I2

Ich weiß nicht, ob du, XXXXXXXXX, das parallel auch aufnehmen kannst. Ich hatte das auch mal. Je nach Zoom-Version geht das. Dann seid ihr sicher, dass nichts verloren geht.

A3

Ja, also wir haben immer Doppelaufnahmen. Ich habe jetzt sogar, weil ich kein Headset benutze, mache ich das einfach auf meinem Handy.

I2

Die Qualität ist okay. Das kann man gut machen.

A3

Genau, also ist auf jeden Fall sicher, hoffentlich. Okay, dann würde ich gerne einsteigen und erst mal sagen, stell du dich doch mal vor. Was ist dein Background?

I2

Ja, mein Name ist XXXXXXX. Ich bin von Hintergrund Wirtschaftswissenschaftler, bin Bachelor und Masterstudent, bin im Fokus über Technologie und Innovationsmanagement. Bin seit 2018 an zwei XXXXXXXXXXXXXXXXXXXXX insgesamt gewesen und habe eigentlich immer im Fokus Technologie und Innovationsmanagement gemacht und habe über drei Jahre als KI-Manager gearbeitet und eben da auch mich mit KI und generativer KI beschäftigt, insbesondere aus der Anwendungsperspektive. Habe aber auch an Studien mitgewirkt zum Thema Sprachmodellentwicklung, bevor Chat-GPT aufgekommen ist, also die Transformer-Modelle. Und ja, habe da ein gutes Grundverständnis von neuronalen Netzen, Deep Learning, Reinforcement Learning, auch von der Seite her, natürlich nicht in der reinen Entwicklungsseite. Und beschäftige mich immer noch mit dem Thema, insbesondere forsche ich im Bereich Sprachassistenzsysteme und generative KI und grundsätzlich auch, ich sag mal, ja, also Decoder-Systeme für beispielsweise Fachanwendungen von Fachsprachen, medizinische Anwendungen. Und darüber hinaus arbeite ich noch in Projekten, die, ich sag mal, Big Data-Themen betreffen, auch von Data Lakes und dann späterer Algorithmen in Wertsetzung. Wie man das heute so schön nennt, also die zukünftige Anwendung von KI und vielleicht dort auch den ein oder anderen Gen-AI-Use-Case.

A3

Okay, sehr gut. Dann hast du auch schon ein paar Nachfragen von mir auf jeden Fall abgehakt. Deine Vorerfahrung mit KI ist also eher nicht wahrscheinlich aus dem Studium, sondern wirklich dann erst in der Praxis, jetzt so in die Arbeit im XXXXXXXXXX gekommen?

I2

Ja.

A3

Okay. Und strebst du denn den PhD an gerade oder bist du auf dem Master zufrieden, sage ich mal?

I2

Ne ne, also ich bin im vierten Jahr im PhD im Bereich Technologie und Innovationsmanagement an der XXXX XXXXXX. Das mache ich in Teilzeit. Also ich bin Teilprojektleiter auf einem großen Forschungsprojekt und den Hauptteil meiner Arbeit bin ich in der IT-Beratung und Entwicklung tätig und unterstütze eben in allen möglichen Industrien, in allen möglichen Bereichen das Thema Big Data, Data Advisory und eben auch künstliche Intelligenz, wo eben solche Themen auch relevant sind, die ihr angeführt habt. Labeling, Zertifizierung, Vertrauenswürdigkeitsbildung, Responsible AI.

A3

Okay, das hört sich nach einem ziemlich abwechslungsreichen Arbeitstag an. Kannst du trotzdem so einen typischen arbeitsalltäglichen Prozess beschreiben? Was machst du häufiger, gerade in der Entwicklung und in der Bereitstellung von KI-Produkten, wenn es darum geht?

