A1

Perfekt. Und dann würde ich dich zunächst mal bitten, dich einmal nochmal kurz vorzustellen.

I16

Ja, hallo, ich bin der XXXXXXX. Ich bin auch Informatiker, habe tatsächlich auch an der XX XXXXXXXX studiert. Und ich habe seit 2018 das Unternehmen XXXXXXXXXX XXXXXXX XXXXXXXXXXXXX. Wir beschäftigen uns mit Digitalisierung für echte Menschen und beschäftigen uns da natürlich jetzt auch verstärkt seit den letzten zwei Jahren sehr viel mit KI. Ich hatte selber auch immer wieder in meinem gesamten Werdegang sehr viele Berührungspunkte, damit ich im Maschinenbau lange unterwegs war, wo das ein großes Thema war. Der ganze Bereich Machiene Learning. Ich war vorher in der Medizin, wo es auch ein riesengroßes Thema war, insbesondere natürlich in der Bildverarbeitung. Und für mich ist es jetzt halt auch ein sehr großes Thema, weil wir sagen, wir bringen sehr verschiedene Aspekte von Digitalisierung zusammen. Wie können wir die menschenwürdig und auch menschengenehm gestalten? Was ist dafür technologisch möglich und auch was für Einsatzfelder und was für Technologien gibt es da überhaupt? Und für uns ist da zum Beispiel das Thema KI im Mittelstand super spannend. Und natürlich auch, welche Sicherheitsbedenken das zum Beispiel mit sich bringt, die ja nicht ohne Grund da sind. Und deshalb auch so das Thema Vertrauenswürdigkeit, Zuverlässigkeit bei KI-Systemen. Das ist was, womit ich mich auch sehr intensiv beschäftige in den letzten Monaten.

A1

Okay. [Schlechte Verbindung] nicht so sehr selber mit KI arbeiten, also ganz wertfrei, sondern vor allem auch mit Leuten sich über KI austauschen und Firmen da ein bisschen beraten und unterstützen. Habe ich das richtig irgendwie?

I16

Ja, wir machen auch ganz klare Modellprojekte, wo wir sagen, okay, wir lernen selber neuronale Netze an für spezielle Aufgabengebiete. Aber sehr oft ist es bei uns halt auch wirklich Beratung oder auch überhaupt erstmal Awareness schaffen. Also ich tingle seit über einem Jahr regelmäßig durch verschiedenste Unternehmer, Netzwerke, halte da Vorträge zum Thema, wie funktioniert KI überhaupt, wie funktioniert Deep Learning, wie funktionieren Transformer-Architekturen und warum muss ich das eigentlich als Geschäftsführer wissen, wenn ich das sinnvoll zur Anwendung bringen will, auch wenn mich das ganze Technik-Thema gar nicht interessiert. Und das ist sehr erhellend, finde ich, einerseits. Also erstmal gefällt das den Leuten sehr gut zu sehen, ah, okay, das ist keine Raketenwissenschaft, das kann mir auch jemand auf Sendung mit der Maus Niveau erklären. Aber wichtig ist es auch, dass die Leute da erkennen, das ist halt kein Selbstläufer. Also das ist nicht so, dass das ohne Arbeit vonstattengeht und dass man einfach da schnippt und sagt, das läuft jetzt. Und das ist, glaube ich, auch eine zentrale Erkenntnis, die die Leute dann mit nach Hause nehmen und da wachsen dann in aller Regel auch die richtigen Projekte draus, wo man dann sagt, was können wir damit eigentlich machen mit vertretbarem Aufwand, wo bringt uns das wirklich weiter. Und können wir der ganzen Sache dann halt vertrauen. Auch ein ganz, ganz großes Thema.

A1

Ja, auf jeden Fall. Aber cool. Also auch durchaus in der Tiefe mal prototypisch irgendwie Lösungen mal ein paar Modelle trainieren, ein bisschen was ausarbeiten und so. Wobei da wahrscheinlich dann die Operationalisierung eher wieder bei den Unternehmen liegt. Also ihr macht praktisch nur so einen Aufschlag. Aber das ist auf jeden Fall cool, dass ihr sowohl in der Tiefe als auch ein bisschen High-Level unterwegs seid. Weil, ja, also kenne ich auch aus meiner Arbeit. Ich bin auch in einem Industrieprojekt mit drin und es ist teilweise sehr unterschiedlich, wie die Leute zu KI stehen und was sie davon denken und so. Und was sind denn so, wenn du so an deine alltäglichen Prozesse denkst oder auch die Prozesse, die du begleitest, sage ich mal, was dann ja auch wieder deine alltäglichen Prozesse sind, die größten Herausforderungen im Kontext von KI?

I16

Also das, worüber ich eigentlich am häufigsten stolpere, ist, dass die Leute versuchen, den zweiten oder dritten Schritt vor dem ersten zu machen. Ich glaube, wenn man sich mit KI beschäftigen will, selbst als nicht-technische Person, dann gibt es einen gewissen Grundstock an Wissen, den man haben sollte. Über die Digitalisierung, also es braucht eine [Schlechte Verbindung] sonst gibt es keine KI-Kompetenz und das versuchen viele Leute jetzt halt zu überspringen und sagen, muss ich ja alles gar nicht mehr lernen, macht ja die KI. Und so funktioniert es halt nicht. Ich habe halt zum Beispiel viele Beratungen, insbesondere im HR-Bereich, wo das Thema super gut ankommt eigentlich. Von wegen Bewerbungen auswerten, Bewerber vorsortieren, optimal schauen, wo können Leute weiterentwickelt werden, welche Fortbildungen sollten die machen. Und da fällt halt zum Beispiel das Thema Datenschutz und das Thema Privatsphäre sehr, sehr oft hinten runter. Und auch so Bias-Effekte, die ja bei KI nun mal leider zwangsläufig auftreten müssen, werden da erstmal geflissentlich ignoriert, weil die Leute das überhaupt nicht auf dem Schirm haben. Weil das statistische Grundverständnis dafür fehlt, weil das technische Grundverständnis fehlt. Und das ist halt sowas wo ich merke, dass das eigentlich dann auch immer erstmal aufgebaut werden muss bevor es da weitergehen kann

A1

Genau also fairness ist auch ein großes Thema natürlich in der KI Forschung, kommt immer irgendwie so ein bisschen darauf zurück zu unsere Daten die statistisch irgendwie erhoben werden, die sind halt voll von biases weil wir halt als Gesellschaft und als Individuen Biases haben und reproduzieren. Und man braucht dann am Ende auch immer Menschen, die diesen Bias überhaupt erstmal erkennen und dann kann man auch versuchen, den zu korrigieren, auch in statistischen Datensätzen, Modellen. Aber ja, glaube ich sofort, dass Leute, die halt nicht dieses technische Wissen haben, das halt gar nicht auf dem Schirm haben und denken, naja, eine Vorauswahl ist doch erstmal super, spart uns Arbeit. Aber introduced halt potenziell starken Bias. Fällt dir irgendwas ein, wo du sagst, das würde die Arbeit vereinfachen? Ganz allgemein, das würde diese, also, wir reden da ja viel von so Wissenstransfer an der Stelle. Wo hast du das Gefühl, könnten wir auch als ML-Forscher oder KI-Forscher irgendwie besser werden, um halt solche Prozesse zu vereinfachen?

