A1

Großartig, genau. Und bei mir funktioniert der Ausschlag auch, das sieht alles sehr gut aus. Ja, dann können wir ins Interview starten, genau. Also Hintergrund bei dir hast du ja gerade schon gesagt, du studierst, du arbeitest nebenbei noch, wenn ich das richtig verstanden habe, oder machst du ein Praktikum?

I4

Momentan Pflichtpraktikum.

A1

Okay, in welchem Bereich?

I4

IT-Sicherheit. Also es ist ein Monitoring-Unternehmen und momentan mache ich Threat Intelligence.

A1

Okay, alles klar. Beschreib doch einmal so deinen Kontakt zu KI bisher oder auch inwiefern du damit vielleicht schon gearbeitet hast. Wenn du nicht damit gearbeitet hast, ist auch okay, aber beschreib einmal so ein bisschen deinen Kontakt zum Thema KI.

I4

Genau, gearbeitet habe ich damit tatsächlich bisher nicht und Kontakt halt, wenn irgendwie was Interessantes rumkommt, was generelle Medien erreicht, dann das natürlich, also so Sachen, natürlich ganz bekannte Sachen sowieso, ChatGPT, keine Ahnung, die ganzen Modelle von irgendwelchen größeren Unternehmen und halt so Videos oder so was. Wenn irgendwie was Interessantes in irgendwelche Feeds von mir gespielt wird, dann gucke ich mir das auch an, also so ein bisschen auch dazu, wie das dann funktioniert.

A1

Okay. Das heißt, aber so irgendwie mal richtig mit einer Library rum experimentiert oder so hast du eigentlich noch nicht. Also du hast jetzt noch nicht so das so praktisch deployed, sondern eher so auf der Nutzer-Schnittstellen-Seite.

I4

Ja.

A1

Also ich meine so Chat-GPT und so ein Kram kann man ja einfach auch, ohne dass man programmieren kann, nutzen und ausprobieren. Das dann schon oder auch sowas noch nicht?

I4

Doch, das schon.

A1

Okay. Wir werden wahrscheinlich ein paar Fragen so ein bisschen umformulieren jetzt und dich dann eher bitten, dass du dich so ein bisschen in die Position begibst von jemandem, weil ansonsten machen die Fragen keinen Sinn. Du machst jetzt ein IT-Studium, das heißt, du machst Softwareentwicklung und ähnliches halt automatisch im Studium, lernst du das und das ist vielleicht was, was jobtechnisch irgendwann für dich perspektivisch funktioniert. Angenommen, du würdest jetzt irgendwann in einer Position sein, wo du KI in irgendeiner Form mitgestaltest, im Unternehmen mitnutzt, mitentwickelst, ähm, kannst du dir da jetzt mit dem Kontakt, den du zu KI bisher schon hattest, irgendwie vorstellen, welche, also wo sich vielleicht Probleme und Herausforderungen irgendwie stellen würden oder was so, wie solche Prozesse ablaufen würden und was vielleicht für dich jetzt so in dieser Vorstellung die größte Hemmschwelle wäre?

I4

Für mich vermutlich als IT-Sicherheitler, je nachdem, wofür es benutzt wird, IT-sicherheitliche Aspekte. Ich habe tatsächlich schon interessante Stellen gesehen dazu, wie KI benutzt wird oder wie jetzt, keine Ahnung, Chat-GPT-Geheimnisse, in Anführungsstrichen natürlich, ausspuckt. Oder sowas, wenn man es nur davon überzeugt. Und sowas ist natürlich auch je nachdem, wofür man das in so einem Unternehmenskontext anwendet. Wenn man das da, keine Ahnung, als unternehmensinternen Assistenten hat oder sowas, kann natürlich auch Leute, die dann nicht sicher sind, was sie diesem Ding preisgeben können und was nicht, kann das dann halt, können da Probleme auftreten, denke ich. Ja, das wäre, glaube ich, so das Größte.

A1

Wie ist es bei dir, also wie zugänglich fandst du bisher denn Informationen zu dem Thema? Also so ein bisschen, wenn man das in Nachrichten mitbekommt, hast du das Gefühl, man kommt da auch mit einem IT-Hintergrund schon ganz gut rein und findet da recht schnell konkrete Informationen, wie man das selber anwenden oder nutzen könnte oder so? Oder ist das für dich eher so eine völlig obskure Blackbox, wo du sagst, das ist irgendwie absolute Magie für mich?

I4

Ich denke schon, dass man tatsächlich relativ leicht, ist so mein Eindruck, in das Ganze reinkommt, dass es relativ high-level Schnittstellen gibt, mit denen man mit sowas arbeiten kann. Und wie weit man sich darin reinsteigern will, ist dann, glaube ich, relativ individuell. Also ich glaube, man findet, denke ich, ganz gut rein. Ja, viel ist, glaube ich, aber auch vielen Leuten nicht klar. Wie genau Machine Learning und so Zeugs insgesamt funktioniert, zum Beispiel wie ChatGPT überhaupt funktioniert, bin ich mir hundertprozentig sicher, dass, keine Ahnung, 80 Prozent der Leute keine Ahnung überhaupt haben.

A1

Ich glaube, das kannst du definitiv noch hochskalieren. Die Anzahl von Machine Learning Experten ist dann doch noch recht überschaubar. Klar, es werden mehr, aber ...

I4

Ja, ich meinte jetzt auch gar nicht so im Sinne von Experten, sondern allein schon, dass das Ding Sätze bildet, indem das Tokens hat und so und dann die Wahrscheinlichkeiten berechnet und da bin ich dann sogar selber schon wieder raus, ob das immer das Wahrscheinlichste nimmt oder ob das dann wirklich mit den Wahrscheinlichkeiten die Tokens nimmt, aber ja.

A1

Ich kann dir auf jeden Fall sagen, dass die Modelle, die im Large-Language-Bereich und NLP-Bereich gerade eingesetzt wird, tatsächlich gar nicht wahrscheinlichkeitsbasiert sind. Das Berechnen von Wahrscheinlichkeiten ist rechnerisch extremst aufwendig. Du müsstest halt dann wirklich Wahrscheinlichkeitsverteilungen über deine Daten irgendwie erstmal bilden. Das ist extremst aufwendig. Und der große Erfolg von neuronalen Netzen, die ja auch für Large-Language-Modelle eingesetzt werden, ist tatsächlich daran begründet, dass die einen sehr effizienten Weg gefunden haben, empirisch in den Daten Muster zu erkennen, die aber nicht auf Wahrscheinlichkeiten und irgendwelchen Base-Verteilungen oder ähnliches basieren. Das heißt, du kriegst zwar so eine Art Wahrscheinlichkeit, also du kriegst halt irgendwie mehr oder weniger wahrscheinliche Sachen, aber das sind keine mathematischen, korrekten Wahrscheinlichkeiten. Die erfüllen nicht die Bedingungen, die du mathematisch eigentlich an Wahrscheinlichkeitserwartungen stellen würdest.

