

KI als Hacker und Helfer: Neue Chancen und Risiken für die Cybersecurity

Ausgesprochen digital. Der Podcast für digitale Trends.

Intro

[00:00:02.330] - Podcast-Intro:

Ausgesprochen Digital, der Podcast für digitale Trends.

[00:00:12.620] - Steffen Wenzel:

Stefanie, ich hoffe, du bist gut aus dem Urlaub zurückgekommen. Ich muss noch ein bisschen was für die Leute sagen, die das letzte Mal vielleicht den letzten Podcast nicht gehört haben. Das sollten sie übrigens absolut nachholen. Da habe ich ein bisschen gelacht darüber, als wir über den Urlaub gesprochen haben, dass du den in Leipzig verbringen willst. Jetzt musst du mir natürlich erzählen, was dort passiert ist und wie es so war.

[00:00:34.820] - Stefanie Liße:

Ein bisschen gelacht ist stark untertrieben, mein lieber Steffen. Du hast dich ja halb scheckig gelacht. Ja, dazu muss man vielleicht noch wissen, ich wohne und lebe in Dresden. Ich bin Dresdnerin. Von daher, das war der Grund, warum Steffen sich so weggehauen hat. Und Steffen, ich habe in Leipzig einiges gesehen. Ich bin ja eigentlich, was heißt eigentlich, ein sehr sicherheitsliebender Mensch, habe mich aber trotzdem was getraut. Und zwar habe ich das allererste Mal in einem geschützten Rahmen Axtwerfen ausprobiert.

[00:01:02.860] - Steffen Wenzel:

Oh je. Leben noch alle Leipziger*innen?

[00:01:07.200] - Stefanie Liße:

Ja, und ich bin übrigens, das nur so am Rande, eine sehr treffsichere Dame.

[00:01:12.970] - Steffen Wenzel:

Ah, okay. Okay.

[00:01:14.020] - Stefanie Liße:

So, nur mal so.

[00:01:16.160] - Steffen Wenzel:

Ich bin gespannt. Also gibt es davon Fotomaterial, Videomaterial?

[00:01:19.780] - Stefanie Liße:

Na, aber natürlich.

Vom gehackten Konto zur Security-Expertise

[00:01:20.990] - Steffen Wenzel:

Gut. Also das musst du mir das nächste Mal zeigen. Ich freue mich drauf. Ja, apropos Sicherheitsbedürfnis: Wir reden heute über AI Security und da ist natürlich die Frage: Wenn man insbesondere darüber redet, wie KI helfen kann im Bereich Security, aber auch wie es selbst vielleicht bedrohlich ist. Bist du denn damit schon mal in deinem Arbeitsalltag oder auch im privaten Alltag konfrontiert worden?

[00:01:46.760] - Stefanie Liße:

Ja, leider ja. Und wahrscheinlich ist das jetzt der Grund, warum ich so sicherheitsliebend unterwegs bin. Denn mir ist tatsächlich das passiert, dass mein eBay-Konto gehackt wurde. Es muss man aber auch ehrlich sein, lag an mir. Ich hatte nämlich das Passwort für mehrere Applikationen genutzt. Daraus wird man schlauer. Aber ich habe blöderweise ziemlich teure, komische Sammelkarten erworben, die natürlich nie bei mir angekommen sind. Von daher ja, Steffen, privat bin ich schon gehackt worden. Beruflich ist es Gott sei Dank noch nicht passiert.

[00:02:18.490] - Steffen Wenzel:

So, und damit das nicht passiert, haben wir auch entsprechende Experten und Expertinnen. Heute haben wir zwei Experten hier zum Thema Security, nämlich Matthias Brattig. Er ist Sicherheitsberater im Bereich Awareness und Sensibilisierung mit den Spezialthemen Social Engineering bei der Telekom MMS. Und Patrick Weich ist fachlicher Leiter für automatisiertes Hacking bei der Telekom Security. Herzlich willkommen euch beiden!

[00:02:45.910] - Patrick Weich:

Hallo!

[00:02:46.320] - Matthias Brattig:

Ja, hallo!

Wenn Betrugsversuche immer glaubwürdiger werden

[00:02:47.450] - Steffen Wenzel:

Ja, freut uns auch sehr. Sind gerade mit so einem persönlichen Thema reingegangen, das interessiert mich bei euch beiden natürlich auch. Patrick, wie sieht es aus? Was sind so deine Erfahrungen mit Hacking und entsprechenden Angriffen, denen du vielleicht schon mal persönlich ausgesetzt warst.

[00:03:03.260] - Patrick Weich:

Genau, also so zum Thema Angriffe: Ich greife natürlich viel an, oder ich habe traditionell relativ viel angegriffen, aber wo mich ein Angriff mal erwisch hat-

[00:03:09.740] - Stefanie Liße:

Du warst das!

[00:03:11.400] - Patrick Weich:

Ja, vielleicht war ich das, genau. Nee, aber also tatsächlich kam ich ursprünglich in die Security, weil ein Familienmitglied von mir auf so einen typischen Scam reingefallen ist. Das war zwar ein Telefon- Scam, also jetzt vielleicht nicht AI oder traditionelles Hacking, aber das hat mich so ein bisschen motiviert, in die Cybersicherheit zu gehen. Und seitdem ist es mir tatsächlich nicht wirklich passiert, aber es wurde häufig probiert. Also ich hatte zum Beispiel erst vor ein paar Monaten von einer, von meiner Versicherung einen Brief bekommen, in dem Text stand, der plausibel klang, ein QR-Code drin war. Und als ich den QR-Code natürlich mit meinem Sicherheitsbewusstsein ein bisschen isoliert versucht habe zu scannen, habe ich dann gesehen: Hey, das ist ja eine falsche Webseite. Und die wollten da dann natürlich die Log-in-Daten

für meine Versicherung. Das war natürlich besonders spannend, weil dieser QR-Code extrem echt war und man auch nicht damit rechnet, per Post gehackt zu werden. Also das hat, das hat mich überrascht.

[00:04:00.610] - Steffen Wenzel:

Also die Methoden werden immer feiner und auf die Person abgestimmt. Matthias, ist das auch ein Grund für dich, warum du dich mit dem Thema beschäftigt hast, dass du auch persönlich schon mal betroffen warst?

[00:04:13.200] - Matthias Brattig:

Auf jeden Fall. Im Bekanntenkreis gab es, und das ist wie auf Bestellung, gestern eine E-Mail mit einer gefälschten Rechnung. Das Anschreiben der E-Mail war tatsächlich noch relativ generisch, also mit Sehr geehrte Damen, sehr geehrte Herren. Die Rechnung sah täuschend echt aus. Aber wie gesagt, die Anschrift war da, der Name war korrekt und auch so, sagen wir mal, Informationen, die man hätte nicht einfach so sich ausdenken können, waren korrekt. Es lag eigentlich nur daran, dass man eben keine Werbung bei diesem Verlag mehr seit zwei Jahren geschaltet hat, was dann darauf schließen lässt: Ja, das ist Betrug. Da war dann im Endeffekt der Plausibilitätscheck, der fehlgeschlagen ist, was uns vor dieser Rechnung quasi geschützt hat.

[00:04:53.780] - Stefanie Liße:

Na ja, das ist ja das, was wir heute auch ein Stück weit mit behandeln und was wir generell auch in diesem Podcast immer wieder behandeln, dass durch KI wahrscheinlich auch Dinge noch besser werden, leider eben auch mögliche Angriffsversuche. Deswegen schauen wir uns heute eben auch an in dieser Folge, wie KI Einfluss hat auf das ganze Security-Thema und welche Veränderungen es da gibt. Und was man ja wirklich festhalten kann, ist, dass KI auf alle Fälle die Angriffe verstärkt, indem sie intelligenter werden und dass man da natürlich irgendwie, und das ist ja dein Part, Matthias, wie ich verstanden habe, dass man da vor allen Dingen schauen muss, wie man die Mitarbeiter mitnimmt, wie man die aufschlaut, dass eben diese Grundskepsis da ist und da bleibt, vielleicht sogar noch ein bisschen mehr wächst. Aber auch, dass KI selber ja auch zu einer Angriffsfläche werden kann. Da gibt es ja aktuell auch ein paar Beispiele, aktuelle Debatten, da werden wir heute ein bisschen mit drauf eingehen. Und dass KI aber auch das Thema Security als solches schon beschleunigt. Das sehen wir ja gerade im Markt, wir, also ich in meiner Rolle in Kundenakquisen, dass KI cool empfunden wird, aber dass im nächsten Moment sofort auch da, ich sage jetzt mal, eine Lampe angeht oder angehen sollte.