I16

Ich glaube, das Thema muss noch viel, viel mehr in den populärwissenschaftlichen Bereich rein. Da gibt es zwar schon tolle Initiativen, ich denke da zum Beispiel an einen KI-Campus, der halt wirklich großartig ist und da ganz, ganz viel Aufklärungsarbeit leistet. Aber ich glaube, das muss auch mit anderen Sachen passieren. Gerade auch jetzt, wo wir in der Technologie halt Fortschrittszeitrahmen haben, die absurd schnell sind. Da hilft es dann halt auch nichts, absolutes Grundlagenwissen zu informieren, wenn die praktischen Systeme drei Generationen weiter sind. Da sehe ich halt schon noch großes Potenzial, dass man da sicherlich auch von akademischer Seite unterstützen kann, indem man sagt, wir bereiten das auf, sodass Leute, die nicht aus diesem Fachbereich kommen, auch überhaupt in der Lage sind, zu verstehen, was da passiert, wie das zu bewerten ist, auch wirklich sinnvoll zu bewerten ist. Weil einerseits gibt es jetzt Leute, die machen da totale Panik und sehen die Terminators schon vor der eigenen Haustür stehen. Und auf der anderen Seite gibt es dann Leute, die sagen, wir müssen ja hier Fortschritt machen, alle Risiken erst mal ignorieren und beides halte ich für völlig falsch. Und das kommt halt auch in der Gesellschaft nicht an. Also ich bin halt auch viel im Schulbereich unterwegs. Ich bin XXXXXXXXXXXXXXXX schon seit vielen, vielen Jahren. Ich sitze in der XXXX XXX XXX XXXXXXXXX XXXXXXX hier in Deutschland für den Bundeswettbewerb. Und da merkt man das auch. KI ist ein Thema, dafür können sich Schulen oder auch Hochschulen ja überhaupt nicht mehr verschließen. Und wenn ich dann sehe, wie dieser eine Fall, den es vor einigen Monaten an der LMU gegeben hat mit dem Studenten, der sich für den KI Masterstudiengang einschreiben wollte und der abgelehnt worden ist, weil er seine Bewerbung mit ChatGPT geschrieben hat, das zeigt für mich, dass wir da vielleicht noch diverse Diskrepanzen drin haben, sogar innerhalb des Hochschulbetriebs, wo die Verwaltung da dem technologischen Fortschritt vielleicht noch Jahre hinterher ist. Also ich glaube, da kann man viel, viel machen, indem man viel Aufklärungsarbeit leistet und das auch in einer Sprache und in einem Umfang tut, den eine breite Masse der Menschen auch einfach erreichen kann.

A1

Also erstmal gehe ich komplett mit, gehen wir glaube ich auch alle hier aus dem Team mit. Das ist ja auch so ein bisschen irgendwie unsere Motivation und deswegen muss das auch interdisziplinär geschehen. Kannst du das noch ein bisschen konkretisieren? Also auch Aufklärungsarbeit leisten in verständlicher Sprache ist ja immer noch relativ allgemein gefasst. Hast du so konkrete Ideen oder Maßnahmen?

I16

Naja, im Prinzip fängt das schon ganz unten an, wenn man so über Wahrscheinlichkeitstheoretische Begriffe redet. Also Menschen sind ja nun mal ganz hundsmiserable, intuitive Statistiker und schätzen Wahrscheinlichkeiten völlig falsch ein. Das heißt, man muss sich konzentriert an die Sachen ran setzen und die Sachen beobachten. Und man muss halt auch verstehen, dass KI-Systeme stochastische Systeme sind und dass sie dementsprechend auch entsprechende Antworten liefern und dass damit einiges möglich ist. Aber eben halt auch nicht alles und vor allem nicht alles mit absoluter Sicherheit. Und das ist so ein Grundverständnis, was eigentlich schon mal überhaupt da sein muss, damit die Leute das verstehen, dass wir auch Dinge haben, die das limitieren. Dass wir zum Beispiel sagen das Gesetz der großen Zahl. Das müssen wir halt erfüllen. Das bringt nichts, wenn du als Personalabteilung sagst, ich schmeiße da mal 30 Bewerbungen rein. Jetzt ist das System angelernt. Das funktioniert halt nicht. Und das sind solche Sachen, von denen ich sagen würde, das müssen die Leute erst mal auf dem Schirm haben und dann vielleicht auch auf dem Schirm haben. Was muss man ganz konkret wissen, auch bei dem, womit man heutzutage so bombardiert wird? Sei das jetzt als Eltern, die sich da für die Schule interessieren, sei das als Angestellte, die in einem Unternehmen arbeiten, seien das Unternehmensinhaber, die sich mit dem Thema auseinandersetzen müssen. Die kriegen ja von allem irgendwelches Marketingzeug übergeworfen. Und das ist ja jetzt größer, schneller, weiter jede Woche was Neues. Und meistens sind die Änderungen nur im Detail und meistens sind es halt nur Benchmarks, die auch günstig aufgebrauscht werden. Und diesen tatsächlichen Stand, was können wir mit KI und was können wir nicht? Und was ist in absehbarer Zukunft vielleicht möglich und was wird nicht möglich sein? Und was haben wir verstanden und was haben wir vor allem auch als Forscher noch nicht verstanden in dem Bereich? Ich glaube, das sind Sachen, die müsste man aufbereiten, um den Leuten auch zu zeigen, ey, das ist ja gar nicht so weit weg von unseren Gedanken. Uns haben nur die wesentlichen Informationen gefehlt. Und das, was die da machen, ist gar nicht Science Fiction, sondern maximal das, was die Marketingabteilung der KI Unternehmen uns erzählen, ist Science Fiction. Aber ich habe die Möglichkeit, das vielleicht selber zu unterscheiden. Weil ich gelernt habe, okay, das funktioniert halt auf statistischer Basis, das hat gewisse Einschränkungen. Ich habe gelernt, wie neuronale Netze grob funktionieren. Ich habe davon eine Vorstellung, wie das geht. Ich habe eine Vorstellung davon, was damit machbar ist und was nicht. Was für Daten erfasst werden und welche nicht. Und ich glaube, da ist noch ganz viel Luft nach oben.

A1

Ja, glaube ich auch. Es ist tatsächlich, ja, es ist wirklich Grundlagenwissen im Endeffekt, was an vielen Stellen fehlt. Es gibt ja, also sage ich dir wahrscheinlich nichts Neues, so Gamblers Fallet, so dass Leute halt einfach so Zufallsexperimente wie einen Münzwurf nicht realistisch einschätzen können und sich dann denken. So das ganze Konzept hinter Glücksspiele im Endeffekt. Also warum Leute ins Casino gehen, weil sie halt denken, dass sie da tatsächlich gewinnen können. Und natürlich können sie das auch, aber mit einer sehr, sehr, sehr niedrig statistischen Chance. Und ja, die meisten Leute verlieren halt Geld. Und dann merkt man, dass wir da als Bevölkerung noch nicht so richtig geschult sind. Ja, ist total richtig. Wir haben ja jetzt uns auch irgendwie eine Sache überlegt, wo wir vielleicht ein bisschen Wissenstransfer irgendwie vereinfachen und verbessern können. Dafür würde ich tatsächlich auch mal eben kurz meinen Bildschirm übertragen. Ich muss nur mal kurz die Powerpoint schon mal starten.