I4

Sind das dann mehr so Konfidenzwerte von der KI?

A1

Eigentlich sind das halt so Softmax-Outputs. Also, aber wir gehen hier zu sehr in die technische Tiefe. Da geht es ja in dem Interview jetzt auch nicht drum. Aber ja, genau. Also diese Internal Workings, das merkt man auch auf Fachkonferenzen so immer wieder. Selbst im Gespräch mit XXXXXXXXX und so. Das ist halt auch nicht einfach, da durchzusteigen. Das erfordert ein paar Jahre Studium. Und insbesondere für Leute, die keinen IT-Hintergrund haben, ist so, da ist schon der Begriff Algorithmus so eine Magic Box irgendwie. Und dann sind ja irgendwie KIs noch deutlich komplexer als nur ein Algorithmus. Aber ja, genau. Also, das ist tatsächlich... Du hast schon recht, als du vorhin gesagt hast, man kommt da schon rein. Man kann da schon recht schnell mit anfangen, was zu machen und so. Man kann relativ schnell ein paar Zeilen Code das mal benutzen und so. Aber so richtig durchdringen ist halt nicht so einfach.

I4

Ja, und herausfinden, was ist das richtige Modell für meinen Anwendungsbereich und sowas, ist natürlich auch immer eine Frage. Und vernünftige Datensets bereitstellen ist dann auch, also, ja.

A1

Ja, ganz genau. Fällt dir so von Haus aus irgendwas ein, wo du jetzt das Gefühl hast, das wäre was, da könnten Machine Learner echt besser drin werden, wenn die das und das machen würden, dann würde der Zugang und das Verständnis für diese Technologie besser werden?

I4

… Fällt mir jetzt gerade nichts spezifisches ein

A1

Das ist völlig okay. Also wie gesagt, das ist ein offenes Gespräch einfach darüber. Keine Prüfungsfrage. Ja, also wir haben tatsächlich, habe ich ja vorhin schon so angedeutet, ich habe mich in meiner Forschung halt schon damit beschäftigt, wie man auch gerade die Entwicklung und die Eigenschaften von Machine Learning Modellen und KI Modellen irgendwie verständlicher machen kann, auch gerade für Leute, die halt irgendwie nicht diese Expertise haben in dem Feld bisher, weil wenn man da mit einem Wissenschaftler darüber redet, der sagt dann direkt, ja, gib mir das Paper oder zeig mir die Formeln und gib mir den Code, so dann verstehe ich das. Aber gerade für Leute, die diesen Hintergrund nicht haben, ist das halt viel zu komplex. Wir haben deswegen, oder ich habe deswegen und wir haben deswegen an einem System gearbeitet, was einen etwas anderen Weg geht und dazu teile ich jetzt dann tatsächlich doch nochmal meinen Bildschirm, sobald PowerPoint sich hier bequemt, die Präsi zu starten. Jawohl, perfekt. Here we go. Ihr seht meinen Bildschirm, hoffe ich. Perfekt. Genau. Und das wäre der Bereich KI Labeling als eine Möglichkeit, um diesen Transfer dieser Technologie irgendwie zu vereinfachen. Ohne, dass wir jetzt hier groß irgendwie was zu sagen, magst du uns einmal so eine spontane Reaktion auf das Label geben? Was siehst du, was sind deine Gedanken dazu, was du siehst?

I4

Das erinnert mich an Nutri-Score. Sonst kann ich mir unter den Sachen schon mal grob was vorstellen. Also PowerDraw ist ja vielleicht verständlich, wie viel Strom verbraucht das, Running Time, Inference bin ich mir jetzt gerade nicht sicher, was das ist. Er hat was mit dem Input zu tun, ich weiß nicht, ob ein Satz an Inputs vielleicht oder so, keine Ahnung. Accuracy und Robustness kann ich mir jetzt nicht so hundertprozentig was drunter vorstellen. Also Accuracy schon könnte ich mir grob was vorstellen, aber kann ich mir vorstellen, dass das schwer zu quantifizieren ist. Ja, Robustness bin ich mir nicht sicher.

A1

Eventuell schaust du dir nochmal den oberen Abschnitt vom Label an, also über diesem Score-ähnlichen, über dieser Skala.

I4

Da sind noch QR-Codes für Informationen.

A1

Sagt dir ImageNet was?

I4

Ne, das nicht.

A1

Okay, dann muss ich das an der Stelle tatsächlich einmal kurz erklären weil tatsächlich ist das MobileNetV3 was wir hier gelabelt haben, ist ein Modell für Bilderkennung, also Bildklassifikation um genau zu sein. ImageNet ist einer der größten Bilddatensätze wurde in der Challenge 2012 vorgestellt wurde in den letzten 10 Jahren rauf und runter gemined da gibt es 1000 Klassen und es gibt halt pro Klasse irgendwie 1000 Bilder und die Bilder sind relativ hochauflösend zumindest zu der Zeit waren das irgendwie 228x228 Pixel, was deutlich größer ist als sonstige Bilddatensätze die in der Praxis damals genutzt wurden, dadurch halt auch sehr große Datenmengen einfach, ImageNet sind so 160 GB an Bilddaten, wie gesagt bei 1000 Klassen, wenn man jetzt völlig random bei jedem Bild sagt, das hier ist eine Katze oder ein Hund oder was auch immer hat man also bei einer random Choice eine Accuracy von 0,1% im Schnitt, weil man immer eine von den 1000 Klassen prediktet und die Chance, dass das die richtige Klasse ist, ist ja entsprechend niedrig. Das heißt, das ist so ein bisschen die Baseline MobileNetV3 bekommt 63% Accuracy hin, wenn man es richtig trainiert und die Inferenz, wo du gerade die Frage aufgeworfen hattest damit bezeichnet man praktisch eine Modellausgabe also du schickst eine Anfrage an das Modell und kriegst eine Klassifikationsentscheidung für dieses eine Bild zurück beziehungsweise in der Praxis würde man nicht ein Bild hinschicken, sondern immer ein Batch von Bildern weil man das auf GPUs halt ganz gut parallel machen kann also man kann ganz gut für 32 oder 64 Bilder je nach Batch Size sich Predictions machen lassen und das heißt, das wäre dann auch eine Inferenz, weil es auch egal wäre, ob du jetzt ein Modell, ein Bild klassifizieren lässt gleich den ganzen Batch zusammen, weil das eh parallel erfolgt. Genau, das sind so ein bisschen die Hintergrundinformationen das heißt wir sind tatsächlich gar nicht im NLP-Bereich hier, sondern in einem KI-Bereich, der das war viele Jahre war das KI-Thema, Bilderkennung so, und irgendwie überall Kameras installieren und Menschen erkennen und tracken und in der Industrie irgendwelche Fertigungsprozesse untersuchen und Fehler erkennen und und und und… Und jetzt so in den letzten zwei Jahren zwei Jahren hat sich natürlich das Verständnis von KI dahin geschoben, dass alle halt über Sprachmodelle reden und das auch der häufigste Kontaktpunkt zum KI-Thema ist. Aber klar, Handykameras und so, Gesichtserkennung, das ist halt auch ein sehr relevantes Thema. Hast du irgendwie als direkte Reaktion, hast du das Gefühl, so was ist nützlich oder denkst du dir, keine Ahnung, wer das, also wem das jetzt nützt, weiß ich nicht. Hast du da irgendwelche Gedanken zu?