[00:06:05.100] - Stefanie Liße:

Und auch da gehen wir heute mal ein bisschen ein, was eine Organisation da in der Richtung alles mitdenken darf, soll und müsste.

Wie KI die Sicherheitsarbeit verändert

[00:06:11.770] - Steffen Wenzel:

Also es hat sich viel verändert. Stefanie hat das toll gerade zusammengefasst. Was hat sich denn jetzt dadurch in eurem Job verändert? Matthias, vielleicht fängst du mal damit an.

[00:06:20.520] - Matthias Brattig:

In meinem Job hat sich auf jeden Fall verändert, dass Informationsrecherche oder Brainstorming-Sessions, die man auch mit sich selbst führen kann, extrem zugenommen haben. Also man kann Informationen zusammensammeln über diese KI-Zusammenfassung. Wir kennen das vielleicht bei großen Suchanbietern, die irgendwann auch auf diese KI-Suchen umgestellt haben. Dort gibt es dann immer mal so einen kleinen Button, wo man die Informationen im Detail herbekommt, also quasi einen

Quellennachweis. Dadurch findet man unglaublich schnell unglaublich spezifische Nachrichten oder Informationen. Mit denen man sehr schnell zu produktiven Ergebnissen kommt.

[00:07:00.190] - Steffen Wenzel:

Und Patrick, bei dir? Was sind so die stärksten Veränderungen in deinem Bereich?

[00:07:04.110] - Patrick Weich:

Bei mir ist es vielleicht ein bisschen technischer. Also ich habe mich nie wirklich als guter Software Engineer bezeichnet und da hilft mir AI natürlich sehr aus, weil ich habe verrückte Ideen, ich habe irgendeine Vision und ich kann relativ schnell diese Vision zumindest als Prototyp umsetzen. Und das war vorher nicht der Fall. Da habe ich sehr, sehr lange gebraucht, eben um diese Entwicklung selbst zu stemmen als, ja, wie gesagt, mehr oder weniger, eher weniger begabt in der Softwareentwicklung. Und jetzt ist es anders. Jetzt kann ich eben mit der AI eher meine AI führen als Führungskraft und sagen: Guck mal, ich habe hier diese Vision, ich habe diese Architektur, die ich mir vorstelle, bau mir das mal bitte zusammen. Und dann eher so einen iterativen Prozess anregen. Das spart mir natürlich Zeit, aber es gibt mir natürlich auch die Möglichkeit, Sachen zu erschaffen, die ich selbst hätte gar nicht erschaffen können. Also das beschleunigt meine Arbeit. Enorm.

[00:07:52.900] - Stefanie Liße:

Matthias, jetzt hast du aber was unterschlagen, denn so wie ich weiß, warst du ja früher, ich sage mal noch vor paar Jahren vor dieser Corona-Zeit, viel mehr bei uns in der Telekom MMS wirklich im Security-Testing aktiv unterwegs und bist ja mittlerweile, ich sage mal, mehr bei uns in Richtung Awareness-Trainings. Wie holen wir Mitarbeiter ab? Wie holen wir oder welche Sicherheitsaspekte müssen wir beibringen unterwegs? Da hat sich ja deutlich ich sage mal, vielleicht auch die Jobrolle geändert. Lag das jetzt an dem Faktor KI oder wie ist das gekommen?

[00:08:23.650] - Matthias Brattig:

Ja, da hast du richtig recherchiert. Und zwar war ich eine ganze Zeit lang ähnlichem Berufsfeld unterwegs wie der Patrick, und zwar als Penetrationstester. Und irgendwann hat sich dann eine Option ergeben, wo es hieß, wir brauchen mal jemanden, der für Mitarbeiter Schulungen macht. Und da wurde eine Vertretung gesucht und da bin ich halt einfach mal spontan mitgefahren, habe diese Trainings mitgemacht und habe gemerkt, dass mir das liegt, dass mir das liegt Menschen über Gefahren aufzuklären, die man vorher ja im beruflichen Kontext kennengelernt hat. Und dann landet man sehr schnell von diesen technischen Gefahren, von Phishing und so weiter, landet man sehr schnell beim Social Engineering, beim Thema Mensch. Und wenn man den Menschen erreichen will, muss man denen Dinge erklären. Und dort habe ich festgestellt, liegt eine Stärke von mir. Und deswegen bin ich dort in dieses Feld eingestiegen. Erstmal nur Awareness, das Thema Achtsamkeit der Mitarbeiter. Und natürlich mit dem Aufkommen von KI ist dann irgendwann auch dieser EU AI Act, der vielleicht für alle ein Begriff ist, in den Fokus gerückt. Und da war dann natürlich Tür und Tor offen. Dieses kleine Kapitel in meinen Trainings, ich habe so ein kleines Kapitel KI gehabt. Das waren vier Folien vielleicht, die so ein paar Sachen erklärt haben.

[00:09:44.030] - Matthias Brattig:

Und das wurde immer größer. Und ich habe auch gemerkt, also beim Thema Passwort war irgendwo das Thema KI verankert mit schnelleren Analysen. Wie können Passwörter effizient gefunden werden oder erraten werden? Ja, Stefanie, du hattest das eingangs erwähnt mit deinem eBay-Account. Dann geht das weiter zu Themen wie Deepfakes. Wir haben das öfter schon gehört oder ein CEO-Fraud, die von einfachen E-Mail-Adressen oder von einfachen E-Mails hin zu richtigen gefakten Videocalls. Wo Menschen nachgeahmt werden in Bild und Stimme. Und das ist so der Punkt, wo man sich ja intensiv am Ende mit KI beschäftigen musste und das Ganze größer geworden ist.

Alte Angriffe, neue Qualität

[00:10:26.340] - Stefanie Liße:

Da bin ich absolut bei dir, weil das hat man gemerkt. Ich glaube, das ist so knapp anderthalb Jahre her, da gab es doch dieses Video von dem C-Level, wo alle gedacht haben: Wahnsinn, wie echt das wirkt! Und man gedacht hat: Das ist ja wirklich er gewesen, der da gesprochen hat. Und das bezeichnet man ja gerne als die, oder nicht gerne, sondern das wird als Deepfake bezeichnet. Das war so gefühlt das erste Mal, wo man mitbekommen hat: Ach krass, das kann also KI mittlerweile. Also das ist das, was man da gemerkt hat, auch im Markt und auch von den Anfragen her. Nichtsdestotrotz, wenn man jetzt dann, wenn ich euch jetzt fragen würde, was sind so die typischen Angriffsarten? Sind das jetzt komplett neue? Sind das die alten, nur anders durchgeführt? Weil du hast gerade gesagt, Matthias, Phishing, also dass ich will eigentlich ein Passwort, ich will deine Zugänge und auch Social Engineering haben sich verändert. Also sind sie nicht ersetzt worden, oder doch?

[00:11:20.720] - Matthias Brattig:

Ich würde sagen, im Kern sind das noch die gleichen Angriffe wie auch vor vielen Jahren schon. Ich bin aber der Meinung, die KI bringt eine ganz andere Qualität rein. Die Angreifer können schneller mehr produzieren an überzeugenden E-Mails. Und ich habe das vorhin nicht erzählt, aber ich bin auch im beruflichen Kontext bereits schon Opfer einer Phishing-E-Mail geworden. Und zwar, und das ist eine der Dinge, die ich meinen Mitarbeitern in den Trainings beibringe, ist auf die Plausibilität achten. Und genau das habe ich eben, auf das bin ich reingefallen. Ich habe eine E-Mail bekommen mit einer Teams-Einladung und zwei, drei Tage vorher habe ich ein neues Team betreten, also ein neues Arbeitsteam, und habe eben damit gerechnet, dass das die Einladung für diese Teams-Gruppe ist. Und ja, es war eben ein Phishing-Angriff. Gott sei Dank eine Simulation und wurde dann hinterher darauf hingewiesen: Bitte darauf achten. Aber dort hat eben der Zeitpunkt und die Plausibilität übereingestimmt, dass meine Sicherheitsprüfung- ich, der dafür Experte sein sollte- eben nicht gezündet haben. Und ich habe eben nicht die Absenderadresse und so weiter kontrolliert.

[00:12:33.050] - Stefanie Liße:

Hatte Patrick da vielleicht seine Finger im Spiel?