A4

Bevor wir jetzt den nächsten Schritt machen, könnte ich vielleicht nochmal nachfragen, was genau jetzt deine Rolle ist? Ich weiß nicht, ob ich es noch nicht ganz verstanden habe oder ob wir noch nicht so im Tiefen darüber gesprochen haben. Was genau machst du, was sind deine Verbindungen zu KI?

I16

Also ich versuche erstmal generell in Unternehmen eine Brücke zwischen Informationstechnologie und normalen Menschen zu schlagen, weil ich beide Sprachen spreche und als Übersetzer fungieren kann und auch den Leuten die Angst davor nehmen kann, sich mit dem Technologiethema überhaupt auseinanderzusetzen. KI ist halt wirklich sowas, was oben jetzt draufgekommen ist die letzten Jahre. Das war zwar schon viel, viel länger vorher präsent, weil wie gesagt, in der Medizin, im Maschinenbau benutzen wir das ja schon seit vielen, vielen Jahren, aber da war es dann auch Spielwiese für die Ingenieure und der Rest musste sich nicht damit auseinandersetzen. Plötzlich muss sich jeder damit auseinandersetzen. Und plötzlich kriegt auch jeder gesagt, wenn du dich nicht damit auseinandersetzt, dann bist du in ein paar Jahren deinen Job los. Und das schürt halt völlig neue Ängste und da komme ich dann halt auch beratend rein. Als Coach für Teams, für einzelne Personen, die sagen, sie wollen nicht so völlig blind vor dem Thema stehen. Also das ist viel Beratungsarbeit, was ich an der Stelle mache. Und dann später halt auch tatsächlich auf der technischen Ebene als Softwarearchitekt begleiten von technischen Teams, die diese Dinge dann konkret umsetzen müssen, Prototypen schaffen, mit denen man Sachen prüfen kann, ob die überhaupt so funktionieren, wie man sich das vorstellt. Das sind so die Themen, mit denen ich mich beschäftige.

A1

Das heißt ursprünglich eigentlich Digitalität im Allgemeinen und jetzt sind in den letzten zwei Jahren, wo KI so richtig marktfähig geworden ist, so wie du gesagt hast, Regression machen wir halt auch schon seit 20 Jahren in der Industrie, aber jetzt wo man sich Chatbots und so einkaufen kann oder ganz neue Möglichkeiten für Lösungen am Markt sind, kommt jetzt das KI-Thema noch zusätzlich obendrauf?

I16

Richtig.

A1

Gut. Ja, genau. Also wie gesagt, wir beschäftigen uns ja auch mit KI und insbesondere Wissenstransfer und haben uns da auch irgendwie was überlegt. Das weißt du, weil du dich ja zur Studie angemeldet hast. Dafür teile ich jetzt wirklich meinen Bildschirm, wie gerade schon angekündigt. Kannst du uns auf das, was du siehst, einfach mal so eine unmittelbare Reaktion geben?

I16

Das ist ein Label mit relativ viel Information. Sieht für mich so ein bisschen aus, als komme ich ganz gut dran zu beurteilen, wie energieintensiv ist das ganze Ding, wie verlässlich ist das ganze Ding, wie schnell ist das ganze Ding. Und ich sehe halt natürlich, es ist A, grün. Das heißt, das ist das Beste, was ich kriegen kann. Das ist natürlich immer schön.

A1

Genau. Also wir haben unten halt verschiedene Metriken, oben so eine Art Gesamtbewertung in der Skala. Sind dir die Begriffe geläufig soweit oder gibt es Sachen, die du so gar nicht zuordnen kannst?

I16

Also mit der Corrupted Robustness kann ich jetzt in der Form nicht so viel anfangen.

A1

Sehr gut. Bist du nicht der Erste. Genau. Um das vielleicht einmal kurz zu erklären. Also Accuracy kennst du ja wahrscheinlich. Das ist das, was man immer maximieren will bei Klassifikation. Bei Corrupted Robustness geht es darum, dass man die Testdaten, die man für die Accuracy-Berechnung genommen hat, praktisch bearbeitet oder nachbearbeitet. Nicht gezielt, sondern eher verschiedene Filter darüber legt, die einfach in irgendeiner Form halt irgendwie verändert, um zu gucken, wie robust die Vorhersagequalität ist, wenn man da jetzt irgendwie so ganz normal irgendwelchen Gauss-Noise drauf hat. Aber vielleicht auch spezifischere Sachen, irgendwelche perspektivischen Verzerrungen oder so, ein bisschen Nebel oder so irgendwas, was halt realistisch bei einem Kamerabild zum Beispiel auftreten könnte, um halt abzuschätzen, wie sehr diese Robustheit vielleicht auch gefittet ist auf diesen Datensatz und in die Knie geht in dem Moment, wo man jetzt ganz andere Daten oder etwas andere Daten da durchjagt. Genau. Gibt es dann zusätzlich noch... Um vielleicht kurz abzuschweifen, gibt es auch eine Richtung im maschinellen Lernen, gerade die sich mit Adversarial Training beschäftigt, wo man wirklich sagt, man will gezielt das schon im Training benutzen, um Modelle robuster zu machen. Und noch ein nächster Schritt ist, wenn man sogar gezielte Angriffe fährt und zum Beispiel die Gradienten aus einem Netz benutzt, um möglichst wenig Veränderungen das Modell auszutricksen sozusagen, was dann auch hinsichtlich Security natürlich irgendwie ein sehr relevantes Forschungsfeld ist. Also angenommen man hat vielleicht nicht die komplette Kontrolle über die Daten, sondern man hat da irgendjemanden, der zum Beispiel irgendwelche Straßenschilder mit irgendwas bekleben könnte. Klar, wenn man das ganze Schild abschwärzt, dann erkennt man nur noch ein Stoppschild, weil das halt zumindest die Form hat. Aber ja, was ist so das, was das Modell aushält? Genau. Oben die Begriffe, MobileNet, infer, ImageNet. Geläufig oder?

I16

MobileNet V3 Small würde ich jetzt mal vermuten. Das ist das Modell, um das es hier konkret geht. ImageNet ist wahrscheinlich die Modellfamilie, bin ich mir aber jetzt auch nicht sicher.

A1

ImageNet ist ein Bilddatensatz für Bildklassifikation. Das ist einer der größten Datensätze für Bildklassifikation und wurde halt sehr stark durchgemined. Gibt es halt ganz viele Modelle, die auf ImageNet oder für ImageNet trainiert wurden. Genau. Und mit Inference meinen wir einfach, dass das Modell halt nur ausgewertet wird. Also es geht ganz normal. Um praktischen Downstream-Einsatz von diesem Modell. Es geht jetzt nicht darum, das nochmal neu zu trainieren. Deswegen tauchen auch unten keine Metriken auf, die sich mit dem Training beschäftigen. Genau.

I16

Und das, was ich da zwischen noch so sehe, A100x8, das ist einfach Hardware und letztendlich Software, auf der das Modell läuft. Richtig?

