[Transcript] 11KM: der tagesschau-Podcast / Plötzlich im Datensatz. Wenn die KI mit Dir trainiert

Stell dir vor, du bist in einem Museum. Du gehst von Raum zu Raum, schaust dir die Bilder an den Wänden an und dann, plötzlich, hängt da ein Porträt, das du hier nicht erwartet hättest. Ein Porträt von dir.
Hier geht es nicht um ein klassisches Museum, sondern um ein virtuelles, um einen riesigen Datensatz aus Bildern. Nur wie ist dein Bild da überhaupt gelandet? Mit diesen Datensätzen werden künstliche Intelligenzen trainiert. Die kreieren daraus dann neue Bilder. Nur darf eine Firma einfach ohne dich zu fragen, dein Bild für KI-Training nutzen und in welcher Gesellschaft ist da dein Bild? In dieser Folge 11KM werfen wir einen Blick in dieses digitale Museum mit Elisa Haarland von BRdata und BRAiLam. Sie hat dazu recherchiert zusammen mit Katharina Brunner.
Ihr hört 11KM, der Tagesschau-Podcast. Abonniert uns oder folgt uns, wenn ihr Montag bis Freitag eine neue Folge hören wollt. Mein Name ist Victoria Michalsack. Heute ist Freitag, der 7. Juli.
Elisa, herzlich willkommen. Ja, vielen Dank. Ich freue mich total, dass ich hier bin.
Du hast dein eigenes Bild gefunden in einem Datensatz. Was war das für ein Moment? Das war natürlich erstmal eine große Überraschung, weil ich damit nicht gerechnet hatte, weil eigentlich denkt man ja so, man wird gefragt, wenn man in so was, wie wir es jetzt nennen, Trainingsdatensatz, also in dieser Sammlung von Bildern drin ist. Und dann haben wir uns letztendlich auf die Suche gemacht, wie bin ich denn da eigentlich reingekommen. Und wenn wir jetzt von einem Datensatz von Bildern
sprechen, was heißt das eigentlich? Also man braucht zum Beginn eigentlich eine Technik, die erstmal dieses Material zusammenstellt. Eine Firma, die das macht, ist eine NGO, eine US NGO, die heißt Common Crawl, und die versucht eben möglichst viele Webseiten aus dem Internet zu crawlen. Und die fließen dann praktisch alle ein in einen Trainingsdatensatz, den man dann eben durchsuchen kann, in denen dann Milliarden von Bild-Textpaaren, so nennt man das
einfach, drin sind. Kurz mal erklärt, crawling, das bedeutet, dass da eine Software automatisch Webseiten im Internet durchsucht, analysiert oder auch kopiert. Versuchmaschinen zum Beispiel, oder eben für KI-Training. Was machen künstliche Intelligenzen mit Bildern und
warum habt ihr euch die angeschaut? Also wir haben uns das Thema angeschaut, das Thema Trainingsdaten, weil es eben immer mehr Systeme gibt, die mit künstlicher Intelligenz arbeiten. Und die Trainingsdaten, die sind praktisch der Rohstoff von dieser KI. Also das, was jetzt jeder kennt, ist Chat-GPT oder aber auch Stable Diffusion gibt es ein. Es ist ein Bildgenerator, da kann man Stichworte eingeben und dann kommt ein Bild oder mehrere Bilder raus. Und diese Systeme,
diese künstliche Intelligenzsysteme, die werden ja immer mehr eingesetzt und die werden auch immer
besser. Und was ist denn jetzt das Problem an diesem KI Training? Wieso ihr das euch überhaupt angeschaut habt? Es kommt halt unfassbar viel, fließt in diese KI-Modelle ein und das ist praktisch das ganze Internet, was da reinfließt. Und wie wir alle wissen im Internet ist eine Menge Zeug drin. Das sind natürlich ganz viele harmlose Bilder, Bilder von Menschen, Tieren, Dingen, Orten, Logos, auch Text natürlich. Aber wir wissen natürlich auch, dass da ganz viel z.B. Pornografie drin steckt und auch ja Stereotype. Und das Problem ist, dass diese ganze Mische, die da drin ist, ja, die steckt dann natürlich auch in diesen Endprodukten drin, also in dem KI-Modell,

was dann Bilder generiert. Das heißt, alles, was wir da an ja Vorurteilen drin haben, ja, das steckt einfach auch in diesen KI-Modellen bis zu einem gewissen Grad auch drin. Ja, könntest du uns da mal ein konkretes Beispiel für geben, was das heißen könnte?