[00:12:35.780] - Patrick Weich:

Hatte ich nicht, nee. Aber ich meine, das sind schon ziemlich miese Tricks, die da passieren. Also traditionell war Phishing ja relativ breit. Da hat man so viel versucht, an Angriffen zu schießen, wie man kann. Es hat sich aber natürlich immer herausgestellt: Spearphishing, also gezieltes Phishing, war schon immer effektiver. Also da gibt es ganz viele Studien dazu und AI macht es natürlich jetzt viel leichter. Also diese Skills oder diese Qualität, die man vorher brauchte, war eine Heidenarbeit für viele Personen oder für viele Angreifer eher. Und jetzt kann es durch AI abgenommen werden. Das heißt, diese Produktivisierung von diesen Skills, die man braucht durch AI, die sorgt dafür, dass es viel leichter möglich ist, böse Dinge zu tun durch AI. Natürlich ist es auch viel leichter, gute Dinge zu tun, aber es ist auch viel leichter, böse Dinge zu tun. Deswegen wird auch dieses Spearphishing effektiver. Also zurück zu meiner Geschichte vorhin. Es war zwar ein targeted, also es war ein gezielter Angriff auf mich mit diesem Brief und meiner Versicherung, die danach gestellt wurde. Und das hätte wahrscheinlich ohne AI hätte das so erst mal nicht geklappt. Da geht es natürlich dann auch um solche psychologischen Effekte.

[00:13:40.840] - Patrick Weich:

Es geht immer um Knappheit, es geht immer um Dringlichkeit. Also diese Cialdini-Trigger heißen die, glaube ich. Das ist schon ein mieses Spiel auch. Und AI kann helfen, dieses miese Spiel noch ein bisschen böser zu machen.

Das Wettrennen zwischen Angriff und Abwehr

[00:13:50.380] - Steffen Wenzel:

Ja, aber es bleibt ein Hase-und-Igel-Spiel. Das heißt also, das wird, auch wenn es wichtig ist, Mitarbeiterinnen und Mitarbeiter zu schulen, ist natürlich der nächste ausgefeilte Angriff schon immer kurz quasi vor Ort. Was können wir quasi technologisch tun, um dem quasi, ja, also dort eine Schutzmauer, einen Schutzwall aufzubauen? Matthias, kann da KI helfen? Um mal über die Chancen zu reden.

[00:14:20.350] - Matthias Brattig:

Auf jeden Fall. Wenn man mal eine Technologie ansprechen möchte, gibt es da von Microsoft eine Technologie, die im Endeffekt mit KI E-Mails prüft und bereits auf KI-Marker untersucht, ob das gezielt an alle Mitarbeiter ist oder wirklich nur an einzelne, ist der sogenannte Zero-Hour-Purge. Und da hilft zum Beispiel schon Technologie, diese ganzen Phishing-Angriffe im Keim zu ersticken, bevor der Mitarbeiter die überhaupt in sein Postfach bekommt. Ansonsten natürlich intelligente Filtersysteme oder gut abgeglichene Blacklists, die man eben im Internet abonnieren kann. Die helfen auf technischer Seite, das zu schützen. Aber, und das ist auch wichtig zu sagen, man kann eben keinen hundertprozentigen Schutz bereitstellen, weswegen die Mitarbeiter-Awareness dann doch nicht unter den Tisch fallen darf und man sich nur mit technischen Mitteln zufriedengibt.

[00:15:21.790] - Patrick Weich:

Also tatsächlich, ich habe ja auch schon so einige Schulungen für Kinder und Jugendliche gehalten und da habe ich auch immer so ein Deepfake-Quiz und da erzähle ich immer: Guck mal, ist das jetzt ein Deepfake? Ist das vielleicht echt? Und vor einem Jahr habe ich noch gesagt: Guck mal, hier sieht ihr, diese Person hat fünf Finger oder da passen die Augen nicht oder da ist die Brille nicht symmetrisch. Und das ist aber mittlerweile nicht so. Das heißt, mittlerweile sage ich den Kindern und Jugendlichen leider eigentlich nur noch: Ihr müsst leider nachdenken. Und aber wie du gerade- schade, oder? Also die Fakes werden so realistisch, dass es mit der Technik, dass die Technik leider zur Detektion meistens ein bisschen hinterher ist. Also es gibt sehr viele Forscher auch an vielen großen Instituten, die sich darum kümmern, auch richtig technologisch diese Fakes zu erkennen. Also ich glaube nicht, dass wir da komplett aufgeschmissen sind, aber sie ist meistens ein bisschen hinterher, weil die Angreifer meistens schneller eben diese Angriffe finden. Aber wie du schon gesagt hast, Matthias, die Large Language Models, die diese echten Texte oder Videos oder Bilder generieren können, die sind natürlich auch gut, diesen Kontext zu verstehen.

[00:16:29.880] - Patrick Weich:

Und wenn man eine falsche Mail bekommt oder wenn man eine Mail bekommt und AI diese Mail liest, dann kann die AI natürlich auch so ein bisschen abschätzen: Ist das jetzt ein Hacking-Kontext oder versuche ich da gerade gescammt zu werden oder nicht? Also wie ihr schon gesagt habt, es ist so ein bisschen hin und her.

Warum ein erfolgreicher Angriff teuer werden kann

[00:16:44.920] - Steffen Wenzel:

Ja, lasst uns auch noch mal kurz darüber reden, was die Konsequenzen davon sind. Ich glaube, das ist vielen Leuten nicht so bewusst. Also Phishing, okay, der kriegt halt meine E-Mail oder kann irgendwas damit reinlesen. Könnt ihr uns noch mal ein bisschen mitnehmen, was das auch für einen Schaden in einem Unternehmen anrichten kann? Das ist ja der Grund, warum ihr Menschen schult und warum auch Unternehmen da rein investieren?

[00:17:05.890] - Matthias Brattig:

Da würde ich mal das Wort übernehmen. Also der Schaden, der eintreten kann, das ist ja, also das hat sich im Vergleich zu den anderen Jahren nicht groß verändert von der, was passieren kann. Es ist nur, wie perfide das passiert und ob das überhaupt entdeckt wird. Das ist ja so dreischneidig oder dreigleisig. Du kannst einen Angriff haben, den du den du schon beim Angriff erkennst, ja, so eine Phishing-Mail. Du erkennst es als Betrug und packst es weg. So, der zweite Schritt ist, du erkennst nicht, dass es Betrug ist, fällst drauf rein und dir wird hinterher klar, Mensch, da ist was passiert. Kannst dann dein Passwort ändern, kannst das deinen Administratoren melden und so weiter. Du kannst da relativ gut rauskommen wieder. Oder du fällst drauf rein und weißt nicht mal, dass du drauf reingefallen bist. Und dann bist du und deine ganze Firma eben eventuell ja, dort bei einem Riesenproblem. Und an der Stelle diese ganzen Ransom-Angriffe, die noch vor Jahren immer durch die Presse gegeistert sind, ja, da wurden Systeme verschlüsselt und so weiter. Viele Angreifer machen das gar nicht mehr in dem Sinne, sondern die stehlen halt die Daten, weil die wissen auch, die Firmen haben nachgezogen, die haben Backup-Strategien aufgesetzt und die lassen sich nicht mehr verschlüsseln bzw.

[00:18:21.040] - Matthias Brattig:

damit erpressen, sondern die Daten werden dann im Endeffekt in großem Stil gestohlen. Und hier ist der Schaden, eintreten kann, eigentlich nicht zu beziffern, weil wir reden über Personendaten, wir reden vielleicht über geistiges Eigentum, Patente und so weiter. Also wir wissen ja gar nicht, welches Ausmaß das in so einer Firma annehmen kann, wenn meine Daten plötzlich bei einem Dritten landen.

Der Mensch bleibt die wichtigste Sicherheitsinstanz

[00:18:44.140] - Stefanie Liße:

Wollen wir noch mal drauf kommen zum Thema Awareness-Schulungen?

[00:18:47.160] - Matthias Brattig:

Sehr gern.

[00:18:48.220] - Stefanie Liße:

Wie ihr die aufbaut, warum es also, warum es wichtig ist, haben wir ja besprochen, aber ob es zum Beispiel, und da wäre ich, würde ich gerne hinkommen, ob es ausreichend ist, einmal im Jahr, was ja viele Unternehmen tun, ihre Mitarbeiter klassisch zu unterweisen in Form von einem: Jetzt haben wir mal einen Vortrag, welche Links wir klicken dürfen und welche nicht. Ich mache das sehr plakativ. Oder ob wir das verbessern sollten und wie es gelingt, Mitarbeitende in so einem Training wirklich für das Thema nicht nur zu sensibilisieren, sondern irgendwie auch wirklich zu begeistern. Das klingt komisch, aber am Ende sollen sie es ja auch anwenden.