A1

Ganz genau. Richtig, genau. Ich nenne das immer das Environment, in dem das Modell betrieben wird. Weil tatsächlich die ganzen Metriken, die unten dargestellt sind, natürlich absolut abhängen davon, in welchem Szenario das Modell eingesetzt wird. Insbesondere natürlich Laufzeit und Energieverbrauch. Ist aber tatsächlich auch so, dass wenn man eine andere Version von TensorFlow installiert, die Qualität halt auch leicht schwanken kann. Weil irgendwelche Algorithmen anders implementiert werden. Genau. Siehst du da irgendwie eine Brücke zu deiner Arbeit? Hast du das Gefühl, dass das was ist, was irgendwie so in deinem beruflichen Alltag, in der Praxis irgendwie hilfreich wäre?

I16

Jein. Also ich glaube, dass das schon hilfreich ist, wenn man irgendwann in den Rahmen des Auswahlprozesses kommt. Was nutzen wir denn jetzt zum Beispiel konkret für ein Modell? Was kommt für uns in Frage? Da hat man, glaube ich, mit so einem Label sehr gute Möglichkeiten zu sagen, was ist uns wichtig? Wie finden wir das hier wieder? Und damit können wir eine gute Auswahl treffen. Und das Problem, was meiner Meinung nach in der Praxis momentan auch sehr stark vorherrscht, ist erstmal, das überhaupt den richtigen Anwendungsfall zu finden. Und ich kann halt nicht irgendein Modell auswählen, wenn ich meinen Anwendungsfall überhaupt nicht kenne. Und da hilft mir das natürlich jetzt erstmal nicht. Also ich glaube, wenn man darüber hinaus ist, dann könnte das sehr gut sein, wenn man sagt, ich habe so einen Katalog, nach dem ich mich richten kann. Wo ich eventuell noch bestimmte Kriterien drin suchen kann. Was ist mir wichtig? Was habe ich für Mindestanforderungen an gewissen Stellen? Dann ist sowas total praktisch.

A1

Ja, genau dafür ist es auch gedacht. Deswegen haben wir ja auch irgendwie in der Interviewstudie geschrieben, dass wir uns halt vor allem mit Entwicklern irgendwie auch zusammensetzen wollen für diese Gespräche. Aber es ist natürlich auch super hilfreich, von anderen Leuten Feedback zu kriegen. Das Design ist ja ein bisschen angelehnt an bekannte Labeling-Formate, also zum Beispiel an die Energielabel aus der EU. Diese Skala sieht ein bisschen aus wie beim Nutri-Score. Das sind also insofern irgendwie auch Sachen, die jedem irgendwie geläufig sind. Gleichzeitig sind wir ja nicht in der Situation, dass jeder einfach im Kaufland um die Ecke irgendwie eine Waschmaschine, also jeder kauft vielleicht eine Waschmaschine irgendwo, aber keiner kauft ja so KI einfach. Sondern das sind ja Entscheidungsprozesse, die bei Firmen liegen in erster Linie. Und ob da jetzt KI drin ist oder nicht in dem Produkt, das interessiert den Endanwender dann nur bedingt. Der kann da auch nur bedingt Einfluss drauf nehmen. Genau, aber unsere Vorstellung ist halt, dass man mit so einem Label vielleicht die Kommunikation vereinfachen kann bei der Entwicklung und Bereitstellung von KI-Produkten. Genau, ein weiterer Vorteil, den man damit hat, ist, dass man auch eine gewisse Form von Vergleichbarkeit schaffen kann. Deswegen hier nochmal ein zweites Label dazu. Kannst du da vielleicht auch einmal kurz zusammenfassen, was du siehst?

I16

Gut, wenn ich mir das jetzt so angucke, würde ich sagen, ich habe natürlich erstmal so einen Vergleich, was im Gesamtfall besser abschneidet, weil ich das sofort oben mit der Skala gut abschätzen kann. Und wenn es mich dann interessiert, was ist mir wichtig, dann gucke ich natürlich unten hin und sehe dann, okay, wenn ich jetzt wirklich eine gute Accuracy haben will, dann würde ich natürlich eher das linke nehmen. Und dann müsste ich aber eventuell damit leben, dass die Power-Consumption halt höher ist. Und das ist dann eine konkrete Entscheidung, die ich treffen muss. Aber das ist halt tatsächlich schon mal was, was einem sehr gut ermöglichen würde, die Sachen zu vergleichen.

A1

Auch irgendwie im Moment ein bisschen schwierig, weil man dann irgendwie sehr, sehr lange irgendwelche Paper lesen muss oder so, in denen die Modelle dann in der Tiefe diskutiert werden. Und insbesondere gibt es ja eigentlich kein bestes Modell. Also das ist ein bisschen kontraintuitiv. Also in der oberen Skala sieht das ja so aus, als ob es ein bestes Modell gibt. Aber das ist ja immer nur gegeben der eigenen Prioritäten. So, wenn man halt sagt, ich will aber das schnellste Modell oder das genaueste Modell oder das robusteste Modell, dann verändert das ja das Ranking komplett. Und das ist auch was, was tatsächlich vielen gar nicht bewusst ist. Also Firmen wollen natürlich immer das Beste. Und das Beste heißt bei denen einfach das, mit dem wir am meisten Geld verdienen. Aber das muss man halt auf technische Metriken übersetzen. Und das ist nicht trivial. Und das heißt, da muss man die Vor- und Nachteile irgendwie besser abwägen. Das ist auch ein bisschen die Motivation hinter meinem Forschungsbereich. Ressourcenbewusstes maschinelles Lernen. Wir haben halt irgendwie jetzt so zehn Jahre Forschung gemacht und immer das neue beste Modell, die neue beste Methode irgendwie vorgestellt. Aber eigentlich heißt das nur, wir haben doppelt so viel Strom verbraucht und jetzt bessere Ergebnisse. Da kann man dann halt auch durchaus mal kritisch hinterfragen. Ist es wirklich besser oder ist es nur anders? Fällt dir irgendwas ein, wo du jetzt sagst, aus deiner Arbeit, aus deiner Expertise heraus, das müsste man anders lösen, auch irgendwie auf Darstellungsebene?

I16

Ja, also was sich mir jetzt halt nicht so wirklich erschließt, ist wie die Wertigkeit in das Gesamtscoring eingeht. Ja, also. Ich würde jetzt persönlich selber auch sagen, ich habe ja auch mit vielen Firmen zu tun, die da deutlich nachhaltiger denken, dass eine hohe Bewertung des Energieverbrauchs auf jeden Fall wichtig ist. Runningtime zum Beispiel ist etwas da würde ich sagen das ist eine schöne Metrik, aber das hängt doch ganz, ganz schwer vom Anwendungszweck ab, welche Relevanz das hat. Und ob das dann schwer in die Wertung eingehen muss oder nicht, weiß ich nicht. Und bei Accuracy ist es genau dasselbe. Es kann sein, dass mir sehr niedrige Genauigkeitswerte gut genug sind. Es kann aber auch sein, dass ich da Anforderungen habe, die halt da immens großen Wert drauf legen, dass die Genauigkeit wirklich gegeben ist. Und bei der Robustheit ist es eigentlich genau dasselbe. Dass ich sagen würde, da hilft mir dann so ein Nebel nicht weiter. Ich glaube sogar, das könnte einen dann schnell erstmal auf eine falsche Spur lenken. Weil das halt nicht wirklich mit dem Anwendungszweck korreliert.

