendungsmöglichkeiten und nicht nur Chatbots, sondern, auch multimodale LLMs z.B. zur Bild- oder Videogenerierung. Wir können Prozesse automatisieren. Die Anwendungsmöglichkeiten sind sehr divers. Und dennoch besteht die Gefahr, dass etwas nicht funktioniert, oder wir durch sehr wahrheitsgetreue Inhalte getäuscht werden. Es besteht also ein Sicherheitsrisiko, ethisches Risiko, politisches Risiko und technologische Herausforderungen. Die Schweiz hat jetzt die Chance, eine führende Rolle im Bereich KI einzunehmen. Das können wir erreichen, wenn wir auch aktiv diese Risiken und Herausforderungen managen.