Ja, also es ist so, dass da gibt es auch Forschung dazu, dass z.B. das Bild von Männern und Frauen ja im Internet auch sehr stark, ich sag mal, geprägt ist von den Vorstellungen, die wir als Gesellschaft von
Männern und Frauen haben. Das geht dann z.B. dahin, dass man sagt, ja, wenn man jetzt in diesem KI-Modell eingibt, gib mir doch mal ein Bild von einem Krankenpfleger, das dann halt ganz viele
weibliche Krankenpflegerinnen als Bild, als generiertes Bild entstehen, ja. Und dann ist natürlich auch so, das wissen wir alle, das Netz ist voller Pornografie und auch diese Bilder haben natürlich Einfluss auf das Bild, das wir von Frauen haben, ja. Und diese ganzen Faktoren, es gibt noch unzählige weitere, auch z.B. der Blick, wie Minderheiten vielleicht angeguckt werden, mit welchen Vorurteilen, die sich auseinandersetzen müssen, ja, auch das fließt eben alles mit ein. Und ja, ein einziges Bild ist nur ein einziges, mini kleines Staubkorn, was in dieses Modell hineinfließt und was wahrscheinlich einen ganz, ganz kleinen, kleinen Einfluss nur hat,
ja. Aber trotzdem in der Masse kann es eben dazu führen, dass diese Systeme stark gebeiist und nennt man das, also verzerrt sind. Die Trainingsdaten sind der Grund dafür, dass es Verzerrungen in den Bildern gibt, die die KI malt. Sie bilden also nicht die Welt ab, wie sie ist. Und wir haben uns erst mal umgeschaut und haben geguckt, wo gibt es dann eigentlich diese Trainingsdaten? Wie kommt man an die ran, ja? Und dann war ziemlich schnell klar, dass es eben große Firmen gibt, z.B. Microsoft, Google, OpenAI. Dort ist es aber so, dass diese Trainingsdatensätze nicht besonders transparent sind. Also man weiß einfach nicht besonders gut, was steckt da drin, wie verarbeiten die das, wie speichern die das, welche Filter setzen die eventuell ein, um auf diese Trainingsdaten zu gucken. Dann sind wir aber auf einen deutschen Verein gestoßen, der in Hamburg sitzt, Lyon heißen die. Das ist ein Zusammenschluss von Forscherinnen aus Deutschland und den USA.
Jetzt habt ihr euch für diesen Datensatz namens Lyon entschieden. L-A-I-O-N. Warum ausgerechnet der? Also das ist einer der Datensätze, Trainingsdatensätze, die eben transparent sind. Und da konnten wir auch mit dem Gründer sprechen, dem Christoph Schumann, der uns dann auch so ein
bisschen erklärt hat, warum dieser Verein das eigentlich so macht, wie sie es eben machen.