I4

Kann ich mir vorstellen, dass es so zur groben Kategorisierung, was genau, worauf möchte ich achten für meinen Einsatz von der KI, die ich entwickeln möchte, dass es dafür schon praktisch ist. Wenn ich jetzt, also dann kann ich zumindest schon mal in den vier Kategorien unten abwägen, okay, worauf davon lege ich am meisten Wert, wie sollen die Relationen stehen, brauche ich irgendwas, was auf jeden Fall akkurat ist, aber ich benutze es so wenig, dass mir Energie und Zeit egal sind, so Sachen.

A1

Ja, das ist, genau, so Entscheidungsprozesse ist ein bisschen auch das, was wir als Target Audience halt irgendwie raus haben, steht ja auch auf der Webseite so, also Menschen, die halt Entscheidungen treffen, welche Modelle für ihren Einsatzzweck am sinnvollsten sind und dann aber halt vielleicht nicht die Domänexpertise oder auch die Zeit haben, erstmal zehn Papiere zu lesen. Genau, also mit diesen Labels kann man halt irgendwie sehr abstrakte oder sehr umfangreiche Informationen ein bisschen vereinfachen und zugänglicher machen und das ermöglicht auch den Vergleich. Hier jetzt zum Beispiel noch ein zweites Modell. Kannst du da vielleicht auch einmal so ein bisschen zusammenfassen und irgendwie wie du, was du so jetzt als direkte Reaktion zu diesen beiden Labels siehst und wie du diesen Vergleich irgendwie siehst?

I4

Ja, als erstes natürlich die zusammengefasste Note dafür. Das ist eine mit A genotet, das ist eine mit C. Und sonst kann man sie dann unten vergleichen. Der eine ist natürlich sehr viel akkurater als der andere, dafür in den anderen Kategorien jeweils schlechter. Was erklärt, wie es zu dem C kommt, aber kann man dann natürlich auch selber entscheiden, okay, wenn mir die Akku... Accuracy auf Deutsch, Akkuratheit...

A1

Ja, die Genauigkeit.

I4

Genauigkeit ist, ja, die Genauigkeit, falls sie die wichtiger ist, kann man das natürlich so eher gewichten. Genau. Also kann man das schon mal relativ leicht, und ich denke mal, das ist für die meisten Benutzer auch schon mal die wichtigsten Kategorien. Also wie genau, keine Ahnung, das intern funktioniert, ist den meisten Leuten, glaube ich, eher egal, als wie lange braucht es, bis es mir ein Output daraus gibt, wenn ich einen Input reinschmeiße.

A1

Ja, genau. Gerade auch so hinsichtlich deines fachlichen Hintergrundes. Hast du das Gefühl, dass hier sinnvoll irgendwie noch, oder welche Aspekte fehlen vielleicht auf dem Label für dich persönlich, wo du sagst, das würde mich wirklich interessieren und jetzt sagst du, ich benutze ChatGPT vielleicht irgendwie mal in meinem Studium, in meiner Freizeit, in meiner Arbeit, was auch immer und das würde mich wirklich interessieren und da gibt es keine Informationen zu. Fällt dir da was ein in die Richtung?

I4

Wofür ist Robustness?

A1

Genau, stimmt. Über Robustness haben wir vorhin nicht gesprochen. Du musst dir vorstellen, Robustheit untersucht immer, dass wenn du deine Eingabedaten irgendwie perturbierst oder veränderst, dass die Vorhersage dann trotzdem noch genau oder korrekt ist. Das ist praktisch der Unterschied. Also die Accuracy, die unten angegeben ist, das ist ja einfach gegeben eines Testsets, das während des Trainings nicht gesehen wurde, damit du praktisch die Klassifikationsgüte deines Modells hast für ungesehene Daten. In der Hoffnung, dass das auch Rückschlüsse darüber gibt, ob in der Praxis gesehene Daten, die vielleicht sich davon ja auch nochmal unterscheiden könnten, auch korrekt erkannt werden können. Und die Corrupted Robustness, um die rauszukriegen, würde man dann diesen Testdatensatz nochmal zusätzlich entfremden, da rauschend drüberlegen, die Bilder irgendwie verschieben, distorten, miteinander mischen und, und, und, um darüber abzuschätzen, wie robust ist denn diese Accuracy. Und wir sehen jetzt hier zum Beispiel bei EfficientNet, dass die Robustheit halt drastisch einbricht, wogegen bei MobileNet die Robustheit noch relativ hoch ist. Die ist relativ nah an der ursprünglichen Accuracy.

I4

Achso also die Werte gehören quasi zusammen

A1

So ein bisschen, ja.

I4

dass je näher der eine an dem anderen Wert ist, desto sinnvoller.

A1

Also man sieht halt, dass EfficientNet immer noch besser in der Robustheit ist. Das heißt, es werden immer noch mehr von den veränderten Bildern korrekt erkannt als bei MobileNet. Aber im relativen Vergleich zu der ursprünglichen Genauigkeit ist der Unterschied halt deutlich schlechter. Das heißt, das EfficientNet scheint ein bisschen weniger robust gegen, also das scheint ein bisschen zu genau auf die Daten zu fitten, sozusagen. Das overfitted so ein bisschen dahingehend. Man kriegt zwar die Validierungsdaten extremst gut, also die Testdaten extremst gut erkannt, aber ist da für Abweichungen und Veränderungen auf diesen Bildern tatsächlich anfälliger.