A1

Ja, also das muss auf jeden Fall ganz individuell entschieden werden. Hier ist es jetzt, um das nur einmal zu erläutern, ist es so, dass wir, also wir haben sogar noch ein paar mehr Metriken als die vier, die wir unten darstellen. Also wir haben zum Beispiel uns auch mal Top-5-Accuracy angeschaut. Die ist natürlich stark korreliert mit der Top-1-Accuracy. Da ist die Frage, ist die Vorhersage unter, also das ist eine der fünf am wahrscheinlichsten vorhergesagten Klassen unter der richtigen Klasse. Da hat man dann natürlich irgendwie nochmal eine andere Zielmetrik sozusagen. Aber auch eine, die für viele Unternehmen relevant ist, gerade wenn man so an Recommendation Systems denkt. Die wollen ja in der Regel nicht nur eine Recommendation, sondern das sind die Top-5-Recommendations hier. Das heißt, je nach Anwendungsfall unterscheiden sich auch die Metriken unten und auch die Relevanz davon. Und in unserem Fall haben wir jetzt einfach gesagt, wir haben, wir haben Metriken, die sich irgendwie auf den Ressourcenverbrauch beziehen, zum Beispiel Zeit als Ressource oder Energie. Und wir haben halt Metriken, die sich auf die Qualität beziehen. Innerhalb dieser Gruppen gewichten wir alles gleich und die beiden Gruppen gewichten wir auch gleich. Das ist so das ganz einfache Ranking, was wir jetzt hier machen. Aber tatsächlich ist auch in den Papern, die ich geschrieben habe, die Wichtigkeit von Interaktivität bei solchen Labeling-Verfahren oder auch einem Framework, was das ermöglichen würde, relativ zentral und wichtig. Und vielleicht ist man in einem Runtime-Constraint-Setting, wo man sagt, wir wollen aber Lösungen oder Antworten liefern alle x Sekunden. Dann ist halt Runtime auf einmal sehr wichtig, oder zumindest, dass es unter diesem Threshold liegt. Genau.

I16

Ja, das sind halt so, wo ich mir denke, wenn jetzt eine Bank-Echtzeit-Analyse, Fraud-Analyse fährt, dann brauche ich halt sehr, sehr hohe Laufzeiten. Und dann nehme ich es vielleicht auch in Kauf, dass das viel Energie verbraucht und dass ich viele falsch-positiv Ergebnisse habe und andere wiederum sagen, ich brauche jetzt die absolute Genauigkeit. Und dafür kann es aber ruhig langsam sein, weil ich lasse das nachts auf unserem Data Lake los und der holt sich da 10.000 Datensätze raus. Und wenn er die ganze Nacht für braucht, ist mir egal. Also das ist dann halt wieder sehr stark der Anwendungszweck, der da mit reinkommt.

A1

Ja, auf jeden Fall. Genau. In deiner Arbeit, wenn du über KI und Machine Learning irgendwie redest oder auch Firmen an der Stelle begleitest, welche Informations- oder Kommunikationsform nutzt du da?

I16

Wenn ich die begleite?

A1

Also, genau, wo kriegst du deine Informationen zu dem Thema her, wenn du dich selber bildest oder halt auch andere darüber informieren willst? Also das hier könnt ihr nochmal nachlesen.

I16

Also ich habe diverse Agenden und Newsletter, wo ich im Prinzip immer wieder so mehrfach die Woche ein paar interessante aktuelle Papers zugeschickt bekomme. Und das ist für mich eine wichtige Informationsquelle. Es gibt diverse Menschen, die sich halt natürlich auch damit auseinandersetzen, wo ich dann sage, von denen habe ich auch Newsletter abonniert, Blogs abonniert, informiere mich da drüber. Und was da so Sache ist. Und womit die sich auch beschäftigen, klar, Firmenseiten, gerade so im LLM Bereich, da auch wirklich mal regelmäßig zu gucken, was erzählt OpenAI, was erzählt Entropic zu den aktuellen Sachen, an denen sie arbeiten. Ich denke, das ist halt auch ein wichtiges Thema. Und dann halt auch in anderen Bereichen, also in den Querschnittsbereichen zu gucken, was sagen die Industrieverbände momentan dazu? Wenn es um die Sicherheit geht, was sagt das BSI dazu? Das sind halt auch so Sachen, wo ich regelmäßig mal draufgucke, um zu gucken, was passiert in dem Bereich überhaupt?

A1

Mhm. Was ist, wenn du jetzt an einem ganz konkreten, also Paper ist ja schon dann ziemlich konkret. Also die anderen Sachen sind natürlich auch völlig valide. Geht ja dann aber eher so in den Überblick behalten. Wenn es jetzt um konkrete Modelle geht oder um ganz konkrete Lösungen, gibt es dann neben Paper noch was anderes, wo du irgendwie dich informierst?

I16

Ja, also das sind dann schon so Sachen, wo ich mir denke, okay, hat jemand in einem bestimmten [schlechte Verbindung] kuratiert Listen zusammengestellt, Modellisten, wo man sagt, du musst dich nicht durch halb HuggingFace durchklicken, um die gängigen Modelle in einem Bereich kennenzulernen, sondern es hat schon mal jemand anders gemacht und das ist ja in dem Bereich heutzutage sehr schön, dass fast alles jemand anders schon mal gemacht hat und damit kommt man eigentlich relativ weit ansonsten ist natürlich auch wirklich Stichwortsuche, ein ganz wichtiges Thema, wenn man sehr konkret weiß, was man macht. Ich hatte halt letztes Jahr zum Beispiel ein Projekt, wo es relativ viel um Text-to-Speech und Sprachsynthese ging, was halt auch ein unglaublich weites Feld ist mittlerweile. Das war relativ problematisch, da Sachen rauszufinden, weil das auch immer, was man dann über die Suchmaschinen findet, sehr oft davon abhängt, was gerade trendet und das müllt einen dann auch die Suchergebnisse zu, wenn man gucken möchte. Mich haben halt wirklich Open-Source-Modelle an der Stelle interessiert, die frei einsetzbar sind, die auch mit relativ wenig Ressourcen auskommen. Das war eine Analyse, die war relativ schwierig, muss ich tatsächlich sagen.

A1

Hast du am Ende was gefunden, was deinem Einsatzzweck irgendwie genügt?

I16

Nee. Also ich habe Sachen gefunden, die haben für einen prototypischen Betrieb erst mal gereicht. Die große Herausforderung war daran aber auch, es ging halt um Deutsch als Sprache im Text-to-Speech und das macht dann auch besonders schwierig, weil man da nicht sagen kann, na, dann nehme ich mal so ein Takotron oder irgendwas, was davon abgeleitet ist und das funktioniert dann, sondern ich glaube, wir haben da zum Schluss auch die Solero-Modelle gesetzt, weil die noch relativ gute Ergebnisse gebracht haben, aber wenn ich jetzt noch mal nachgucken würde, würde ich wahrscheinlich auch schon wieder völlig andere Dinge finden.

A1

Ja, mit Sicherheit. Wir haben uns im Kontext von der Studie natürlich auch irgendwie angeguckt, ob wir zumindest das Transkribieren zum Teil automatisieren können und benutzen da jetzt einen Ableger von Whisper. Whisper ist so ein LLM dafür. Ein Ableger davon, der von der RWTH Aachen entwickelt wurde, der auch lokal halt irgendwie laufen kann. Das war halt für uns ein Kriterium natürlich für die Auswertung. Damit sind wir ganz zufrieden tatsächlich. Ja, XXXXX kann da besser was zu sagen. Die geht nämlich über alle Transkripte noch mal drüber und korrigiert alles, was die KI nicht hinbekommt. Aber soweit tut's das schon, oder?