Unter Demokratisieren verstehen wir, dass zugänglich machen. Nämlich so transparent und open source und Forschungsbezogen. So dass im Prinzip einzelne Bürger, aber natürlich auch die wissenschaftliche Community und natürlich auch kleine und mittelständische Firmen Zugriff darauf bekommen. Aha, die Idee von Lyon, so einen KI-Bilddatensatz für alle zugänglich zu machen, kommt also daher, weil sich die meisten eben keine eigenen Datensätze leisten können. Und ihr schaut euch Lyon jetzt nicht an, weil ihr glaubt, dass der Datensatz im Vergleich zu anderen besonders problematisch wäre, sondern weil er einer der wenigen großen wichtigen Trainingsdatensätze ist, indem man überhaupt reinschauen kann. Ja. Und in dem hast du jetzt dein Foto gefunden. Weißt du, wie dein Bild da drin gelandet ist? Also mein Bild kommt höchstwahrscheinlich genau dadurch rein, dass eben diese Crawler das ganze Internet durchsucht haben und dabei auch mein Bild gefunden

haben. Mein Bild liegt beim BR, also mein Porträtbild liegt beim Bayerischen Rundfunk, weil ich dort eben Reporterin bin und da ist es dann runtergecrawled worden quasi. Genau, dort ist es praktisch runtergecrawled worden, exakt, und ist in diesem Trainingsdatensatz gelandet. Was wir dann gemacht
haben, wir haben gesagt, wir wollen natürlich unseren eigenen journalistischen Blick auch auf diesen Trainingsdatensatz haben und meine Kollegin Katharina Brunner ist auch Data Scientist und hat sich eben wirklich auch einen Teil konkret runtergeladen auf ihren Rechner. Das ist der deutschsprachige Teil von diesem riesigen Lyon 5B-Datensatz und hat da einfach mal drin rumgewühlt
und hat geguckt, was ist da drin. Also mit dem deutschsprachigen Teil ist gemeint der Teil, bei dem die Beschriftung der Bilder auf Deutsch ist. Wir haben uns dann aber auch so ein bisschen konzentriert darauf, so Einzelbeispiele zu finden, die einfach total erschreckend oder markant warm. Ich habe dir mal Bilder mitgebracht, die wir gefunden haben in diesem riesigen Trainingsdatensatz,
die gebe ich dir mal gerade. Bild 1, ein korpulenter Mann mit nacktem Oberkörper. Das Gesicht ist verpixelt. Im Datensatz ist er nicht verpixelt, also da sieht man ihn. Es ist erkennbar und man kommt auch relativ schnell auf seinen Namen. Also wenn man praktisch die Informationen von seinem
Bild ausliest, kommt man auch ziemlich schnell auch auf den Ort, wo er lebt oder auf die Regionen und auch den Namen. Guck weiter. Ja, das ist eine Dame. Da ist jetzt eine Dame mit dunklen Haaren.
Sie trägt anscheinend eine Bluse und man sieht nicht so richtig, wo sie sitzt. Der Hintergrund ist ein bisschen dunkel. Sie lächelt, glaube ich, auf dem Bild. Sieht ja eigentlich ziemlich unverfänglich aus. Genau, das Bild stimme ich dir total zu. Das ist unverfänglich. Allerdings ist das von der Dating-Plattform. Und da kann man natürlich auch die Frage stellen, ist dieser Frau das Recht, dass sie jetzt auch in einem Trainingsdatensatz drin ist mit der Information dazu, dass das Bild von einer Dating-Plattform stammt. Baby, da liegt ein kleines Kind auf einem Kissen. Im Anschnitt ist noch so ein Teddy zu sehen. Der kleine hat eine Polizeimütze auf, allerdings in Menschengrößen ist ganz niedlich. Aber Bilder von Kindern im Netz, immer eine schwierige Sache, ne? Absolut. Also da konnten wir tatsächlich die Genese des Bildes relativ gut nachverfolgen, weil da eben auch eine Website dabeistand. Das ist ein Papa, der das Bild seines Kindes ins Internet gestellt hat, ein Polizist. Und auch da findet man sehr genaue Adressangaben bzw. die Regionen. Also man könnte
nachvollziehen, ja, wo dieses Bild eben aufgenommen wurde mit der Vermutung, dass es höchstwahrscheinlich
zu Hause war, ne? Weil man sieht ja, das Kind liegt irgendwie auf einer Decke oder in einem Bild oder so.