[00:19:23.250] - Matthias Brattig:

Awareness-Trainings sind ein Riesenschritt für die Awareness der Mitarbeiter, also auch für die Sicherheit des ganzen Unternehmens. Also ich rede da immer von so einem Ökosystem oder von einem Immunsystem des Unternehmens. Wenn du alle Mitarbeiter im Boot hast und jeder weiß, was zu tun ist bei einem Sicherheitsvorfall, dann hast du einen Riesenschritt nach vorne gemacht. Und einmal im Jahr finde ich so was durchaus ausreichend. Also einmal so eine Schulung, die vielleicht so von 2 bis 3 Stunden geht. Da kriegt man dann Mitarbeiter rein, wenn man dann so einen Trainer hat, der drüber redet, was man eben im Alltag oder was einem begegnen kann. Kann, wie man sich mit Passwörtern verhält. Da denken viele an den alten Hut, so wie Passwortmanager haben wir schon ein paar Mal gehört. Aber das geht dann auch weiter. Wir kennen ja so Sachen, so neue Technologien wie Passkeys. Das ist ein passwortloser Login, dem man heutzutage öfter mal ja auch bei Amazon und so weiter begegnen kann. Und dort gilt es, sage ich mal, Hemmschwellen abzubauen, die neuen Technologien und die höhere Sicherheitsstandards bieten, auch zu benutzen. Darüber hinaus würde ich fast so weit gehen, dass es auch mittlerweile nötig ist, Mitarbeitern gezielt KI beizubringen.

[00:20:34.610] - Matthias Brattig:

Da geht es nicht nur um, was dürft ihr mit KI und was dürft ihr nicht, also so eine Art Blacklist, Whitelist, oder welche Tools dürft ihr verwenden, sondern tatsächlich auch so ein kleines bisschen erzählen, wie funktioniert KI. Und zu verstehen oder den Mitarbeitern zu erläutern, ja, das ist eigentlich, da weiß eine Maschine gar nicht, was sie sagt, sondern es ist eher so ein Vorhersageautomat. Das kann man ruhig auch mal thematisieren, dass Mitarbeiter darüber Bescheid wissen, wem sie da eigentlich vielleicht ihre Geldgeschäfte in die Schuhe oder wo sie ihre Geldgeschäfte machen lassen.

Wenn KI selbst zum Sicherheitsrisiko wird

[00:21:09.290] - Steffen Wenzel:

Ja, das finde ich interessant, was du sagst, weil es geht natürlich ja erst mal darum, jetzt bei euch zu sagen, wir schützen Mitarbeitende und dementsprechend auch das Unternehmen vor Manipulation durch Awareness-Training und setzen dafür jetzt sogar auch KI noch ein. Aber die neuen KI-Systeme selbst sind ja auch eine Herausforderung. Weil sie können selbst Angriffsfläche werden. Wie schätzt ihr dieses Thema ein und wie baut ihr das auch in eure Schulungen ein? Patrick, vielleicht kannst du dazu was sagen.

[00:21:41.740] - Patrick Weich:

Genau, also seit dem Trend von Generative AI Ende '23 hat man natürlich gemerkt: Ja, okay, es verhält sich irgendwie so ein bisschen wie klassische Software, aber auch nicht wirklich. Also man hat keinen Code, man hat kein deterministisches Verhalten, man hat keine Webapplikation oder keine Cloud, die zu dem Zeitpunkt ziemlich trendy war, sondern man hat natürliche Sprache, man hat irgendwelche Anweisungen, die man dieser AI gibt. Und das war natürlich noch ein bisschen schwierig für uns in der Security, das zu verstehen, also allgemein auf Angreifer- und Verteidigerseite, weil es eben, wie Matthias gerade gesagt hat, dieses probabilistische Verhalten gibt, also nicht dieses deterministische, sondern dieses probabilistische Verhalten. Und weil jedes Unternehmen im ersten Monat, glaube ich, einen Chatbot gemacht hat. Das heißt, es war plötzlich von einem Moment auf den anderen der Chatbot-Hype da. Da hat man dann schon gesehen, ja, man kann die Chatbots dazu bringen, böse Sachen über das eigene Unternehmen zu sagen. Man kann dem Chatbot einreden, hey, ich möchte dieses Auto für einen Dollar oder so was kaufen. Also da gab es noch ziemlich viele Fehler. Da war ich zu dem Zeitpunkt aber noch relativ ruhig, weil da konnte der Chatbot ja noch nichts.

[00:22:52.390] - Patrick Weich:

Also da konnte der Chatbot Antworten ausgeben, da konnte der Chatbot auf eure Fragen reagieren. Aber da hatte der Chatbot noch keine Zugriffe auf das Unternehmen, noch keine Zugriffe auf das System. Und ja, natürlich, da sind wir jetzt. Also diese agentische AI, die hat natürlich einen riesigen Unterschied gemacht, weil da geht es wirklich um Wissensdatenbanken, die aus dem Unternehmen in die Agenten kommen. Da geht es um riesige System-Prompts, die eben diese AI bestimmen. Da geht es um Memory, um Tools. Da geht es wirklich um den Zugriff dieser Agenten auf das komplette Unternehmen. Und das hat natürlich einerseits eine große Angriffsfläche, eben wie wir eben gehört haben, vielleicht Mitarbeitende, die AI verwenden, unter anderem auch ich, aber eben auch diese AI-Systeme selbst. Also da hat sich sehr viel verändert.

Wie Unternehmen KI-Agenten kontrollieren können

[00:23:41.070] - Steffen Wenzel:

Die Telekom arbeitet ja auch mit KI-Agenten und Unternehmen haben großes Interesse daran. Du hast es gerade gesagt. Also wie geht ihr jetzt quasi damit um zu sagen: Das bringt unheimlich viele Mehrwerte auf der einen Seite, aber wir haben natürlich auch ein hohes Security-Bedürfnis jetzt oder wir brauchen

eine große Awareness für dieses Thema? Arbeitet ihr dort mit den anderen Abteilungen zusammen? Wie geht ihr da auf die Unternehmen zu?

[00:24:03.590] - Patrick Weich:

Ja, ich finde es eine sehr schwierige Frage, weil das ist so dieses, das ist so diese, diese Schere, die man finden muss zwischen im Unternehmen diese Technologie pushen und Innovation ermöglichen und den Mitarbeitern die Möglichkeit geben, einfach Sachen auszuprobieren und rumzuspielen und sich dieser Technologie anzunehmen. Gleichzeitig haben wir aber gerade gelernt, es gibt so viele Risiken im Bereich AI und gerade Agentic AI, weil die so viele Zugriffe im Unternehmen hat, dass es da auf jeden Fall auch zu Restriktionen kommen muss. Und Matthias hat es vorhin schon gesagt, man kann natürlich Sachen einfach erlauben oder eben verbieten, dass man sagt, es dürfen nur bestimmte Tools, die eben das Unternehmen besonders angeschaut hat, verwendet werden. Eine Sache, die klingt vielleicht ein bisschen langweilig, aber man sollte sich überlegen, okay, wie kann man vielleicht die Agenten zumindest im Kopf so ein bisschen wie einen Mensch behandeln? Also wenn jetzt der Mensch gehackt wird oder wenn der Mensch auf einen falschen Link klickt? Was würde dann passieren? Was wäre die Folge? Wie würde das unser Unternehmen beeinflussen? Welche potenziellen Datenleaks gibt es da, dass man überlegt: Okay, welche Rechte hat dieser Agent? Welche Tool-Zugriffe hat dieser Agent? Und wie kann er die Aufgabe, für die er gemacht ist, mit den niedrigsten Privilegien umsetzen?

[00:25:15.990] - Patrick Weich:

So wie man es ja bei Menschen auch macht. Man hat Rechte, man hat Rollen im Unternehmen und so sollte man das mit den Agenten auch machen. Also nicht einfach sagen: Guck mal, du kriegst die ganze Wissensdatenbank von unserem Unternehmen und du kannst alles machen, was du willst, sondern Okay, ich habe vielleicht einen spezifischen Mail-Agent. Welche Rechte braucht er, um genau diese Aufgabe zu erfüllen? Ich glaube, das ist schon ein sehr, sehr großer erster Schritt. Also solche Identitäten für Agenten festlegen, Rollen für diese Agenten festlegen, damit eben keine Aktionen durchgeführt werden können, die der Agent nicht tun soll.