A4

Ja, also manchmal sind ein paar Absätze nicht dabei oder so, aber ansonsten funktioniert das eigentlich echt sehr gut. Also ich bin auch erstaunt, weil manche Worte werden halt erkannt, wo ich mir denke, das hätte ich nie im Leben rausgehört. Und dann andere Worte, wo ich mir denke, es ist offensichtlich, sind dann einfach falsch. Aber ansonsten erleichtert das die Arbeit zu 100%.

A1

Ja, das ist wieder das Thema Robustness an der Stelle. Also es ist halt nicht ganz deterministisch, welche Aussprache funktioniert und welche nicht. Naja, wir haben uns auf jeden Fall natürlich im Rahmen unserer Studie auch so ein bisschen gefragt, welche Alternativen gibt es im Moment zu Labeling, die schon am Markt sind, die eingesetzt werden, was ist im Moment der Workflow bei den Leuten. Das ist so die Sache, die Sachen, die wir irgendwie gefunden haben. Also einmal links zum Beispiel ganz klassisch gibt es auch ein Paper zum Mobile-Net, das sind dann 40 Seiten, die man sich anschauen kann, wenn man wirklich in der Tiefe verstehen will, wie dieses Modell funktioniert. Rechts ist die zugehörige Model-Card so angegeben auf Hugging Face. Da ist ja dieses Model-Card-Format, was ursprünglich auf dem Google-Paper basiert, inzwischen auch angekommen, dass man einfach so die wichtigen KPIs und Informationen zusammenfasst. Hier jetzt gemessen auf einem Galaxy von Samsung. Das heißt, Mobile-Net natürlich auch als Modell insbesondere ausgelegt für den Einsatz auf mobilen Endgeräten, weil es halt relativ wenig Parameter braucht oder relativ wenig Speicher braucht und relativ schnell ist. Genau, dann haben wir passend dazu natürlich noch die Übersicht von Papers with Code. Das ist eine Seite, die ganz praktisch ist für alle Leute, die auch wirklich auf die Code-Ebene wollen, weil die, da kann im Endeffekt jeder freiwillig sein Paper mit seinem Code irgendwie verlinken und man kriegt dann für alle angegebenen Modelle zum Beispiel aus den Papern auch wieder eine Liste von allen Papern, in denen dieses Modell benutzt wird und so. Ist halt so eine offene Datenbank, sage ich mal, zu dem Thema. Ähnlich offen sind Blogposts. Gibt es inzwischen auch für alle großen Technologien irgendwie auf Medium bzw. Towards Data Science entsprechende Blogposts. Das ist natürlich immer das Risiko bei solchen Sachen, da kann halt auch jeder alles reporten und man hat halt so gar keine Trustworthiness dahinter, gar keine Vertrauenswürdigkeit. Dann haben wir links noch die aktuelle Dokumentation von Keras, also so wie das Modell dann in TensorFlow Keras zur Verfügung steht und genutzt werden kann. Auch wieder insbesondere für Entwickler relevant, weil man dann direkt die Software-Schnittstelle da auch hat. So ein paar KPIs werden da meistens auch reported und rechts noch ein Factsheet, nicht für MobileNet tatsächlich, weil dieses Factsheet-Format von IBM eigentlich nur für IBM-Produkte genutzt wird und für IBM-Modelle. Das ist auch patentiert und so, ist aber sehr ähnlich wie die Model Cards eigentlich. Wenn man sich das ja jetzt in Gegenüberstellung mal anschaut, kannst du da vielleicht nochmal drauf eingehen, wo du Vor- und Nachteile siehst und wie dieser Vergleich von diesen verschiedenen Sachen für dich so ein bisschen ausschaut?

I16

Ja, gut. Also bei Papers ist natürlich so der Punkt, die sind vor allem dann relevant, wenn es in die Tiefe geht, wobei ich da auch sagen würde, so ein Paper mal eben schnell in LLM reinstecken und sich eine Zusammenfassung generieren lassen das ist auch Teil meines Standard Workflows, damit ich überhaupt sehe macht das Sinn sich damit auseinanderzusetzen und das zu lesen oder kann ich das gleich links liegenlassen? Sodass man da halt auch nochmal Zeit sparen kann. Aber das ist halt trotzdem schon in aller Regel was, das ist relativ komplex. Die Modelcards, sind halt was, was ich auch regelmäßig benutze. Da ist halt einfach das schöne, dass das klassifiziert ist, das heißt wenn ich eine bestimmte Problemstellung habe und weiß nach was für Klassifizierungen ich suchen muss, dann sind die ganz gut um entsprechende Lösungen zu identifizieren. Bei den anderen Sachen, ja das IBM Ding das kenne ich jetzt nicht, das kann ich nicht so gut beurteilen wie wertvoll das ist. Bei der Developerdokumentation würde ich sagen das ist ja eher dann für den praktischen Einsatz, wie kommt es dann im größeren Kontext zum Einsatz, wichtiges sehe ich in einem späteren Prozessschritt(?). Blogposts finde ich total wichtig, sind für mich aber eher dann Sachen wo ich sage da geht es um allgemeine Informationen. Ich würde jetzt nicht nach spezifischen Modellen über Blogposts suchen. Also sowas wie Towards Data Science habe ich auch abonniert und lese die Sachen da aber ich überspringe solche Artikel die sehr spezifisch sind meistens wie ich da auch den Eindruck habe, das ist relativ biased was da drin steht. Und da würde ich dann eher tatsächlich in die zugehörigen Paper gucken um mir da einen Überblick zu verschaffen. Die Frage ist ja immer die der Fragestellung, die ich eigentlich habe. Brauche ich jetzt eine Lösung, dann würde ich an der Stelle sagen, dann ist ein gehen über eine Modelcard der erste Schritt, der irgendwie Sinn macht, wenn man halt irgendwie so eine Datenbank hat wie das was wir da jetzt oben rechts sehen. Das könnte natürlich auch nochmal total hilfreich sein aber da, das Models with Code natürlich sehr davon abhängt, ob die entsprechenden Researcher oder Entwickler die Sachen tatsächlich dort auch zur Verfügung stellen sind das natürlich auch Informationen die relativ unvollständig sein können. Wenn ich dann so grob habe, was überhaupt in Frage kommt, dann würde ich sagen kommen halt wirklich sowas wie paper oder nochmal die zusätzlichen Dokumentationen, die vielleicht auch in der Modelcard dann weiter unten zu finden ist zum Tragen, zu sagen ist irgendetwas geeignet oder nicht. Und das wäre jetzt zum Beispiel auch genau der Punkt, wo ich sagen würde wo ein AI-Label dann schon Sinn machen würde, weil das natürlich helfen könnte eine gute Vorauswahl zu treffen. Wenn man jetzt fünf unterschiedliche Sachen identifiziert hat die überhaupt in Frage kommen und fünf würde ich jetzt sagen ist eine sehr niedrige Zahl, weil die Zahl der Modelle die einem zur Verfügung stehen für bestimmte Anwendungszwecke die wird ja auch immer größer. Dann kann noch sowas helfen um halt wirklich die Zahl der infrage kommenden Modelle relativ schnell zu reduzieren. Also da käme das dann meiner Meinung nach gut zum Einsatz und danach erst wenn sowas passiert, [schlechte Verbindung] so Sachen wie die Entwicklungsdokumentation tatsächlich eine Rolle