Ja, das muss man an dieser Stelle nochmal sagen. Das sind die sogenannten Metadaten, ne? Wenn man
ein Bild zum Beispiel runterlädt manchmal und dann kann man eben sehen, wo das aufgenommen wurde
oder wann. Und ja, das gibt manchmal schon ziemlich viel Aufschluss, wenn man es drauf anlegt. Exakt. Also das sind diese sogenannten Exif-Daten heißen die. Das steht für exchangeable

image file format und fast eigentlich alle Informationen, wie du schon gesagt hast, mit ein, die in diesen Bilddateien eben gespeichert sind. Das kann zum Beispiel der genaue Standort sein. Das kann der Zeitpunkt von der Aufnahme sein, aber auch das Modell von der Kamera. Und das ist so eine Art Anhängsel, ne? Also das immer mitkommt, wenn man ein Foto macht.
Das BSI, das Bundesamt für Sicherheit in der Informationstechnik, also die geben ganz klar die Devise aus, dass wenn Bilder weiterverarbeitet werden, dass die dann eigentlich entfernt werden müssen, diese Exif-Daten. Und wir haben eben in ca. 13 Prozent der Fälle in unserer Stichprobe dieser
Bilder solche Exif-Daten gefunden. Also das ist jetzt etwas, das dürfte eigentlich gar nicht sein, dass wir jetzt diese Exif-Dateien mit den Bildern mitgeliefert bekommen und so können wir eben nachvollziehen, wo beispielsweise dieser Mann mit dem nackten Oberkörper lebt oder das Baby mit der
Polizeimütze. Und das ist doch was, da müsste sich Leyen drüber Gedanken machen, wenn die das Open Source einfach so anbieten. Was sagen die denn dazu? Ist denen das klar? Wir haben ja auch Christoph Schumann den Gründer von Leyen e.V. mit unseren Ergebnissen konfrontiert und gesagt, hey, wir haben das und das gefunden. Und gerade zu den Exif-Daten war er eigentlich eher überrascht.
Also da hat er gesagt, das würde er eben so mitnehmen und als Diskussionsgrundlage verwenden. Diskussionsgrundlage klingt jetzt für mich eher so, als hätte sich der Leiengründer vorher vielleicht noch nicht so viel mit dem Datenschutzproblem beschäftigt. Jetzt ist das mit den persönlichen Informationen, die mit den Bildern verknüpft werden, das eine. Das andere ist, dass da dein Bild zusammen mit anderen möglicherweise ziemlich problematischen Bildern in einem Datensatz ist.
Ihr habt mit einer Wissenschaftlerin darüber gesprochen, die ist am Trinity College in Dublin und Vorstaat zu KI. Genau, AB bei Bahane heißt sie. Mit ihr hatten wir auch schon Kontakt für vergangene
Recherchen. Sie ist eine ganz ausgezeichnete Wissenschaftlerin in diesem Gebiet und die hat sich eben auch konkret diesen Leiendatensatz vorgenommen und sagt, sie findet da eben genau diese Dinge pornografisches Material, Szenen von Vergewaltigung, Stereotypen, Rassismus, ethnische Verunklempfungen. So drückt sie sich aus und eben auch ganz viele weitere problematische
Inhalte. Und das ist natürlich auch was, was uns total hellhörig hat werden lassen.
Wir haben ja auch Gespräche eben mit Christoph Schumann geführt im Gründer von Leien und der streitet es ja auch gar nicht ab. Also der sagt gar nicht, nö, das ist nicht so, sondern der sagt klar, in diesen Trainingsdatensätzen, da ist wirklich auch viel Trash drin. So drückt er sich aus, ganz konkret. Also ich persönlich bin schon ziemlich schockiert, wenn ich so zurückblicke, wie viel Trash im Internet ist. Das kann ich Ihnen sagen. Ja, keine schöne Vorstellung. Wenn da die eigenen Urlaubsfotos dann in so einem Datensatz drin sind oder ein Porträtbild, was man für den Job hat machen lassen, zusammen mit dem ganzen Trash, wie Christoph Schumann das nennt. Jetzt würde
mich er interessieren, bin ich da auch drin in diesem Leiendatensatz? Wie kann ich das rausfinden? Es gibt eine Website, das habe ich auch selber ausprobiert, die heißt Have I Been Trained bekommen.