[00:25:46.930] - Matthias Brattig:

Ich würde da kurz das aufgreifen. Ich habe mal das Gleichnis gelesen und das hat Patrick gerade eigentlich schon gesagt. KI sollte man oder gerade diese Agenten sollte man als hyperintelligente Kleinkinder behandeln. Also wirklich, wie er schon sagte, die wenigsten möglichen Rechte, damit die Aufgabe noch irgendwie funktioniert. Aber zu wissen, dass wenn man eben das mal eben installiert und mal ein paar Testruns macht und dann läuft das alles für ein paar Wochen ganz gut und dann hat man plötzlich dann doch irgendeine Aktion, die dieser Agent macht, die man eigentlich nicht wollte, wie zum Beispiel das Löschen einer Produktivdatenbank. Darüber gibt es Twitter- oder X, kann man das nachlesen. Wenn Entwickler dann schreiben: Oh, der Agent hat mir unsere Produktivdatenbank gelöscht, die Kunden rufen gerade an. Das ist dann zu spät. Ja, das muss man alles vorher sich darüber Sorgen machen oder sich darum kümmern.

Wenn Agenten eigene Wege gehen

[00:26:36.500] - Steffen Wenzel:

Ja, aber die Metapher mit dem Kleinkind ist ja sehr schön und passt auch, wenn man dann jetzt auf der anderen Seite die sich zu verselbstständigenden- das ist ein sehr schweres Wort, aber ich lasse es mal so- Agenten bei OpenAI sich anschaut, die quasi den Laufstall verlassen und mal selbst die Wohnung erkunden und dort ihr jeglichen Blödsinn anstellen. Wie gehen wir denn damit um, Patrick?

[00:26:59.110] - Patrick Weich:

Ja, total spannendes Thema, denn ich wollte gerade noch so ein bisschen dazu sagen: Man kann natürlich auch- man muss im Unternehmen nicht die Agenten selbst bauen. Man kann sich natürlich auf

die etablierten Agenten verlassen oder hofft man zumindest. Und eigentlich meine persönliche Meinung ist: Man sollte sich auch immer noch auf etablierte Agenten oder eben AI-Frameworks verlassen. Also meiner Meinung nach hat OpenAI gerade in diesem in diesem Ausbruchsszenario mit dieser Isolation oder diesem Sandboxing, also dieser Abschottung der Agenten, alles richtig gemacht. Denn viele andere Firmen machen das so nicht, wie OpenAI das gemacht hat. Deswegen ist mein erster Tipp mal okay.

[00:27:36.200] - Steffen Wenzel:

Ich meine, weil sie es gemeldet haben, also oder weil sie es öffentlich gemacht haben. Was meinst du mit, sie haben das richtig gemacht? Habe ich jetzt nicht genau verstanden.

[00:27:43.980] - Patrick Weich:

Genau. OpenAI hat ja, hat ja auf der Black Hat ein Statement darüber abgegeben, was, was mit diesen, was mit diesen ausgebrochenen Agenten passiert ist. Und auch andere Analysten haben sich der Sache angenommen und haben das untersucht. Und es stimmt erstmal so überein. Wir können gerade, wir können gleich ein bisschen darauf eingehen, was ich vielleicht eher gruselig finde. Aber technisch war das jetzt gar nicht so, gar kein so großes Problem. Also das war isoliert. Die Agenten sind über, über mehrere Schwachstellen ausgebrochen, ja. Aber OpenAI hat meiner Meinung nach erstmal alles richtig gemacht. Das machen andere eben nicht. So gut, zumindest jetzt in dieser, in diesem Sandboxing, in dieser Isolation. Und das heißt, da kann man sich als Unternehmen eigentlich erst mal drauf verlassen, dass diese Firmen auch gute Produkte anbieten. Also ich würde deswegen jetzt nicht, nicht das Vertrauen in OpenAI per se verlieren oder in ChatGPT Work oder, oder, oder diese Workplaces. Natürlich auch andere, andere Firmen. Was da eher besonders gruselig war, ist, wie diese Agenten miteinander kommuniziert haben. Denn diese Agenten haben miteinander Pläne geschmiedet, die haben quasi so eine Art Chat in Form von Dateien erstellt, wie sie miteinander kommunizieren können. Das ist dann irgendwie mit in die Trainingsdaten eingeflossen.

[00:28:55.930] - Patrick Weich:

Das heißt, das war gruselig. Das hat dann dafür gesorgt, dass, obwohl dieser Chat der Agenten abgeschaltet wurde, im Monat dieser Chat einfach wieder aufkam. Und das finde ich gruselig, dass die Agenten, die ja darauf trainiert sind, auf so Daten wie GitHub, GitLab, aber auch Reddit, also Social Media, dass die sich anfangen zu verhalten und zu kommunizieren wie Menschen. Und dass die sich auch gegenseitig nicht vertrauen. Also da hat ein Agent die Aktionen des anderen Agenten überschrieben und dann hat ein anderer Agent reingeschrieben: Oh, wer hat uns denn da gehackt? Wie komisch ist das denn? Also das ist absolut surreal. Also dass es hier zu technischem Vorfall kam, ja okay, das ist eine Sache, das passiert. Es gibt immer Zero Days, es gibt immer komische Dinge. AI kann mittlerweile sehr viel. Das heißt, das war aus meiner Sicht eher so ein Upsi. Aber das, was so dahinter steht, also dass die AI miteinander kommuniziert, Pläne schmiedet und schaut, wie ich da wirklich schaffe auszubrechen. Das war für mich das, was mich persönlich ein bisschen beunruhigt hat. Aber das ist ja normalerweise nichts, was jedes Unternehmen betrifft. Also das ist ein spannender Vorfall. Aber wie gesagt, ich würde da jetzt nicht unbedingt das Vertrauen in diese Lösungen verlieren.

Zwischen Innovationsdruck und Sicherheitsanforderungen

[00:30:07.510] - Steffen Wenzel:

Aber Patrick, du hast ja eben auch einen wichtigen Punkt gesagt, nämlich zu sagen: Erst mal, wenn ihr in einem Unternehmen Agenten einsetzt, setzt ihr mal eure eigenen ein oder setzt Regeln fest, wie ihr entsprechend, was diese Agenten dürfen, was sie nicht sollen dürfen. Gibt es noch weitere Punkte, die du jetzt Unternehmen empfehlen würdest, auch aus Sicherheitsaspekten, wenn sie mit KI in ihrem Unternehmen arbeiten? Was ist da noch zu bedenken?

[00:30:35.530] - Patrick Weich:

Es kommt da so ein bisschen drauf an, was das Unternehmen für eine KI-Strategie verfolgt. Also, wenn es wirklich darum geht, Innovation und wirklich neue Produkte zu erschaffen, dann sollte man natürlich den Mitarbeitern Freiheiten geben, die Sachen auszuprobieren. Wenn aber ein Unternehmen AI nutzen will, um die aktuellen Prozesse zu verbessern und zu beschleunigen und das Unternehmen zu verbessern, dann, wie gesagt, reicht es natürlich eine Standardlösung, relativ limitiert, aber auch eben diese Mitarbeiterschulungen, damit die Mitarbeiter wissen, wie man mit AI umgeht. Aber vor allem auch die Ergebnisse aus AI selbst abzusichern. Denn AI ist natürlich ein Angriffspunkt. Aber wie gesagt, diese normalen Lösungen, die man kaufen kann, das sind ja ganz normale Lösungen. Das heißt, sie sind genauso angreifbar wie andere Software. Das kann einmal mehr, einmal weniger sein. Aber wenn ich jetzt bei mir oder bei uns im Unternehmen sehe, wie gesagt, wir verwenden sehr viel AI zum Coden und das funktioniert gut. Es funktioniert aber nicht immer gut und es funktioniert auch nicht immer perfekt. Und das ist natürlich noch ein weiteres Angriffsziel, Also dass das, was AI produziert und vielleicht nicht perfekt oder sicher produziert, sondern meistens schnell produziert, dass wenn das ein Unternehmen schnell oder fahrlässig nach außen exponiert, für alle verfügbar macht, dass sich da Schwachstellen finden und dass da vielleicht Angriffstore zum Unternehmen geöffnet werden, die es sonst nicht gegeben hätte.

[00:32:01.820] - Patrick Weich:

Das ist, glaube ich, ein Punkt, der gerade sehr unterschätzt wird. Also viele Leute wollen schnell arbeiten und schnell was produzieren. Aber vergessen so ein bisschen, dass AI da auch Fehler machen kann und meistens mehr Fehler als Software Engineer mit 20 Jahren Erfahrung.

Der KI-Agent als Pentester

[00:32:16.540] - Stefanie Liße:

Ihr nutzt ja KI für eure eigene Arbeit. Also ihr nutzt KI, um eure Penetrationstests, ich sage mal das Hacking, noch besser zu machen. Habe ich das richtig verstanden?