A1

Ja wo du jetzt selber sagst, also so Entscheidungsprozesse, das werden immer mehr Modelle die dafür geeignet sind. Ist natürlich prinzipiell richtig, ist aber teilweise im Moment, also wir haben das auch in den Interviews gemerkt, dass das Verständnis von was ist, denn ein KI Modell halt sehr unterschiedlich ist. Es gibt halt auf der einen Seite dieses alles was mit großen Bild oder halt noch viel größeren Sprachmodellen zusammenhängt und da ist natürlich auch gerade das wo am meisten Aufmerksamkeit irgendwie hin ist. Ganz oft, wenn man mit Firmen arbeitet, und das hast du ja auch schon angeschnitten, haben die aber halt auch viel kleinere Datensätze wo halt selbst neuronales Netz allgemein in der Regel schon völlig overpowered ist weil man halt irgendwie mehr Parameter als Trainingspunkte hat. Ja kann man halt irgendwie alles lernen und naja generalisiert aber halt gar nicht mehr weil man alles auswendig lernen kann Wie oft erlebst du denn solche… Also erlebst du oft solche Entscheidungsprozesse?

I16 00:040:04

Ja, schon, weil da spielt dann auch wieder so ein bisschen diese allgemeine Unwissenheit mit rein, weil dann Leute sagen, Mensch, ich habe da so ein total tolles White Paper gesehen oder ich habe so ein Signal gehört von OpenAI und die sagen, man kann das und das und das und alles damit machen und man muss da ja nur mal eben die API ansprechen und damit haben wir alle Probleme gelöst und wenn man dann auf die Problemstellung drauf guckt, dann sieht man ja, das ist ein sehr spezifischer Fall. [Schlechte Verbindung]… Besser irgendwie ein einfaches Modell dranzusetzen. Also ich habe das schon oft genug gehabt, dass Leute versucht haben irgendwie mit neuronalen Netzen was zu lösen wo so ein K-Means clustering völlig ausgereicht hätte und auch die sinnvollere Alternative gewesen wäre, weil, wie du gerade schon sagtest, wenn die Datenmenge so klein ist, hast du sofort overfitting und damit erkennst du dann eigentlich gar nichts mehr außer den Fällen die du antrainiert/handtrainiert hast und das ist natürlich auch blöd. Da geht es dann hauptsächlich darum die Leute zu (unverständlich) und zu sagen, schießt hier nicht mit Kanonen auf Spatzen, sondern guckt euch doch an was ihr an der Stelle macht und wir können diese Recherchearbeit, welche Modelle richtig richtig gut sind eigentlich sparen, wenn eure Andwendungszwecke so eng ist, dass wir wahrscheinlich mit einem selbsttrainierten Modell viel schneller zum Ziel kommen oder mit einem klassischen ML-Algorithmus, der sehr viel einfacher gestrickt ist

A1

Ja, genau. Das ist also sehe ich genauso. Ich wollte da nochmal nachhaken, weil ich glaube, diese... Da haben wir jetzt so viele verschiedene... Also man hat halt immer super viele verschiedene Optionen. Man kann halt auch alles mit Random Forests machen oder mit K-Means oder halt mit neuronalen Netzen und dann ist aber noch wieder nicht gleich neuronales Netz, weil es halt zig verschiedene Layer-Typen gibt und man kann da halt irgendwie auch... Man kann auch sehr einfache, interpretierbare Neuronale Netze konstruieren mit so ein paar Hidden Neuronen und das war's. Dann hat man auch irgendwie noch überschaubar viele Parameter und kann jede Entscheidung einigermaßen nachvollziehen. Aber das, womit halt Deep Learning (unverständlich) ist halt einfach riesige Datenmengen und unfassbar komplexe Aufgaben, die dann bewältigt werden.

I16

Ja, das ist aber auch halt wirklich zum Teil das Problem, das heutzutage gesagt wird, ey, mit KI kann ich ja alles lösen. Ja, mag sein. Erstmal würde ich dem widersprechen, aber selbst wenn es so wäre, ist das nicht sinnvoll. Und das ist halt genau dieses Problem. Ich muss nicht mit Machine Learning an irgendwas rangehen, wo ich mit linearer Optimierung zu einem besseren Ergebnis komme. Ja, das ist halt sowas, wo ich dann sage ne. Man muss auch auf den Anwendungszweck draufgucken.

A1

Ja, genau. Und dann hat man bei neuronalen Netzen, bei so großen KI-Modellen, verliert man halt dann dafür Themen wie Interpretierbarkeit der Ergebnisse, Robustheit der Ergebnisse, weil minimale Änderungen in den Daten dann wieder an irgendeinem Punkt komplett andere internen Prozesse irgendwie auslösen. Das wäre jetzt auch nochmal das, wo wir so ein bisschen am Ende drauf zu sprechen kommen können. Vertrauenswürdigkeit, wir haben vorhin über Fairness und Bias geredet, haben wir jetzt so natürlich auf dem Label im Moment noch gar nicht drauf. Kannst du da vielleicht ein bisschen was zu sagen? Wie siehst du diese Bereiche?

I16

Also Vertrauenswürdigkeit hat ja relativ viele Aspekte, denke ich. Das ist sicherlich erstmal so, das kann ich dem vertrauen, was die KI mir ausspuckt, oder habe ich da eventuell ein Problem, dass, also gerade wenn ich jetzt an große Modelle denke, dass wir Halluzinationseffekte haben, wo die KI sich einfach irgendwas ausdenkt, ja, salopp gesagt, oder habe ich vielleicht auch einen ganz klassischen ML-Algorithmus, der einfach keine besonders gute Accuracy hat, weil er nicht gut trainiert ist, und wo ich dann plötzlich feststellen muss, dass die Ergebnisse, die da rauskommen, falsch sind, und ich halt eben damit nicht leben kann. Das ist ja ein Aspekt davon. Ein weiterer Aspekt ist aber auch, glaube ich, was passiert mit den Daten weiter, und führen diese Daten, zu einer Verantwortung, und trägt diese Verantwortung irgendein Mensch, oder trägt die dann plötzlich niemand mehr? Und ich glaube, das ist ein weitaus größerer Bereich, der da dran hängt. Das ist ja was, was wir jetzt schon sehr sehen, wenn darüber diskutiert wird, sollen KI-Inhalte irgendwie gekennzeichnet werden, oder nicht? Können wir das überhaupt automatisch lösen, oder nicht? Was passiert, wenn wir das nicht tun, und wer ist dann dafür verantwortlich, wenn dann irgendwas Schlimmes damit passiert, wenn ja zum Beispiel politische Kampagnen, dadurch gezielt manipuliert werden, oder man halt gezielt falsche Informationen verbreitet? Das sind halt auch so Sachen, von denen ich sagen würde, dass die natürlich in diesen Vertraulichkeitsaspekt sehr stark reinspielen. Und das ist aber Anwendung dieser Systeme, und das ist jetzt nichts, was den Systemen irgendwie inherent ist. Also das könnte man ja jetzt nicht an den Systemen selber messen.