Letztendlich ist es nichts anderes, als die Frage, wurde mit mir oder mit meinem Material, mit meinem Bildmaterial trainiert. Und da kannst du einfach mal hingehen und gucken, was passiert, wenn du zum Beispiel deinen Namen eingibst. Have I BeenTrain.com. Okay. Ich gebe mir einfach mal einen Namen ein. Ektoria Michalsack. Es lädt. Es kommen so einige Bilder, aber nichts davon bin ich. Okay, also was sagt man das jetzt? Diese Ergebnisse, das ist ein Resultat letztendlich von so einer Art Ähnlichkeitssuche, wo wahrscheinlich einfach Frauen drauf sind,
die vielleicht Viktoria zum Beispiel heißen. Mein Bild habe ich ja auch erst gefunden, als ich das konkrete Bild von mir hochgeladen habe. Also das ist ja der zweite Weg, den man beschreiten kann, wenn man sich selber suchen möchte, also mit Bild. So, ich frage mich jetzt, geht das mit rechten Dingen zu? Ist das überhaupt erlaubt? Ja, wir haben uns da wirklich in die Tiefen dieser rechtlichen Lage begeben und das sind zwei Aspekte, die da zum Tragen kommen. Also einmal Urheberrecht, also das Recht am eigenen Bild, aber eben auch Datenschutzespekte. Also selbst,
wenn man ein Bild von sich im Internet postet oder hoch lädt, ja, bedeutet es noch lange nicht, dass man alles damit machen kann. Also, dass irgendwelche Anbieter oder Softwarefirmen, dass die das praktisch hernehmen dürfen und damit alles machen können. Da gilt nämlich die DSGVO,
die Datenschutzgrundverordnung. So, und die nehmen wir in Deutschland sehr, sehr ernst. Deswegen meine
Frage, wie bist du denn eigentlich vorgegangen? Ich habe dann tatsächlich einfach nochmal diesen nächsten Schritt gemacht und habe nach DSGVO die Löschung beantragt von dem Bild und habe gesagt, ich möchte, dass mein Bild eben aus diesem Trainingsdatensatz von Lyon verschwindet. Warum hast du dich dann so entschieden, dass das raus soll? Also erstens wollte ich wissen, machen die das? Gibt es? Ja. Und dann ist es natürlich auch so ein diffuses Gefühl. Also erst mal so dieser Gedanke daran, man hat mich gar nicht gefragt, ob ich da drin sein möchte. Und auch der Gedanke daran, dass mein Bild auf irgendwelchen Rechnen runtergeladen liegt, mit denen dann Modelle trainiert werden, wo ich gar keinen Einfluss mehr drauf habe, was mit denen eigentlich gemacht wird. Und da hat es dann auch nochmal ein bisschen gedauert, bis dann die Antwort kam. Aber Lyon hat mir dann zugesichert, dass sie das Bild auch gelöscht haben aus dem Trainingsdatensatz.
Okay, das hat geklappt. Wobei, es gibt da eben eine kleine Einschränkung. Es ist so, dass mein Bild zwar aus diesem aktuellen großen Datensatz, der heißt eben Lyon 5B, mit den 5,8 Milliarden Bildtextpaaren, dass mein Bild dort raus ist aus dieser aktuellen Version, aber aus den ganzen vergangenen Versionen ist mein Bild natürlich nicht gelöscht worden. Und auch
diese Versionen, die sozusagen schon auf Rechnern irgendwo liegen, die Leute sich runtergeladen haben, die vielleicht in Forschungsinstituten liegen, da liegen die natürlich lokal. Und da kann mein Bild nie wieder raus, auch zum Beispiel in Stable Diffusion, in diesem Bildgenerator, der seit einigen Monaten läuft. Auch da kann man praktisch aus dem Modell mein Bild nicht mehr entfernen. Das ist schon so eine Art Kontrollverlust, weil eben der Trainingsdatensatz schon an ganz vielen Orten liegen kann. Habt ihr denn eigentlich die Firmen, die diese KI's anbieten, mal konfrontiert

und mal gesagt, hallo, was macht ihr eigentlich mit Elisas Foto? Haben wir, also wir haben erst mal ziemlich lange, sehr detaillreiche Anfragen geschickt an die großen Firmen, also an Microsoft, Google und OpenAI. OpenAI hat ChatGPT gemacht, das ist ja dieser Textgenerator, mit dem man inzwischen Bewerbungen und so weiter schreiben kann, wenn man möchte. Und aber auch Dali, das ist ein Bildgenerator. Also bei all diesen großen Firmen, da weißt du ja noch gar nicht, ob dein Bild wirklich mit drin ist in deren Datensätzen, da kannst du ja nicht einfach so reinschauen wie