I4

Das ist vielleicht tatsächlich ein interessanter Faktor vielleicht sogar interessanter als Genauigkeit, je nach Anwendungszweck vermutlich, wie sich das eine zu dem anderen unterscheidet, weil Accuracy sagt ja jetzt nicht unbedingt viel aus wenn dann halt eben nicht also man hat ja im Anwendungsfall eigentlich nie oder ja okay ich könnte mir schon Anwendungszwecke vorstellen aber weniger die wirklichen Bilder aus dem aus den test cases als beliebige Bilder

A1

Ja, also das ist tatsächlich auch, also man müsste natürlich für jedes Modell, bevor man das in ein Produkt einsetzt, gucken, ob die Bilder, so wie wir bei dem Produkt irgendwie aufkommen, überhaupt irgendwie auch nur annähernd damit zusammenhängen. Also klar, wenn man jetzt drüber nachdenkt, wir haben irgendwie ein Fließband, darauf werden irgendwelche Objekte produziert und man hat zwischendurch einen Schritt, wo eine Kamera ein Foto davon macht und du willst anhand dessen erkennen, ist das korrekt gefertigt worden oder nicht. Da hilft dir wahrscheinlich der ganze Imageschnitt-Datensatz gar nicht. Weil da geht es ja nicht um Hunde und Katzen und Flugzeuge so, sondern halt um komplett andere Sachen. ImageNet ist eher so gedacht als Bilder von allem Möglichen, was dann so ein bisschen interessant wird, zum Beispiel in Handys. Also was man jetzt gerade halt irgendwie hat bei den Google Pixel und auch bei Apple Phones, ist man kann halt mit der Kamera ein Bild von etwas machen und sich dazu Informationen holen oder erkennen lassen, was das für ein Objekt ist, solche Sachen. Und natürlich ist die, also du hast vollkommen recht, beide sind interessante Eigenschaften. Es ist sehr schwer zu sagen, was von beiden wichtiger ist als das andere. Die bedingen sich gegenseitig so ein bisschen und am Ende ist es super Application-spezifisch, wie wichtig jetzt das eine oder das andere ist.

I4

Ja. Aber was du vorhin auch meintest mit dem, dass sie sich nicht nur bedingen, sondern dass quasi der Unterschied zwischen den beiden auch relevant ist. Also dass bei dem Mobile-Net jetzt zum Beispiel die Robustheit zwar niedriger ist, aber noch näher an der Genauigkeit, wäre ja schon wieder so ein Schritt weiter quasi. Also ein Schritt weiter gedacht, als nur ich gucke drauf und habe Information. Aber ist, denke ich, schon ein relevanter Wert, um quasi zu bestimmen, wie robust irgendwas ist. Ich bin am überlegen, ob die Robustheit im Allgemeinen mehr aussagt als quasi das Abweichen von der Genauigkeit.

A1

Ja, ich verstehe, was du meinst. Das wäre durchaus etwas, was man auch kritisch anbringen könnte. Da ist halt auch wieder das Problem, wenn man jetzt einen anderen Anwendungsfall als Image-Classification nimmt, dann hat man auch wieder ganz andere Metriken. Also im Sprachmodellbereich würde man gar nicht über Top-One-Accuracy sprechen, weil da wird ja nichts klassifiziert. Das ist ja alles unüberwacht und generativ und nicht Klassifikation von einzelnen Werten. Aber um nochmal auf die Frage zurückzukommen, die ich vorhin gestellt hatte. Also dir fallen jetzt so keine weiteren Sachen ein, die auf so ein Label sinnvoll draufkommen sollten? Oder fällt dir noch irgendwie was ein, wo du sagst, das würde mich eigentlich interessieren und ist jetzt hier gerade völlig ausgeblendet sozusagen?

I4

Ne sonst würde mir nichts einfallen

A1

Okay, ich hätte jetzt gedacht, so mit IT-Sicherheit im Hinterkopf und so, dass so Themen wie, inwiefern wird bei dem Modell vielleicht auch auf Differential Privacy geachtet oder inwiefern wurden die Daten kuratiert oder so, dass das vielleicht auch irgendwie interessante Aspekte wären.

I4

Ja, das schon. Ich habe mich da halt noch nicht tiefgehend genug beschäftigt, um da Aussagen darüber treffen zu können, wie man das quantifizieren kann und ob das quantifizierbar ist.

A1

Ja, aber so im Falle von ChatGPT jetzt zum Beispiel, im Falle von so einem Sprachmodell. Also du kriegst ja gar keine dieser Informationen, wenn du ChatGPT nutzt, ohne weiteres so ohne, dass du dich jetzt wirklich auf exzessive Recherche begibst, und für manche von diesen Sachen findest du tatsächlich keine Informationen online. Gibt es da noch irgendwelche Eigenschaften wo du sagst ich würde da gerne mehr drüber wissen aber die Informationen fehlen halt irgendwie?

I4

Vielleicht für viele Sachen denke ich mal wie genau lernt das sind… Ich meine, das ist bei den meisten Sachen vermutlich sogar fast eine der ersten Sachen, die man erfährt, aber sind meine User-Inputs Teil von dem Lernsatz der KI also muss ich Angst haben, dass wenn ich in Chat-GPT reinschreibe keine Ahnung, irgendwelche persönlichen Sachen oder sowas dass die bei irgendjemand anderem mit der richtigen Frage rauskommen oder da Einfluss drauf nehmen

A1

Es geht so ein bisschen in die Richtung, was jetzt bei einigen von den bildgenerativen Modellen war, dass da gezeigt wurde, dass mit den richtigen Prompts man einzelne Bilder der Trainingsdaten fast perfekt rekonstruieren kann. Also so, wenn man den Prompt richtig formuliert mit generiere mir eine Person mit der und der Haarlänge, Haarfarbe, den und den Eigenschaften, in dem und dem Setting, dann hat man tatsächlich es geschafft, reverse, ohne dass das Modell diesen Link hat, ein sehr, sehr ähnliches Bild zu rekonstruieren aus den Trainingsdaten, wo man auch genau erkennt, welche Person das ist auf dem Bild. In die Richtung geht das ja so ein bisschen. Also kriegst du einen Prompt so formuliert oder wie anfällig ist, ChatGPT dafür, dass man einen Prompt formuliert und hinterher eins zu eins Zitate aus irgendeiner Masterarbeit findet, weil die halt im Netz ist. Aber ChatGPT das natürlich nicht entsprechend referenziert, sondern einfach das als angeblich…

I4

Hat es ja nicht als direkte Referenz, weil es guckt ja nicht aktiv nach.