I2

Also einen klassischen Tag habe ich so gar nicht. Also wenn ich das als Woche beschreibe, kann ich sagen, ein Tag ist Forschung, Projektmanagement, Vorbereitung von klinischen Studien, so User-Interaktionen, Testings, dann sowas wie Anträge schreiben, dann ist das mal sowas wie so BIT-Vorbereitung. Das bedeutet, man möchte eine Angebotsproposal dem Kunden näher bringen, dann ist das tatsächlich operative Beratungsarbeit, Prozessanalysen, dann sind das vielleicht interne Sachen, sowas wie Qualitätsmanagement, Vorbereitung zum Einsatz von generativer KI in Verwaltungssoftware, sowas wie in Conference Tira, also ganz, ganz unterschiedlich. Und dann die inhaltlichen Projektthemen sind, wie gesagt, branchenübergreifend. Das ist von, welchen strategischen Einsatz möchte man oder welche strategische Rolle möchte man künstlicher Intelligenz geben im Bereich generativer KI. Oder ganz konkret Dinge, die noch gar nicht, ich sag mal, den Reifegrad haben, irgendwelche Algorithmus-Verarbeitungen zu machen, sondern da geht es ja eher um Data-Analyse-Projekte. Also das ist so die Bandbreite, die hoffentlich klar genug ist für euch.

A3

Okay, ja. Was ist so der Teil, der am nächsten noch an dieser Entwicklerperspektive drin ist? Beziehungsweise arbeitest du auch manchmal mit Code oder ist das eher Architektur ein bisschen größer gedacht?

I2

Also wir haben meistens in agilen Projekten so eine Mischung aus, ich sag mal, Proxy, Programmmanager, Product Owner und Development Team und da arbeite ich immer als Bindeglied, also ich schaue mal auch mit in den Code rein, ich muss aber nicht selber programmieren, sondern bin eher auf der Perspektive wie in einem Dreieck, also wenn das so die Anwendungsperspektive ist, dann ist das die Anforderungsperspektive und das ist die erste Entwicklungsperspektive und je nachdem, ob Dinge gut laufen oder schlecht laufen, muss man da noch tiefer reingehen oder nicht, aber ich bin nicht derjenige, der Code schreibt, sondern ich muss das Vertrauen haben, was die Kolleginnen und Kollegen machen, dass das Hand und Fuß hat, um eher diesen problemlosen Ansatz zu verfolgen, dass wir schauen, was ist jetzt sinnvoll, um vielleicht so ein Troubleshooting zu betreiben oder um nächste Anforderungen abzubilden, um die dann auf einer Entwicklungsebene runterzubrechen.

A3

Okay. Und was sind so, vielleicht auch in dieser Konstellation, dieses Dreieck, was du gerade genannt hast, was sind da so die größten Herausforderungen, die euch so im Alltag entgegenkommen?

I2

Also bei Projekten grundsätzlich ist es erstmal eine Mischung aus, sprechen alle Abteilungen miteinander, also sprich Business mit IT, ist man da sozusagen aligned, sind es Projekte, die aus einem speziellen Resort kommen oder sind es übergeordnete Projekte, also die fehlende Interdisziplinarität ist, finde ich, auf einer Hürde, dann ein fehlender Pragmatismus zum Technologie-Einsatzbeurteilung und vor allem auch zu, ich sag mal, den organisatorischen Apparaten möglichst gering zu halten, was in der Praxis nicht so oft gemacht wird. Das bedeutet, es wird sehr oft in Hürden gedacht oder in Limitationen oder vielleicht auch in bürokratischen Hindernissen. Also ich will gar nicht sagen, was die DSGVO, sondern wirklich... Können wir das so machen? Ist das technisch möglich, anstatt, ich sag mal, iterativ vorzugehen und kleinere Bausteine zu machen, sozusagen Baby-Steps? Wird das erstmal alles durchdacht? Und ich glaube, das kommt aus der... Also wenn ich noch den Satz zum Interpretationsraum bekomme, ich glaube, das liegt daran, dass wir in der Wirtschaftsinformatik sehr stark immer in diesem Business-Process-Modelling-Prinzip gearbeitet haben und vielleicht auch etwas ältere Kolleginnen und Kollegen gerne eben diese Prozessmodellierungen haben wollen, erstmal muss der ganze Prozess durchexerziert sein. Obwohl, ich sag mal, Technologieeinsatz, insbesondere so algorithmische Verarbeitung oder GenAI vielleicht, diese Prozesse deutlich verschlanken könnte, und hier es einfach so eine Diskrepanz gibt von den Sichtweisen her und meistens eher die, ich sag mal Traditionelle Sicht, durchgedrückt wird und das ist, finde ich, so eine Haupthürde in vielen Bereichen.