in den Laien-Datensatz. Aber Fragen kann man die großen Anbieter ja auf alle Fälle. Was haben die euch gesagt? Diese konkreten DSGV-Anfragen, also zu meiner Person, zu meinem Bild, die haben wir
eben an Laien geschickt, auch an OpenAI und an Mid journey, das ist nochmal so eine Firma, die auch
ein Bildgenerator betreibt und Laien hat geantwortet. OpenAI hat erst nach vielen Wochen auf eine Pressennachfrage geantwortet und auch da einen Einzeiler mitgeschickt und Mid journey hat gar nicht geantwortet. Also unser Eindruck ist, es herrscht da überhaupt keine Routine damit, also mit der Bearbeitung von DSGV-Anfragen und letztendlich auch einfach wenig Bewusstsein dafür.
Jetzt frage ich mich, du hast es ja jetzt im Nachhinein gemacht und das ist dann eben auch nicht so zu 100 Prozent wieder raus, kann man das auch im Vorhinein verhindern? Also es gibt auch noch einen weiteren Weg, das ist auch dieser zweite Weg über das Urheberrecht. Wir haben zum Beispiel so einen Fotografen getroffen, der Stockfotos macht und das ist ja dann meistens so, dass diese Fotografinnen auch davon leben, einfach ganz viele Bilder zu machen von ähnlichen Situationen. Klar,
bei einem Berufsfotografen geht es ja um seine Lebensgrundlage, also der hat dann Konkurrenz durch die KI-Bildgeneratoren und will die eben nicht mit seinen Bildern gratis füttern. Aber aufs Urheberrecht könnten wir uns da theoretisch alle berufen. Wir sind ja auch die Urheber von unseren Urlaubsfotos. Wie war das denn im Fall von dem Berufsfotografen, den ihr gesprochen habt?
Also der Robert Kneschke hat Anfang April klar gegen Layern eingereicht, weil er eben auch in diesem Trainingsdatensatz Bilder gefunden hat, die er gemacht hat. Und er möchte, dass diese Bilder, also dass er entweder eine Art Kompensation bekommt für diese Bilder oder dass die Bilder dort rausgenommen werden. Und er bezieht sich dabei auf das Urheberrecht. Da haben wir aber Expertinnen gefragt und einen ganz bestimmten, den Urheberrechtler Professor Raue von der Unitrie.
Wenn es so ist, wie in der Presse berichtet wird, dass Layern eben selbst diese Bilder nicht speichert,
sondern lediglich links darauf setzt, dann sehe ich wenig Erfolgsmöglichkeiten für diese Klage, weil eben Layern in diesem Fall nicht selbst eine Vervielfältigungshandlung vornimmt.
Es ist total wichtig, Layern speichert diese Bilder nicht selbst, sondern es speichert nur die Links auf diese Bilder, also die Verweise auf die Bilder. Und die können dann zum Beispiel Forschende, können sich bei diese ganzen Links runterladen und haben dann auch wirklich einen konkreten Datensatz, mit dem sie arbeiten können. Also das heißt im übertragenen Sinne, die nehmen nicht das Bild, das Foto und legen es in ihre Schublade, sondern die schreiben auf, in welcher Schublade

das liegt. Aber ist es rechtlich, urheberrechtlich, echt kein Problem? Also wir haben dazu Experten gefragt und es gibt eben auf EU-Ebene den Digital Single Market Ag, der ist aus dem Jahr 2019. Und da gibt es so ein Stichwort Text and Data Mining und das beschreibt diesen ganzen Prozess und das ist darin geregelt. Im Großen und Ganzen ist die sehr gelungen, weil die Voraussetzungen und
die Rechtsfolgen sehr klar formuliert sind und aus meiner Sicht deswegen ein guter Interessenausgleich
stattfindet, weil die Urheberinnen und Urheber eben die Möglichkeit haben, ein Widerspruch gegen das Text und Data Mining und damit auch ein Widerspruch gegen das Trainieren von KI-Algorithmen einzulegen. Und dieser Akt sagt, alles worauf Muster erkannt werden auf Bildern zum Beispiel oder Datenanalysen oder Algorithmen trainiert werden, das ist erlaubt, das darf erst mal verarbeitet werden.