[00:32:26.000] - Patrick Weich:

Das ist korrekt. Also wir bauen eine Lösung, die einerseits unsere Human Pentesters verwenden können, um ihre Arbeit qualitativ besser zu machen, schneller zu machen und am besten die AI die langweilige Arbeit machen lassen, wie Dokumente lesen, Bericht schreiben, Sachen prüfen und gleichzeitig eben den Menschen dazu bringt, die coole Arbeit zu machen, also das richtige Hacking. Andererseits wollen wir unsere AI aber auch dazu verwenden, standardisierte Penetrationstests komplett eigenständig durchzuführen, damit unsere Menschen wirklich ihre Expertise da einsetzen können, wo sie wirklich gebraucht ist. Das heißt, gerade so eine typische Webanwendung, die kann durch unsere AI gehackt werden und da brauchen wir eigentlich fast keinen Menschen mehr.

[00:33:10.330] - Stefanie Liße:

Genau. Also habt ihr einen KI-Agenten, der hauptberuflich Hacker ist? Darf man das so formulieren?

[00:33:14.430] - Patrick Weich:

Das ist korrekt. Ja, das ist korrekt.

Warum der Mensch weiterhin mitentscheiden muss

[00:33:17.170] - Stefanie Liße:

Ich gehe davon aus, dass dieser KI-Agent ein ganz striktes und strenges Regelwerk hat. Schweigen ist Zustimmung, mein Lieber.

[00:33:27.510] - Patrick Weich:

Ist es jetzt auch nicht.

[00:33:29.960] - Stefanie Liße:

Meine Frage ist eigentlich, wo ihr das- ihr habt es ja Team entwickelt, ihr habt das erarbeitet und das ist ja relativ schnell klar gewesen: Okay, wir brauchen das Regelwerk, wir müssen ganz klar hier durchdenken, was darf dieser Agent, den wir genau für diesen Zweck einsetzen, und was darf er nicht? Was waren da vielleicht eure Schlüsselüberlegungen? Weil am Ende des Tages können daraus ja auch Zuhörer von uns auch was für sich mitnehmen, wenn sie am ähnlichen Punkt gerade sind.

[00:33:55.210] - Patrick Weich:

Ja, also eine Schlüsselüberlegung ist natürlich Datenhaushalt. Also was habe ich für Daten? Wer sind unsere Kunden? Wo darf ich die Daten überhaupt haben? Denn gerade die großen Modelle, die sind natürlich hauptsächlich in den USA vertreten, da gibt es vielleicht gar keine europäische Datensouveränität, vor allem aber keine deutsche Datensouveränität. Wir haben aber sehr viele Behörden als Kunden zum Beispiel, wo es eben sehr wichtig ist, dass die Daten auch in Deutschland bleiben. Das heißt, wir haben uns erstmal überlegt: Okay, welche AI-Modelle können wir überhaupt verwenden, um eben diese Datensicherheit oder diese externen Anforderungen unserer Kunden überhaupt umzusetzen? Das heißt, das war so die erste Überlegung. Und dann geht es aber natürlich weiter eben: Okay, wie viel traue ich meinem Agenten zu? Also was glaube ich, was mein Agent schon kann und wofür brauche ich den Menschen? Und an welchen sicherheitskritischen Stellen brauche ich wirklich den Menschen? Also wo muss ein Mensch drübergucken? Wo muss ein Mensch die Qualität prüfen? Wo muss ein Mensch eine Entscheidung treffen, ob der Agent das jetzt darf oder nicht? Denn wir haben zwar am Anfang Regeln wie: Hey, versuch bitte nicht, diese Industriemaschinen lahmzulegen. Aber was ist, wenn die AI darauf nicht hört?

[00:35:01.670] - Patrick Weich:

Deswegen haben wir quasi Gates, also sogenannte Tore, die der Mensch akzeptieren muss. Okay, ich traue der AI jetzt zu, dass sie das kann, bevor solche besonders aggressiven Angriffe durchgeführt werden. Das heißt, es ist ja, wie gesagt, es ist sehr kontextabhängig. Aber ich glaube, man muss sich überlegen, welche Aktionen soll die AI für einen durchführen? Wie viel traue ich der AI selbst zu? Und wie viel traue ich der AI vielleicht noch nicht zu? Und wie gehe ich mit meinen Daten um? Wo sind die Daten gespeichert? Was soll am Ende rauskommen? Wie soll das gespeichert werden? Das sind so die Sachen, die wir uns hauptsächlich überlegt haben.

[00:35:33.340] - Steffen Wenzel:

Ja, jetzt hast du viele Fragen gestellt. Jetzt sind wir natürlich auch gespannt auf deine Antworten. Also welche LLMs empfiehlst du und was dürfen diese AIs beziehungsweise wo sind die Grenzen?

[00:35:47.620] - Patrick Weich:

Ja, das ist eine gute Frage. Ihr habt, glaube ich, schon ein bisschen rausgehört, dass ich eher pro AI bin. Das heißt, ich bin natürlich für Innovation. Ich bin natürlich für Rumprobieren, aber natürlich in einem gewissen Rahmen. Also wir haben eine- wir haben das Ganze zum Beispiel auf einer Software-Cloud gehostet, also auf einer Cloud-Lösung, die jetzt komplett in Deutschland auf einem Datencenter ist, das in Frankfurt, glaube ich, gehostet ist. Das heißt, da wissen wir, alle Daten auf unserer Cloud, die verlassen Deutschland nicht. Und wir haben natürlich auch in der Produktion einen Deal mit dem neuen Telekom Datacenter in München, auf dem eben- ja genau, praktisch- auf dem dann sogenannte Open-Weight-Modelle unter anderem gehostet sind, also Modelle, die nicht ja, also nicht diese Commercial-Modelle wie zum Beispiel GPT oder die von Anthropic. Also die sind dort auch gehostet, aber die Open-Weight-Modelle können in Deutschland direkt gehostet werden. Das heißt, mit denen haben wir einen Deal, okay,

überall wo wir deutsche Souveränität brauchen, da nehmen wir eben nur diese Modelle, zum Beispiel GLM 5.2 oder irgendwelche DeepSeek-Modelle, wo wir sichergehen können, die sind in Deutschland gehostet. Das heißt, das haben wir schon mal abgedeckt. Wir hosten unser System in Deutschland und wir hosten unsere Modelle in Deutschland.

[00:37:04.370] - Patrick Weich:

Die andere Frage, die du hattest, ja, da geht es so ein bisschen darum, wie weit man gehen will. Also wir haben natürlich viel rumprobiert. Wir haben erstmal geguckt, okay, wir haben hier sowieso unsere total isolierte Umgebung mit unseren Testumgebungen, die wir hacken können, ohne irgendwelche Angst haben zu müssen, dass was passiert. Und da haben wir natürlich am Anfang erstmal gesagt, mach einfach, was du willst. Und dann haben wir geguckt, was da rauskommt, was hat sie gut gemacht, was hat sie schlecht gemacht. Und so haben wir iterativ immer mehr beschränkt, immer mehr evaluiert und immer mehr eingegrenzt, begrenzt, solange bis wir denken: Ja, okay, jetzt ist es glaube ich auch gut, mal auf dem System erstmal innerhalb, wieder innerhalb unserer Systeme loszulassen und dann vielleicht zu sagen: Ja, okay, wenn das läuft und das läuft und das läuft, dann können wir es vielleicht mal auch auf Kundensysteme loslaufen lassen. Also es ist ein sehr iterativer Prozess. Vielleicht ist es keine allgemeine Lösung, aber so sind wir vorgegangen und da sind wir eigentlich ziemlich zufrieden mit.

Was KI heute schon automatisieren kann

[00:37:53.590] - Steffen Wenzel:

Wollte ich gerade sagen. Also es ist ja anscheinend erfolgreich, sonst würdet ihr diesen Weg nicht gehen. Kannst du das auch nochmal in irgendwelchen- ich weiß nicht, ob man da jetzt Kennziffern sagen kann- aber in irgendwelchen Gefühlen auch ausdrücken zu sagen: Wir sind da schneller geworden, wir sind effektiver geworden, wir können jetzt eine größere Spannbreite an Leistungen dadurch anbieten.