A1

Genau, es schlägt nicht nach, aber wenn das halt eine plausible Antwort ist, die eins zu eins basiert eigentlich auf einem Datapoint im Training, ist das natürlich irgendwie ein Problem, gerade was Privacy angeht.

I4

Ja, das stimmt. Das wäre noch ein interessanter Faktor, aber ich weiß nicht, wie sich das quantifizieren lässe.

A1

Ja, weiß ich tatsächlich auch nicht. Dafür bin ich nicht genug LLM-Experte, aber ich bin mir sicher, dass hier an solchen Metriken und auch was Richtung Halluzination und so angeht, da werden Metriken für ausgearbeitet werden, weil natürlich muss das irgendwie quantifiziert und bewertet werden, wie problematisch das bei diesen Modellen ist. Ähm, genau, ähm, das heißt, ansonsten zur Gestaltung des Labels gäbe es da jetzt von deiner Seite aus erstmal keinen...

I4

Sonst vielleicht noch irgendwie, was mir gerade eingefallen ist, Anwendungsgebiet oder so. Ob sowas irgendwie machbar wäre, keine Ahnung.

A1

Anwendungsgrad von dem Modell?

I4

Ja, dass man drauf hat, dass es für Bilderkennung, Bildgeneration, so Sachen ist, denke ich, wäre Sinnvoll.

A1

Ja, das hängt natürlich immer dann komplett zusammen mit den, also eigentlich stehen diese Informationen in Textform da schon drauf, Inferenz auf ImageNet, dann ist halt klar, was eigentlich gemeint ist, aber natürlich könnte man auch das in Form von so einem Batch oder so irgendwie versuchen nochmal abstrakter darzustellen, absolut. Das habe ich tatsächlich auch schon mal überlegt.

I4

Der Schritt, den man dem Nutzer dann noch abnehmen würde.

A1

Ja, und den ansonsten auf den ersten Blick auch nicht ersichtlich ist, ne?

I4

Ja, der dann überprüft werden müsste

A1

Ja, das ist tatsächlich ein sehr guter Einwand, sehr korrekter Einwand, ja. Cool. Wenn du sagst, du hast schon dir über KI so ein bisschen Wissen angeeignet und interessierst dich so ein bisschen dafür, wo kriegst du deine Kommunikation, deine Informationen zu dem Thema her? Du hast vorhin schon gesagt, so klar, Medien, Nachrichten, wenn da was aufkommt, aber wenn du dann sagst, ja, das ist ja ganz interessant, jetzt möchte ich da doch nochmal ein bisschen in die Tiefe gehen, was sind so die Ressourcen, die du da zurate ziehst?

I4

Hauptsächlich YouTube, also Videos zu verschiedenen Themen. Dann teilweise auch, wo irgendjemand irgendwas implementiert mit KI. So, das hauptsächlich. Und sonst halt die Feeds.

A1

Mhm, weil wir haben tatsächlich da auch noch mal so eine kleine Liste gemacht, tatsächlich YouTube ein guter Punkt, den wir nicht so explizit bisher auf dem Schirm hatten und habe ich jetzt bisher auch in den Interviews nicht gehört, aber geht so ein bisschen in den Punkt 4, Blogposts, weil also YouTube ja eigentlich so eine Videoblog-Plattform ist. Ja, wir haben natürlich auf der einen Seite ganz klar die wissenschaftlichen Publikationen, glücklicherweise ist viel auf Arxiv, aber da hat man natürlich so ein bisschen die, ja, da muss man erstmal reinkommen, das liest man nicht mal eben so weg. Dann hat man auf Huggingface zum Beispiel, weiß nicht, ob du die Plattform kennst, das ist eine Plattform für Sprachmodelle, insbesondere, also die ist insbesondere sehr populär geworden aufgrund ihrer Anwendung für Vision Transformer und auch für NLP-Modelle, wo man so Model Cards drauf hat. Ja. Ursprünglich geht die Idee aber zurück auf Google. Papers with Code, weiß nicht, ob dir das was sagt, ist so ein bisschen eine Mischung aus den beiden, wo man auf freiwilliger Basis Ergebnisse einschicken kann, wenn man eine neue Methode entwickelt hat oder einen neuen Datensatz entwickelt hat und dann kann man so die Key Takeaways aus dem Papier da in so Leaderboards und so darstellen. Dann gibt es natürlich noch Blogposts, die ganzen Libraries haben alle irgendwie eine Dokumentation und von IBM gibt es zum Beispiel auch noch Factsheets. Für IBM-Produkte. Um hier das so ein bisschen zu visualisieren, wir haben zum Beispiel links hier die Publikation zum MobileNetV3, also den ersten Teil davon, die erste Seite von irgendwie 50 oder so. Rechts haben wir eine Model Card, da sieht man dann zum Beispiel auf zwei verschiedenen Devices mal so ein bisschen so Messwerte. Wobei, ne, es sind sogar tatsächlich die gleichen Devices, aber unterschiedliche Modelle. Das eine ist ein MobileNet Small als tier Flight und das andere... Als Malis. Genau. Papers with Code, da hat man dann so auch das Paper natürlich. Man hat aber tatsächlich dann auch noch eine Liste von allen möglichen Papieren, wo das Modell auch genutzt wurde. Teilweise kriegt man da auch so kleine Graphen gezeigt, wie sich das so verhalten hat. Rechts ein Blogpost zum MobileNetV3. Dokumentation aus Keras und ein anderes IBM Factsheet, weil IBM diese Factsheets nur rausgibt für ihre eigenen Produkte und Modelle. Deswegen gibt es nichts zu MobileNetV3 von IBM. Es ist halt nicht von IBM. Wenn man sich das jetzt alle so im Vergleich anschaut, wo siehst du da Vor- und Nachteile in diesen Visualisierungs- und Reportformen?

I4

Ja, hauptsächlich natürlich darin, wie in Depth sie sind oder eben nicht, also ich denke man kann sich da gut von oben nach unten quasi durcharbeiten je nachdem wie tief man in das Thema einsteigen möchte also ich würde eher sagen, dass es halt für verschiedene Nutzer ist quasi die verschiedenen Sachen, also das Paper ist halt dann letztendlich eher für die Leute, die sich damit beschäftigt wollen, wie es funktioniert, warum funktioniert es, was tut es, was sind wie ist man an Werte gekommen. Die, wie hießen sie von IBM die…

A1

Fact Sheets

I4

Wenn ich das jetzt richtig gesehen habe sind das dann eher so High-Level-Informationen mehr

A1

Ja, genau, es ist das, was der Name sagt. Es ist ein Factsheet. Die Model Cards sind auch ein bisschen ähnlich dazu.