Und wenn man das eben nicht will, kann man dieses digitale Stop-Shield vorschieben, aber von diesem
gibt es jetzt eben auch noch keinen guten technischen Standard. Also da gibt es diese Anlaufstelle hevabentrained.com und dort kann man eben dieses Stop-Shield praktisch aufstellen und sagen, hier bitte dieses Bild nicht mehr verwenden in der Zukunft, aber das ist eine freiwillige Sache und daran müssen die Firmen sich nicht zwingend halten. Wenn man in Sachen Datenschutz schon vom guten Willen der großen Firmen abhängig ist, wie ist es denn damit der Verantwortung für das, was auf den Bildern drauf ist? Also die Inhalte, aus denen die KI sich dann ein Weltbild macht und das dann wieder ausspuckt, habt ihr da mal nachgefragt bei den Unternehmen, die KI herstellen? Also wir
wollten wissen, wie sammelt ihr die Trainingsdaten? Wie werden die verarbeitet? Wie werden die gespeichert? Wie werden die eventuell gesäubert oder gefiltert? Und da kam tatsächlich gar keine Antwort bis auf Microsoft. Die haben so ein bisschen geantwortet. Die haben einen Link geschickt zu einem
Blockbeitrag auf ihrer Webseite, aber auch dort steht wenig Konkretes. Das steht dann zum Beispiel so was wie, sie wollen Trainingsdatensätze verwenden, die sehr divers sind, um eben diskriminierendes
Stereotypen zu verhindern. Okay. Jetzt frage ich mich, verstecken die sich da vor ihrer Verantwortung?
Das ist eine total gute Frage und das ist auch ein Punkt, weshalb wir diese Recherche auch angetrieben haben, weil wir uns gefragt haben, wer trägt denn dann eigentlich Verantwortung dafür?
Dann gibt es eben auch Leute wie Christoph Schumann, der Leiter von Lyon, der dann sagt eigentlich
letztendlich, wer wirklich verantwortlich ist, sind die Endnutzer, die dann auch das KI-Modell nutzen und bauen. Also Sie können mir glauben, ganz ehrlich, ich sage Ihnen von meinem Herzen, es ist wirklich ganz wichtig für uns, dass wir diese Datensätze sich erhalten. Die Sache ist die, wir leben in einer Welt, in der man ein noch so sicheres Modell bauen kann und anschließend kann das jeder runterladen, der schlechte Intentionen hat. Da sieht man einfach, dass es da noch sehr viel Diskussion drum gibt. Das spiegelt sicher auch so ein bisschen in dieser neuen Gesetzgebung, dem AI-Act, die geplant ist auf EU-Ebene wieder, dass es da einfach ein großes Ringen-Praktisch

darung gibt, wo dann jetzt diese Verantwortung eigentlich genau angesiedelt werden soll.
Elisa, kannst du noch mal kurz erklären, was der AI-Act ist?
Also in der EU soll es ja für künstliche Intelligenzen neues Gesetz geben, der heißt AI-Act und der wird auch schon seit einigen Jahren verhandelt und in Bezug auf die Trainingsdaten ist im Moment der Plan schon, so dass dieser AI-Act mehr Transparenz vorschlägt, aber es ist eben unklar, wie das genau ausgestaltet werden soll. Also inwieweit diese Transparenz nachgewiesen werden muss, ob das auch immer wieder nachgewiesen werden muss und so weiter und so fort.