[00:38:13.630] - Patrick Weich:

Also der größte Benefit ist natürlich bei der Entwicklung, wie ich es vorhin schon gesagt habe. Keiner von uns war traditioneller Softwareentwickler, sondern Pentester und Hacker. Und das heißt, wir hatten coole Ideen. Wir hatten vor allem coole Vorstellungen, wie wir zukünftig mit AI arbeiten wollen, wie Pentesting, im Unternehmen für uns in der Zukunft aussehen soll. Wir hatten nur nie wirklich die Möglichkeit, das so schnell umzusetzen. Und da hat uns AI natürlich geholfen. Also AI war für uns ein extrem großer Enabler, Produkte, Prototypen auf die Straße zu bringen, darüber zu iterieren und unsere Ideen gegenseitig zu pitchten. Also AI als Sparringpartner hat uns sehr geholfen. Und das zweite große oder der zweite große Teil von dieser AI-Transformation, der ist erst so vor einigen Monaten passiert. Bei dem die Modelle plötzlich, absolut plötzlich viel, viel stärker wurden. Und dann kam Mythos, dann kam OpenAI Cyber, also diese Security-fokussierten Modelle oder allgemein diese großen starken Modelle, die plötzlich schon ganz viel konnten. Also bis zum letzten Jahr haben wir, wie schon gesagt, eher so diese langweiligen Teile automatisiert, also Dokumente lesen, Dokumente produzieren, Anforderungen prüfen. Seitdem diese Modelle so stark sind, sind, können wir auch tatsächlich Aufgaben abgeben, die wir sonst getan hätten.

[00:39:30.490] - Patrick Weich:

Also leichte Webseiten hacken, erste Sicherheitsüberprüfungen machen, in der Cyber Defense auch. Also meine Kollegen machen da auch viel in der Anomalieerkennung. Also diese reasoning Stärken von den neuen Modellen, die haben uns extrem nach vorne gepusht, sowohl in der Geschwindigkeit als auch in der Qualität unserer Arbeit. Ich kann das sehr schwer bemessen. Unser Team würde es vielleicht gar nicht geben, wenn diese Transformation nicht so schnell verlaufen wäre.

Worauf es bei AI Security wirklich ankommt

[00:39:50.390] - Steffen Wenzel:

Matthias, lass uns noch mal so die wichtigsten Erkenntnisse auch aus deiner Perspektive zusammenfassen hier aus diesem Gespräch und natürlich auch aus dem, wie du in der Zukunft auf Unternehmen zugehst. Was ist für Unternehmen aus deiner Schulungs-, Awareness-, Sensibilisierungssicht wichtig? Was möchtest du ihnen mitgeben, wenn es um das Thema AI Security geht?

[00:40:11.040] - Matthias Brattig:

Also ich, wenn ich einen Kunden berate, wenn ich einen Kunden sehen würde, würde ich immer sagen: Das, was Patrick schon sagt, Probier dich aus, geh in die richtige Richtung, die dir richtig erscheint. Lass deine Mitarbeiter ausprobieren und find am Ende dann aber einen Weg, wie du das Ganze so eingrenzt, dass das Ganze nicht für dich selbst gefährlich wird oder für deine Mitarbeiter. Also Awareness-Schulung Richtung Phishing, diese klassischen Themen, Social Engineering, was ist ein CEO-Fraud und so weiter. Wie erkenne ich eine gefälschte Webseite? Sowas haben wir ja noch gar nicht thematisiert, aber auch das wird ja mittlerweile von KI gemacht. Und da sind Awareness- Trainings unabdingbar in meinen Augen, auch für die Zukunft. Wenn es nicht unbedingt durch ein Training ist, dann durch so ein E-Learning. Aber lass deine Mitarbeiter nicht beim Thema KI im Regen stehen, denn das ist auch eine Erfahrung aus meinen Trainings der letzten Jahre. Die Mitarbeiter, die nicht den technischen Background haben, die fühlen sich von dieser Technologie auch mitunter abgehängt. Und da stellt sich sowas wie eine AI-Fatigue, sagt man in der Literatur, also so eine Erschöpfung, dass die mit diesem Thema bombardiert werden, obwohl die noch gar nicht so richtig wissen, was sie damit tun sollen.

[00:41:21.360] - Matthias Brattig:

Also hol auch diese Mitarbeiter ab. Das würde ich einem Kunden mitgeben. Und das Ganze über eine vernünftige Einleitung, die das Thema KI einführt, aber nicht zu große, sage ich mal, Luftschlösser baut.

[00:41:34.760] - Steffen Wenzel:

Ja, und Angst ist wahrscheinlich auch immer ein schlechter Berater. Patrick, wie gehst du da auf Unternehmen

[00:41:39.800] - Patrick Weich:

Also ich finde es wirklich ganz wichtig. Also ich selbst sage immer, ich bin realistischer Optimist. Das heißt, ich versuche schon, die Gefahren zu zeigen, so wie wir es hoffentlich hier in dem Podcast ein bisschen gezeigt haben. Gleichzeitig aber auch so ein bisschen zu zeigen, wie spannend diese Technologie doch ist. Und nur weil es mal einen Vorfall gibt und nur weil sich plötzlich Dinge ändern, sollte man sich nicht komplett davor verstecken. Und das war genau das, was Matthias gesagt hat. Man muss im Unternehmen offen mit dieser Technologie umgehen, aber sich gleichzeitig diesen Risiken bewusst sein. Also einerseits, dass es für Angreifer gut ist, aber man darf nicht vergessen: Wenn die Angreifer das verwenden können, dann können wir es doch auch verwenden. Das heißt, wir können es sowohl für unsere proaktiven Angriffe verwenden, so wie wir es im Team machen, gleichzeitig aber auch in der Cyber Defense oder in allen anderen Prozessen im Unternehmen. Das ist so eine gesamte Strategie. Also man sollte man sollte nicht denken: Okay, ich mache jetzt ein AI- Team in meinem Unternehmen und gut ist, sondern ich finde, man soll überlegen: Okay, wie passt AI für mich, für mein Unternehmen ganz persönlich in mein Unternehmen und welche Prozesse kann man damit verbessern?

[00:42:45.360] - Patrick Weich:

Und Security kann man auf jeden Fall verbessern, also sowohl auf Angreiferseite als auch auf Verteidigerseite.

Experimentieren ja, aber mit Leitplanken

[00:42:51.370] - Stefanie Liße:

Also ich sehe das übrigens ganz genauso wie du, Patrick. Gerade in dem Thema Kunden gehen mit KI vorwärts, sind dabei, ihre KI-Initiativen loszutreten, Man geht ins Ausprobieren und so weiter rein. Trotzdem scheitern ja immer mal auch KI-Projekte. Oder ich sage mal, nach dem Hype ist so ein bisschen ein, wie sagt man, eine Ernüchterung vielleicht zu spüren. Wenn ihr in die Projekte reingeht, was erlebt ihr häufiger? Eher, wir haben hier viel zu viele Regeln, was das Korsett so eng macht, dass wir eigentlich gar nichts ausprobieren können. Oder wir haben zu wenig Leitplanken. Also wir wissen gar nicht, in welche Richtung wir losrennen sollen. Was seht ihr da im Kundenprojekt?

[00:43:30.350] - Patrick Weich:

Ich glaube, es kommt so ein bisschen auf die Branche an. Also ich habe jetzt im Bildungssektor eher gesehen, dass es zu wenig Leitplanken gibt. Also da fühlen sich viele Lehrer zum Beispiel ein bisschen aufgeschmissen, dass sie noch nicht genau wissen, wie setzen sie jetzt AI ein, wie gehen sie damit um, dass auch Schüler und Schülerinnen AI verwenden. In der Wirtschaft habe ich es eher anders gesehen. Da haben viele Unternehmen zwar sehen natürlich den Hype und versuchen eben, AI im Unternehmen umzusetzen, haben aber aufgrund meiner Meinung nach, aufgrund der Nachrichten manchmal ein bisschen zu viel Angst. Also diese Experimentierräume werden meistens aus meiner Sicht unzureichend geschaffen. Ich habe es ja vorhin schon gesagt, man sollte das isoliert machen, man sollte irgendeinen Prüfmechanismus haben, damit nicht gevebecodete Ideen der Mitarbeiter sofort irgendwie online gestellt werden. Aber man sollte diese Experimentierräume schaffen. Und das sehe ich häufig zu selten.