I4

Genau, wo man dann weniger sieht, wie kommt es zu den Werten, sondern einfach nur hier hast du die Werte, dass man das grob kategorisieren kann, einordnen kann, vergleichen kann. Also ich denke mal, die Vor- und Nachteile sind je nachdem, wie genau oder ungenau sie sind. Das sind, würde ich sagen, Vor- und Nachteile, je nachdem, ob du es genau oder ungenau haben möchtest.

A1

Ja, ein wichtiger Punkt ist noch, dass die Label, die wir jetzt hier irgendwie gezeigt haben, die werden tatsächlich in ähnlicher Form so von einer Software Library ausgespuckt, die ich irgendwie über die letzten Jahre entwickelt habe. Da kann man einfach so ein Pandas DataFrame reinladen, hat man da pro Zeile drinstehen, das Modell, das auf dem Datensatz hat, ist die und die Eigenschaften. Und das ganze Scoring und auch die Generierung von diesem PDF erfolgt halt voll automatisch auf der Basis. Wogegen ich zumindest bisher das Gefühl habe, dass sowohl das Paper schreiben als auch das Dokumentations schreiben, Blogposts und Co., das ist halt sehr viel manueller Aufwand. Wobei das jetzt durch Sprachmodelle vielleicht auch einfacher wird, wobei man da natürlich auch wieder die Frage stellen kann, ist das so sinnvoll, dass das einfacher wird, weil wenn man halt ChatGPT dafür benutzt, um da Blogposts schreiben zu lassen, dann ja.

I4

Ist das dann auch richtig.

A1

Man hat dann trotzdem den Manual-Effort, das nochmal zu Proof-Checken, sag ich mal, wenn man es richtig macht. Und bei mir, ich habe halt irgendwie auch viel im Bereich Benchmarking gemacht, um halt so Modelle automatisiert durchzutesten. Genau. Ja, ich glaube, ganz wichtig festzuhalten ist auch, dass wir natürlich jetzt nicht glauben, dass Label die ganzen anderen Sachen irgendwie ablösen wird. Es ist halt eine zusätzliche Repräsentationsform, die aber halt so ein bisschen den Vergleich und die Einsichten at a glance irgendwie ermöglichen, was halt einfach bisher nicht möglich ist.

I4

Quasi wie der Factsheet, nur ein bisschen noch mehr Highlevel. Ja. Und so die Note nimmt einem dann sogar den, kann einem auch den Vergleich von den Kategorien abnehmen und so.

A1

Genau, also es ist auch so ein bisschen, da kann dann auch so ein Unternehmen theoretisch einfach sagen, wir legen die und die Werte für unsere Produkte fest, so natürlich ist Genauigkeit immer wichtig, so klar. Es gibt auch tatsächlich auch Firmen, die sagen, ob das jetzt 80 oder 85 Prozent sind, ist eigentlich nicht so wichtig. Wenn das Modell dann aber halt nur die Hälfte der Energie kostet, dann ist das schon was, wo dann viele Anwender sehr interessiert wären, weil das natürlich am Ende alles Energiekosten sind, die dieses Produkt halt irgendwie hat und diese Trades transparenter zu machen, wäre halt glaube ich im großen Interesse für uns alle. Ja, ansonsten, das machen halt alle. Also wollen das genaueste Modell, aber diese Kosteneffizienz, die fällt so ein bisschen hinten runter.

I4

Wenn ich jetzt Suggestions haben möchte beim Nachrichten schreiben oder sowas, würde ich natürlich es ist mir egal, ob das jetzt 80 oder 85 Prozent genau ist, was der mir vorschlägt, was das Wort ist, weil letztendlich überprüft der Nutzer das ja sowieso, aber wenn es dafür halt nicht zwei Sekunden laden muss, sondern nur eine Sekunde oder sehr viel weniger Energie kostet, bei meinem Handy oder so, dann ist es natürlich sinnvoller,

A1

Ja, das ist ja auch gerade irgendwie im Mobile-Bereich total relevant gerade. Also tatsächlich pushen die auch da drauf, die Modelle kleiner zu machen. Das ist auch die ganze Idee hinter MobileNet, by the way. Also das ist ein Modell, was ja explizit gemacht wurde für Smartphones oder wo explizit das Deployment auf Smartphones irgendwie auch fokussiert wurde. Aber ja, es ist in der Wissenschaft auch sehr viel, dass man tatsächlich sehr viel über Genauigkeit nur spricht und die Leute meinen, sie haben das neu beste Modell, was sie auch auf dem Papier, wenn man sich die Qualität anguckt, vielleicht haben. Aber dafür haben die halt irgendwie einen Faktor 5, den höheren Laufzeit, Stromverbrauch, Ressourcenverbrauch, was man durchaus auch mal hinterfragen kann. Also Web-Modelle. Das ist bei uns hier im XXXXXXXXXXXXX auch sehr stark für Resource-Aware-Machine-Learning. Das heißt nicht, dass man keine Ressourcen ausgeben darf. Wir dürfen auch irgendwie Big-NLP-Models und so machen, aber man sollte halt transparent sein dabei und dass auch diese Trades auch irgendwie beleuchten. Also da setze ich mich auch stark für ein. Genau. Kommen wir noch am Ende so ein bisschen drauf zu sprechen, Thema Vertrauenswürdigkeit. Vertrauenswürdigkeit ist ja auch was, was bei uns im XXXXXX XXXXXXXX und auch im XXXXXXXX XXXXXX XXX XXXXXXXXXXX XX, hier in XXXXXXXX und tatsächlich auch an der XXX auch sehr groß ist, sehr groß thematisiert wird. Hast du da Gedanken zu jetzt auf Basis von allem über das wir schon gesprochen haben? Vertrauenswürdigkeit?