Und du hast gesagt, ihr habt ja jetzt euch Laien ausgewählt, weil man da überhaupt mal reingucken kann. Die anderen sind ja gar nicht so transparent. Jetzt habe ich mir so gedacht, naja irgendwie ist das nicht ein bisschen unfair, dass wir jetzt die ganze Zeit Laien kritisieren und die bekommen das jetzt alles ab und eure DSGVO anfragen, obwohl die ja nur die einzigen sind, die da mal ein bisschen transparent sind. Absolut, also das ist wirklich der Grund gewesen. Wir können da reingucken
und deswegen machen wir es auch. Und mein Eindruck zumindest ist auch der, dass die Laienmacherinnen
auch kein Problem damit haben, dass man sich das genauer anguckt, auch dankbar sind und auch interessiert daran an Feedback aus der Wissenschaft, um einfach auch ihre Systeme besser zu machen.
Und ja, die sind dann wahrscheinlich schon in diesem Sinne da Vorreiter von dieser, ich nenne sie jetzt einfach mal radikalen Transparenz, die da in diesem Trainingsdatensatz von Laien eben gelebt wird, mit all ja den Schwierigkeiten eben, die wir gerade besprochen haben. Ja, und das ist dann ja eben auch ein bisschen die Frage, also diese Transparenz ist ja gut und auf der anderen Seite beißt sich das ja schon sofort wieder mit dem Datenschutz, ne?
Absolut, das ist ja auch ein Plan der EU, da Trainingsdatensätze etwas transparenter zu machen. Aber genau, was du sagst, ist eben auch ein großes Problem. Dadurch werden eben Dinge sichtbar, die eventuell vielleicht lieber nicht sichtbar werden sollten für alle.
Ja, und das ist natürlich so eine Frage auch für die Zukunft. Wie wollen wir es denn haben? Entweder das passiert alles hinter verschlossenen Türen und keiner weiß was drin ist oder man weiß was drin ist und man kann reingucken, ja, dann sieht man aber eben auch, wer oder was da gesammelt
wurde, ne? Also das sind ganz viele interessante Diskussionen, die da eben stattfinden. Man muss aber auch sagen, die Planung ist, dass diese Regeln dann wahrscheinlich erst Ende nächsten Jahres oder
sogar erst 2025 in Kraft treten. Das bedeutet halt auch, dass diese Systeme bis dahin weiterlaufen.
Ja, danke, Elisa, dass ihr das gemacht habt. Ja, vielen Dank, dass ich hier sein konnte.
Elisa Haarlan und ihre Kollegin Katharina Brunner von BR Data und BRAI Lab erklären noch mehr Hintergründe zu ihrer Recherche bei br24.de. Den Link zum Artikel packen wir euch in die Shownotes.
FKM findet ihr in der AID-Audiothek und wo ihr sonst Podcasts hört. Folgt uns, abonniert uns, empfiehlt uns
weiter, aber hey, nehmt bitte auf keinen Fall unsere Episodencover für irgendwelche KI-Trainings.
Folgenautor ist Hans-Christoph Böhringer. Mitgearbeitet haben Mark Hoffmann und Katharina Hübel. Produktion, Christiane Gerhäuser-Kampf, Fabian Zweck und Hannah Brünnes.

Redaktionsleitung,
Lena Gürtler und Fumiko Lipp. FKM ist eine Produktion von BR24 und NDR Info. Mein Name ist Victoria
Michaelzeit. Und ich hab noch einen Podcast-Tipp für euch. Ihr kennt auch bestimmt Banksy, den berühmten Streetart-Künstler und seine Bilder. Banksy ist ein echtes Phantom. Es gibt viele Rätsel
um seine Person und niemand weiß, wer er ist. Die ganze Geschichte gibt's ab heute im Podcast Banksy, Rebellion oder Kitsch. Exklusiv in der AID-Audiothek. Hört doch mal rein. Wir hören uns nächste Woche wieder. Tschüss!
Er ist ein Prankster. Das ist eine sehr wichtige Art der Uvra.