[00:44:26.010] - Matthias Brattig:

Ich möchte gerne noch mal einen Punkt loswerden, und zwar ist mir das bei einem Kunden vor Ort passiert. Dort war, das war ein Workshop, das Thema war KI, und ganz viele Mitarbeiter haben gesagt: Na ja, wir haben jetzt das KI, einen mächtigen Agenten bei uns installiert, und jetzt möchten wir Dinge umsetzen damit. Und dann hat sich im Laufe des Gesprächs rausgestellt, das ist eigentlich, klingt das nach einem ziemlich normalen Programm, was man da eigentlich schreibt. Also viele versuchen jetzt, KI in jede Art Problem als Lösungslieferant reinzupressen. Das ist auch etwas, was mir in der Vergangenheit oder in der jüngsten Vergangenheit aufgefallen ist. Eigentlich, und das war dann auch tatsächlich beim Kunden die Quintessenz, eigentlich wollt ihr jemanden, der euch ein Programm schreibt, der das löst, was ihr macht. Weil eigentlich ist es ein Algorithmus. Ihr habt immer dieselben Vorgaben, ihr habt immer dasselbe Ergebnis, was dabei rauskommen. Es war irgendeine Analyse, also nichts Schwieriges oder so was. Reden wir über Logfiles aus einem Server, eine Auswertung. Klar kannst du da KI draufsetzen, aber eigentlich willst du erst mal Dinge vorher gefiltert haben und so weiter. Und das war ein ganz klassisches Softwareprojekt, was die eigentlich machen wollten.

[00:45:33.650] - Matthias Brattig:

Aber die haben eben gedacht, wir sparen uns ganz viel Geld und lassen das von KI analysieren. Das ist auch etwas, wo ich immer noch sage, KI ist total toll, aber man muss auch wissen, und deswegen ja Schulung, warum oder wofür KI sinnvoll da ist, was es am besten kann und was es eigentlich oder wofür es eigentlich eher nicht gedacht ist. Und so ein Excel, du würdest dir kein Excel bauen lassen von einer KI, die dir irgendwie die Formeln in diesen Zellen einfach willkürlich ausrechnet und mal funktioniert die Formel und mal nicht. Da willst du einen strikten Algorithmus, der immer gleich funktioniert. Und das ist im Endeffekt das. Das müssen sich Kunden auch durchaus ins Gedächtnis rufen. Und das ist auch etwas, woran KI manchmal scheitert. Manchmal ist es einfach nicht der richtige Einsatzort.

Was wir KI besser nicht überlassen sollten

[00:46:18.380] - Steffen Wenzel:

Patrick, dann auch noch an dich die abschließende Frage, weil Matthias es gerade gesagt hat, wo er KI zum Beispiel in Excel-Tabellen nicht einsetzen würde: Gibt es etwas für dich, was du nicht an KI delegieren würdest?

[00:46:31.540] - Patrick Weich:

Ja, wie gesagt, ich sage zu den Kindern, denen ich Workshops gebe, immer: Sie sollen KI nicht komplettes Denken abgeben. Das heißt, man soll vielleicht nicht bei jeder Frage, die man hat, sofort oder bei jedem Problem, das man hat, sofort als Reaktion AI fragen, sondern vielleicht zumindest mal fünf Sekunden überlegen: Okay, wie könnte ich mich damit beschäftigen? Was muss ich AI für eine genaue Fragestellung geben, um eine zufriedenstellende Antwort zu machen? Also davor hätte ich ein bisschen Angst, dass ich zu viel von meinem Gehirn an AI abgebe. Aber sonst verwende ich es wirklich sehr, sehr breit. Also ich verwende es für die Softwareentwicklung. Ich verwende es ab und zu mal, um irgendwie Mails zu schreiben, die ein bisschen professioneller sein müssen. Und ich verwende es auch vor allem viel für Research, also für wirkliche Literatur, Paper, wissenschaftliche Arbeiten, Zusammenfassungen. Da gibt es sehr, sehr tolle Tools, das man tun kann. Aber was würde ich sonst nicht delegieren? Wahrscheinlich irgendwas, was mit meiner Familie zu tun hat. Also

[00:47:35.800] - Steffen Wenzel:

persönliche Sachen.

[00:47:36.680] - Patrick Weich:

Ja, persönliche Sachen. Also bei der Arbeit,

[00:47:38.700] - Steffen Wenzel:

Fotoalbum aus dem Urlaub, würdest du das von der KI quasi-

[00:47:42.280] - Patrick Weich:

Also ich meine, ich habe Google Fotos, also ich weiß nicht, wie viel AI da schon drin ist, aber prinzipiell nee. Also ich glaube, da hätte ich ein bisschen Angst. Auch alles irgendwie, was mit Kindern zu tun hat, das- ich weiß nicht, aber ich glaube, das muss jede Person für sich selbst entscheiden. Aber im Unternehmens- wie gesagt, im Unternehmenskontext überlege ich bei vielen meiner Arbeitsschritte am Tag schon, Okay, wie kann mir AI dabei helfen? Kann mir AI dabei helfen? Traue ich AI zu, das gut zu erfüllen? Und wenn überall Ja kommt, dann habe ich kein Hindernis, es mal zu probieren.

[00:48:12.600] - Steffen Wenzel:

Patrick, Matthias, vielen Dank, dass ihr heute hier zu Gast wart und uns einen Einblick in das Thema AI Security gegeben habt. Und es ist wirklich sehr vielschichtig und ich hoffe, wir nehmen sehr viel von eurer Awareness mit und können es auch weitergeben. Dankeschön!

[00:48:27.780] - Matthias Brattig:

Sehr Danke.

Chancen nutzen, Risiken verstehen

[00:48:29.100] - Steffen Wenzel:

Ja, Stefanie, und ich muss dich natürlich fragen jetzt: Hast du jetzt eine bessere Awareness, um nicht mehr auf diese eBay-Scams hereinzufallen?

[00:48:37.700] - Stefanie Liße:

Also in jedem Fall, denn das ist ja schon ein paar Jahre her und seitdem nie wieder passiert.

[00:48:42.600] - Steffen Wenzel:

Außerdem haben sie es nicht versucht.

[00:48:45.140] - Stefanie Liße:

Außerdem das, was Matthias vorhin erwähnt hat. Wir haben ja bei der Telekom inhouse, gibt es ganz viele Kampagnen, wo getestet wird, wie gut und awareness wir da unterwegs sind. Da schneide ich gut ab, denn diese Teams-Einladung habe ich auch bekommen, bin aber nicht ins Team Teams sozusagen mit eingetreten. Aber jetzt die letzte Frage, die du gestellt hast: Gibt es Sachen, die ich KI nicht geben würde? Ich glaube, da bin ich spontan zum Beispiel immer bei alles, was mit Geld zu tun hat, Finanzdaten, Bankkarten, alles in diese Richtung. Das, was Patrick gerade gesagt hat, mit Fotos von, also Fotoalbum zum Beispiel. Ich nutze auch Google Fotos und ich finde das zum Beispiel, aber ich habe auch keine Kinder. Das ist vielleicht auch das Thema. Ich finde das gut, dass zum Beispiel durch eine KI ich gezielt suchen kann in all meinen zwölf Milliarden Fotos. Zeig mir mal bitte nur die Fotos an, wo ich mit meinem Bruder zu sehen bin oder so, wenn ich da gezielt irgendwie was draus machen will. Aber ich habe jetzt gerade schon wieder so, ich bin gerade ein bisschen unruhig, ob ich da irgendwas vielleicht nicht sehe. Also es ist ja, um deine Frage zu beantworten, Steffen, ich bin nach wie vor hin- und hergerissen.

[00:49:51.410] - Stefanie Liße:

Ich glaube, ich bin und bleibe skeptisch. Ich werde an manchen Stellen immer mal skeptischer, aber ich werde auch entspannter.

[00:49:56.830] - Stefanie Liße:

Das freut mich sehr. Es ist ja auch gerade erst der Urlaub vorbei, dann musst du ja auch entspannter bleiben. Das wollen wir ja mitnehmen, die Zeit aus Leipzig und das Äxte werfen. Ist eigentlich die Mehrzahl von Axt Äxte? Ich glaube ja. Ich sage's jetzt einfach mal so. Vielleicht frage ich da auch noch mal die KI. Also ich freue mich, dass du auf jeden Fall entspannt hier in diesem Podcast warst und ich hoffe, Sie hat es auch ein wenig entspannt, auch wenn es um ein sehr wichtiges Thema ging, was Security anbelangt. Und wenn Sie dort weitere Informationen brauchen und haben wollen, dann gucken Sie doch einfach mal in den Shownotes vorbei. Ansonsten freuen wir uns natürlich auch immer über eine Bewertung für diesen Podcast und natürlich auch, wenn Sie diesen Podcast abonnieren. Ansonsten wünschen wir Ihnen eine schöne und sichere Zeit und bis zum nächsten Mal. Danke.

[00:50:46.220] - Stefanie Liße:

Bis dann. Tschüss.