I4

Ja, also hier vorhin hatten wir ja schon über gesprochen über was passiert mit meinen Inputs, werden die weiterverarbeitet oder nicht und können die irgendwo ausgegeben werden, das ist natürlich für Vertrauen wichtig beziehungsweise Vertraulichkeit ist ja eher die Geheimhaltung (ungenau) dass meine Inputs vertraulich sind beziehungsweise auch eben nicht sein können dass das bewusst sein sollte ob Inputs weiterverarbeitet werden, nicht weiterverarbeitet werden, was genau mit dem passiert ist denke ich schon relevant oder sollte den Endnutzern von einem Produkt denke ich schon wichtig sein weil wenn jetzt, keine Ahnung bei meinem Praktikum zum Beispiel dass man keine Kundendaten mit ChatGPT verwenden sollte, weil man letztendlich ja die dann an ein drittes Unternehmen weitergibt und man nicht genau weiß was damit passiert da genau aber ich denke das ist hier jetzt ein bisschen was anderes

A1

Ja, also Informationen zu Vertrauenswürdigkeit wäre auf jeden Fall was, was auch irgendwie auf den Labels natürlich irgendwo dargestellt werden könnte. Wobei das oft ja nicht an dem Modell selber hängt, ne? Das Modell, das läuft auf irgendeinem Rechner. Und das kann erstmal nicht die Daten sehr vertrauenswürdig oder nicht vertrauenswürdig einsetzen. Da wäre es dann mehr das Produkt oder der Service, der mit diesem Modell zusammenhängt, so wie bei OpenAI, so eine Web-Schnittstelle praktisch. Da müsste dann diese Information halt irgendwie transparent gemacht werden. Das Modell selber macht erstmal die Daten jetzt nicht vertrauenswürdig, also behandelt die nicht mehr oder weniger vertrauenswürdig. Für das Modell ist das nur ein Haufen Bytes, das wird da einmal durchgejagt und fertig. Wie ist das denn, wenn man darüber nachdenkt, wer so ein Label ausstellen könnte? Weil als Hintergrund, also XXXXXXXXX hat letztes Jahr eine Studie dazu gemacht, inwiefern so Trust-Siegel für KI helfen würden. Und in der Studie kam netto raus, ich fasse das grob zusammen, Menschen, also es war eine quantitative Umfrage, also 40.000 Befragte, nicht qualitativ, die haben Geld dafür bekommen, dass sie diesen Fragebogen online ausgefüllt haben. Und da war so das generelle Fazit: ich vertraue KI eigentlich nicht. Die haben dann irgendwie zwei Anwendungsfälle gebaut, so einmal irgendwie, ich glaube, Krebsfrüherkennung, das andere weiß ich aus dem Kopf nicht mehr, und haben dann einmal praktisch der einen Hälfte das Produkt so vermittelt und gefragt, wie vertrauenswürdig ist das, und den anderen Leuten haben sie das vermittelt, und da war aber so ein Trust-Siegel dabei. Und da war tatsächlich so das Feedback im Schnitt, ist mir eigentlich egal, ob da ein Trust-Siegel drauf ist oder nicht, aber im Zweifelsfall vertraue ich dem erstmal nicht. So eine sehr technologie-kritische Meinung, was auch nachvollziehbar ist, weil das sind ja eben keine Experten, die halt irgendwie diese Entscheidung treffen müssen. Lustigerweise setzen sich ja auch alle in Flugzeuge. So, und wenn man dann aber, also dieser Technologie vertrauen die Leute, da haben wir auch einfach jetzt schon viele Jahre, wo das gut funktioniert. Aber man sieht das auch so ein bisschen im Bereich E-Autos und so, autonomes Fahren. Viele sind sich da halt einfach, sind da nicht so, haben da nicht so das Vertrauen in die Technologie. Wir fragen uns halt auch, wie das bei denen ist.

I4

Ich habe das Gefühl, das liegt am Namen. Dass sich viele Leute unter KI was mehr vorstellen, als es letztendlich ist. Also dass der Otto-Normalverbraucher dann eher so ein Hollywood-artiges Verständnis von KI hat, dass da wirklich was, ja kann man drüber streiten, dass da wirklich was ist, was denkt und Gedanken hat zu irgendeinem Thema oder sowas und nicht, dass es ein System ist was eine Menge an Inputs und Outputs hat an denen halt Verfahren, also bestimmt wird, was getan wird. Zudem wer so ein Label ausstellen könnte ist immer schwierig im IT-Sicherheitskontext gibt es da ja ein passendes Pendant mit Zertifikaten für Websites und sowas zu dass da letztendlich auch überragend viel Vertrauen von Endbenutzern in alles reingesteckt wird, was IT-Sicherheit angeht was irgendwo verständlich ist und irgendwo auch nicht, weil jeder sich, nicht jeder sich tiefgehend damit beschäftigen kann, wie funktioniert Verschlüsselung, wie funktioniert dieses ganze Certificate Authority System was ist ein Root Certificate und wem vertraue ich eigentlich überhaupt implizit also letztendlich müsste das ja ein Unternehmen sein oder oder, oder irgendeine Autorität, die unabhängig von den Unternehmen ist und wo man sicherstellen muss dass die einzelnen Leute, die das einordnen von den Sachen durchführen, dass die auch äh nicht geschmiert sind und ich meine, viele von den Sachen sind natürlich auch relativ gut per Peer Review Bar so die Zeit lässt sich natürlich einfach überprüfen okay, mit den angegebenen Merkmalen hat das wirklich diese Laufzeit oder nicht und Accuracy ist denke ich auch, Power Draw wird vermutlich auch gut überprüfbar sein mit… genau eigentlich sollte ja bei allen den Werten allen angegebenen Faktoren bisher gut nachvollziehbar sein oder gut überprüfbar sein, ob die angegebenen Sachen korrekt sind und ich denke mal schon, dass sich organisch irgendwas entwickeln kann dem dann im Allgemeinen vertraut wird, das ist ja dann letztendlich wie bei den meisten anderen Sachen auch, wenn man genug Zeit hat und sich einfach rausstellt, okay, mit diesen 80 Modellen hat es bisher gestimmt, warum soll es dann mit dem 81. Modell nicht stimmen und irgendwie, man hat eine Autorität, wo dann viele Experten sagen, ja okay, das passt, dann wird das schon passen und wenn sich sowas gebildet hat, dann denke ich schon, dass die Vertraulichkeit da rein, oder das Vertrauen darein, sorry, schon einen Unterschied machen würde. Also ich kann mir schon vorstellen, dass mich das beruhigen würde, irgendwie so eine Autorität zu haben, die dann Sachen beurteilt und sowas ausstellt, das ist, denke ich, schon sinnvoll.

A1

Okay, cool. Also die Frage haben wir uns halt auch so ein bisschen gestellt. Das war auch was, was bei XXXXXXXXXX Studie aber rumkam. Also die Leute haben halt irgendwie auch oft gesagt, ich weiß nicht, wo das herkommt, also vertraue ich dem nicht. Und wir haben ja generell auch gerade so ein Problem gesellschaftlich mit mangelndem Vertrauen in Technologie, Politik, Authorities, wo wir ja auch als Machine Learner einfach nichts machen können. Also dass die Leute eine Impfskepsis zum Beispiel haben, da hat man ja ein komplett durchstrukturiertes System von wie man Impfprodukte irgendwie entwickelt und an den Markt bringt. Klar, das hat auch seine Fehler, das ist auch nicht perfekt, aber es ist ein ziemlich gut strukturiertes, lang etabliertes und auch organisiertes. Und auch irgendwo durchsichtiges System. Also wer will, der kann relativ gut nachvollziehen, also der kann halt wirklich sehr viel Informationen dazu finden, wie so eine Studie durchgeführt wird und so.