A1

Ja. Genau, also ganz viel davon ist halt schwierig, auf Messwerte runterzubrechen, und auch so Fairness zum Beispiel, von einem KI-Modell, ist halt sehr schwer zu beziffern. Ist aber natürlich auch was, wo wir im Kontext von dem Label natürlich drüber nachdenken, ob man das irgendwie mit aufnehmen kann. Man könnte dann zum Beispiel ja irgendwie sagen, dass der Datensatz vielleicht von Experten hinsichtlich Fairness oder Bias gescreent wurde, und solche Themen halt überhaupt bewusst behandelt wurden bei der Kuratierung und des Trainings des Modells. Genau, also da sind wir halt auch einfach so ein bisschen, das ist so eine offene Frage, die wir uns auch selber stellen. Wie kriegen wir das irgendwie hin? Wichtig ist ja auch im, also wir hatten ein paar Gespräche, wo Leute dann auch gesagt haben, naja, das ist ja so ein bisschen wie so ein Zertifikat dann auch, wenn man da jetzt konkrete Werte drauf hat. So wie das EU-Label für Energieverbrauch, das ist ja auch eine Form von Garantie dafür, wie dieses Produkt funktioniert. Und dem vertrauen die Leute dann einfach und sagen, na gut, okay, es hat irgendwie eine Energie, die Effizienzklasse von B, dann ist das schon recht gut und A wäre halt besser so. Ist halt die Vertrauenswürdigkeit bei so einem Label überhaupt gegeben? Oder würden Leute das halt anzweifeln? Ich weiß nicht, wie siehst du das?

I16

Ich glaube, da muss man halt kommunizieren, was das Label nicht aussagt. Also wenn so ein Label, und ich glaube, das ist tatsächlich sehr, sehr schwierig, keine Aussage darüber macht, wie fair ein Modell ist oder wie vertrauenswürdig so ein Modell ist, wie sicher ein Modell ist, und man das von vornherein ausklammert, dann kann das an dieser Stelle halt schon helfen, aber da muss halt quasi unter dem Label sofort ein Warnlabel drunter, dass eben diese Sachen dadurch nicht abgedeckt werden und nicht betrachtet werden. Und da hat man dann psychologisch natürlich wieder das Nicht-Drinstehen, was dann gerne mal überlesen wird. Und das macht es dann vielleicht nochmal ein bisschen schwierig. Also generell würde ich an der Stelle schon sagen, das ist ja auch beim NITRI-Score zum Beispiel so, dass die Leute dann sagen, ach hier, NUTRI-Score ist ja wunderbar grün, den nehme ich jetzt einfach mal, und gucken sich die Inhaltsstoffe dann gar nicht mehr an. Und ich glaube, da muss man halt schon so ein bisschen aufpassen und die Leute dann wiederum darauf lenken und zu sagen, das ist schön, um eine Erstauswahl zu treffen, aber guck dir danach die Details an, weil da passiert vielleicht nochmal einiges, was für dich wesentlich ist.

A1

Hast du denn das Gefühl, wenn jetzt jemand, also man würde jetzt diese Aspekte rauslassen, bleibt eher auf diesen technischen Messwerten, glaubst du, wenn du das jetzt einsetzen würdest, also das Software-Framework, um diese Labels zu generieren, dass du schon Public auf GitHub, weil das habe ich irgendwie im Rahmen meiner Promotion entwickelt. Man kann da einfach praktisch eine Tabelle reinladen von Messwerten und die generiert einem dann so ein Label für alle Modelle, die man getestet hat. Glaubst du, dass Leute, wenn man das jetzt so, wenn du das jetzt zeigen würdest bei dir in deinem Alltag, dass Leute sagen würden, ja, gut, das steht auf dem Label drauf, dann wird das schon stimmen? Oder würden die fragen, wo kommt das denn her? Würden die denn diesen Informationen vertrauen?

I16

Ich glaube, ja. Das liegt aber auch daran, dass ich halt nicht dieses Label zum Beispiel einem Entscheider zeigen würde, sondern ich würde das Label einem Entwicklungsingenieur zeigen. Ja, weil ein Entscheider kann mit den Sachen, die hier von dem Label momentan gemessen werden, nichts anfangen. Und das ist halt so ein Problem. Der bräuchte im Prinzip so ein Business Value Label.

A1

Ja, und Kosten Nutzen.

I16

Und das geht hieraus natürlich nicht hervor. Ein Entwicklungsingenieur würde sagen, ah, okay, hier sehe ich Energieverbrauch und Robustness und sowas, damit kann ich was anfangen, das gucke ich mir an und dann kann man da natürlich nochmal fragen, ja, wo kommt denn das her? Und dann kann man aber sagen, hier gibt es einen offenen Berechnungsalgorithmus, den kann man einsehen, den kann jeder betrachten und auf der Basis erfolgt halt entsprechend die Einordnung. Ich glaube, das ist halt schon was, was relativ vertrauenswürdig ist.

A1

Also ich meine, muss ich jetzt natürlich auch sagen, aber natürlich ist das so vertrauenswürdig, als dass man sich einfach meine Software angucken kann. Und dadurch, dass ich ja nicht selber die Messwerte und Mittel, sondern nur Berichte habe, ist da auch, also ich kann da auch nur bedingt viel rumtricksen. Also das ist, ja, meine Software ist offen, insofern ist die irgendwo auch vertrauenswürdig. Ob Leute dem vertrauen oder nicht, ist davon aber auch nicht unbedingt abhängig. Aber das ist, ja, also wahrscheinlich ohne das untere Teil wäre das wahrscheinlich was, was auf so einer Entscheidungsebene fast schon interessanter wäre. Dann steht da nur so Gesamtwertung, okay, passt. Aber sobald es diese technische Information mit dabei hat, ist das vielleicht etwas, was Leute auch irritieren, verwirren könnte.

I16

Ja, also da stehen hier jetzt natürlich sehr viele Sachen drauf, von denen ich sagen würde, ah, okay, Datensatz, mit dem das trainiert worden ist, Hardware, auf der das läuft, Softwareversion, mit der das läuft, damit kann der Entscheider natürlich überhaupt nichts anfangen. Andererseits, wenn man diese ganzen Details weglässt, dann finde ich, ist es halt wieder relativ schwierig, dem Label eine Gesamtbedeutung zuzuweisen, weil das A, was hier jetzt steht, das ist halt irreführend, weil wenn Accuracy für mich total wichtig ist, dann ist das definitiv das falsche Modell.

A1

Ja, da wäre ja die Idee, dass derjenige, der die, der im Vorfeld praktisch die Modelle entwickelt, auch das natürlich gegeben der Prioritäten in dem Use Case irgendwie tun würde. Und, ähm, ja, dann gegeben dieser Gewichtung halt das Gesamtscoring irgendwie ermitteln lässt. Wir sind fast durch mit der Stunde. Deswegen würde ich tatsächlich, glaube ich, an der Stelle das Interview zum Ende bringen. Danke für die Aufmerksamkeit auf jeden Fall. Die Folien hier würde ich dir im Nachgang schicken, falls du da nochmal durchblättern willst oder das irgendwie...