I4

Man kann das ja auch reproduzieren, wenn man erstmal genug studiert hat.

A1

Ja, genau. Also der Prozess wird da tatsächlich recht offen gelegt. Trotzdem sagen die Leute, ich vertraue dem nicht. Einfach weil ich habe online gelesen, dass man dem nicht vertrauen kann. Das heißt, solche Probleme können wir ja nicht wirklich lösen. Aber wir überlegen natürlich auch, wie kann so ein Label vielleicht auch was werden, wo Menschen dann auch wirklich sagen, ja, okay, das hat Hand und Fuß, dem vertraue ich einfach mal. Und wie gesagt, wir haben so Sachen wie Autos und Flugzeuge, wo erstmal alle so ein Grundvertrauen irgendwie haben, bis auf ganz wenige Ausnahmen. Und in den letzten Jahren hat sich die Technologie gerade im IT-Bereich so rapide entwickelt. Und da merkt man, dass die Leute tatsächlich Vertrauen komplett über Bord werfen, solange sie coole Benefits haben. Also sind halt alle sofort dabei, sich eine Alexa zu Hause hinzustellen, weil das ist irgendwie total lustig. Und irgendwie ...

I4

Internet of Things. Ganz schwierig IT-Sicherheitstechnisch

A1

So Entertainment und Praktisch. Aber ich brauche dir nicht sagen, aus sicherheitstechnischer Sicht ist das alles ganz schön wild und eigentlich abstrus.

I4

Dass jedes Gerät und dein Kühlschrank und deine Heizung und keine Ahnung der Fressnapf von deiner Katze alles Computer haben und im Internet sind schwierig

A1

Ja, genau. Und gleichzeitig gesagt, Usability und Entertainment überwiegt halt, ne? Die Leute kaufen es halt trotzdem und haben da Bock drauf, so. Und KI ist gerade so lustig, weil alle Firmen gerade sagen, wir müssen KI machen und ganz viele Leute in der Bevölkerung sagen, ich mache nichts mit KI, das ist mir alles zu gruselig, so. Also da haben wir gerade so voll den Break in der Technologie irgendwie, wo ich gerade tatsächlich nicht das Gefühl habe, dass die neue KI im Pixel 7 ein Verkaufsargument ist. Also ich glaube nicht, dass die Leute deswegen das Ding kaufen wollen, so, weil da KI draufsteht.

I4

Also ich glaube auch nicht, dass… Oh meine Kamera ist stehen geblieben… ich glaube auch nicht, dass ein Verkaufs- oder eben Nicht-Verkaufs-Argument ist für die meisten Leute. Oh, mein Handy ist leer. Oh doch nicht. Warum macht meine Kamera nichts?

A1

Sonst ist es auch nicht wild alles ist gut…

I4

Ja, also ich denke mal, bei zum Beispiel der KI im Google Pixel ist es jetzt nicht wichtig, dass da KI drin ist. Und ich glaube, für die meisten Leute auch relativ egal, weil es so eine Abstraktionsstufe weiter ist. Du redest nicht direkt mit der KI oder die KI macht keine Bilder, sondern du machst ein Bild und das geht durch eine KI. Und dann ist es schon wieder so weit hinten, dass es den meisten Leuten vermutlich eher egal ist. Hauptsache, das Endergebnis stimmt.

A1

Das war ja total lustig, zum Beispiel, dass bei den Samsung-Smartphones war das irgendwie vor ein paar Monaten mal Thema, dass die automatische Bildkorrektur von denen bei Sonnenuntergängen und so halt einfach angefangen hat, die benutzen halt generative KI, um das cooler aussehen zu lassen und das heißt, am Ende ist das nicht mehr das Foto, was du gemacht hast, sondern einfach so eine Synthese aus deinem Foto und allen möglichen Bildern, die so im Internet sind

I4

Und jedem Foto

A1

Ja, genau, und dann sieht das halt irgendwie cooler aus, aber wie gesagt, Usability überwiegt halt an der Stelle wirklich dem kritischen Denken, so Genau, ja, aber cool Dann sind wir tatsächlich auch mit unserem Interview jetzt durch

I4

Vielleicht noch zu dem Label. Ich könnte mir vorstellen, dass Vertrauen in das Label leichter ist als jetzt bei, vorhin hatten wir das Beispiel von Impfung, weil es leichter nachzuchecken ist. Das könnte ich mir schon gut vorstellen, dass das dahingeht. Der Barrier of Entry ist, glaube ich, ein ganz Stück tiefer. Du kannst halt die Zeit relativ einfach checken. Du brauchst halt nur das entsprechende System und kannst dann abchecken, okay, ist das in etwa in dem Bereich. Im Gegensatz zu einer repräsentativen Studie zu den Effekten von, keine Ahnung, welcher Impfung jetzt auch immer, kann man das einfach laufen lassen und checken, okay, stimmt das? Und ich denke mal, da lässt sich dann leichter ein Vertrauen in so ein Label aufbauen.

A1

zumindest bei Open-Source-Modellen. Man muss auch in meinem Hinterkopf behalten, dass zum Beispiel die GPT-Varianten von OpenAI sind ja eben nicht so open. Das heißt, die kann man nicht mal eben auf eigener Hardware durchmessen. Da müsste man einfach hoffen, dass das plausibel ist. Und tatsächlich haben die ja in ihrer, bei GPT-4 in der Technical Documentation auch explizit reingeschrieben, aufgrund der Competitiveness im NLP-Bereich disclosen wir jegliche Informationen zu Trainingsdaten, Trainingsaufwand, Trainingsrechenarchitektur, Netzarchitektur, Hyperparameter fürs Training und, und, und. Das heißt, da ist gar keine Reproduzierbarkeit gegeben. Nicht mal im Kern. Abgesehen davon, dass man für Reproduktion von dem Modell auch irgendwie, weiß ich nicht, 8000 GPUs bräuchte oder so, was ja auch keiner zu Hause hat. Aber klar, bei kleineren Modellen kann man das schon besser reproduzieren. Ja, klar. Genau. Tatsächlich, wie gesagt, sind wir dann mit dem Inhalt, des Interviews durch. Ich würde dann die Aufnahme an der Stelle einmal beenden.