A3

Ja, ich meine, das ist ja ein legitimes Problem aus der Wirtschaftsinformatik, glaube ich, diese verschiedenen Sichtweisen. Ich würde gerne nochmal auf einen Punkt, den du in der Mitte genannt hast, zurückkommen. Und zwar die Technologie-Einsatzbeurteilung. Darum geht es ja auch. Oder zumindest dieses Problem wollen wir ein bisschen mit unserer Vision lösen. Kannst du da nochmal drauf eingehen, was genau hier die Hindernisse sind, die du wahrnimmst?

I2

Also das ist sehr spezifisch für Technologien. Also wenn du gerne auf jetzt KI und Gen-AI kommen möchtest, dann bitte. Okay, also ich finde, da muss man nochmal ganz klar unterscheiden vor dem aktuellen KI-Summer, den wir jetzt seit zwei Jahren haben und davor. Also ich würde sagen, jetzt ist generative KI das Thema, das überall durchs Dorf gejagt wird. Und ich habe jetzt auch schon einige große Investmentbanker sagen hören, dass Gen-AI nicht das verspricht, was es machen soll. Also eine wahnsinnige Effizienzsteigerung und eigentlich dadurch auch eine Produktivitätserhöhung, die so nicht eingetreten ist, was totaler Humbug ist, weil die Technologie erst so kurz eigentlich im richtigen Einsatz ist oder auch in der Einsatzverprobung. [3 SEKUNDEN FEHLEN] Und schon so langsam wieder ab, wenn ich das mal so sagen kann. Und das merkt man auch in Projekten. Also ich habe auch schon öfters die Frage gehört, ja, was kommt dann nach KI? Was ist denn das Nächste nach KI? Und ich finde, die Frage ist auf der einen Seite natürlich verständlich, weil das so polarisiert war in den letzten Monaten und vielleicht auch anderthalb Jahren, aber auch total hirnrissig, weil wir noch in der Entwicklung relativ weit hinten dran sind mit dem Thema. Bedeutet, früher KI schwierig, Einzelprojekte. Man probiert mal was aus, man ist sehr explorativ unterwegs. Man hat vielleicht Forscherinnen und Forscher, die dabei sind und testen das. Also das, was eigentlich jetzt Quantum Computing ist. Und Gen AI als Stellvertreter Technologie für Künstliche Intelligenz im Ganzen wird jetzt schon wieder so ein bisschen stiefkindlich behandelt, weil das Wertversprechen der Effizienzerhöhung noch nicht so eingetreten ist. Und eben auch, was ja, glaube ich, mit euer Thema ist für Labeling, ein Vertrauensverlust da ist, weil, ja, auch bekannt geworden ist, dass mit Daten trainiert worden ist, die vielleicht gar nicht öffentlich sind, es nicht Freigaben erfolgt und nicht mehr Erlaubnis gebeten wurde und der Bereich Halluzination ja total relevant ist, was aber für die allermeisten Anwendungsfälle, die nicht in kritischen Bereichen sind oder in kritischer Infrastruktur vorhanden sind, erstmal zu vernachlässigen sind. Und es gibt mittlerweile viel bessere Modelle, beispielsweise Perplexity für Research-based Gen AI, viel besser als Open AI. Die sind ja sogar, die sind ja live ans Internet angebunden, die können ja sogar die Quellen, die ihr vorlesen. Aber diesen Vertrauen zurückzubekommen, das, finde ich, merkt man schon in Projekten. Und was noch dazu kommt, es fehlt eine strategische Ausrichtung für diese ganzen Dinge. Es wird immer Technologien aus der Fachbereichsecke oder auch von dem CTO oder teilweise sogar aus der, ich sag mal, Informatikecke vorgestellt. Da wird gesagt, da können wir das gut machen. Und das ist, finde ich, ist der falsche Weg. Eigentlich muss es aus beiden Richtungen gehen.

A3

Okay, interessant. Okay, was könnte denn eine Lösung dafür sein, für jetzt auch dieses spezifische Problem, was du zuletzt genannt hast? Kannst du dir da vorstellen, was auch vielleicht aus der KI-Forschung, aus der Informatik-Forschung kommen könnte?

I2

Also ich denke, dass man Technologieprojekte oder dass wir grundsätzlich Technologieprojekte nicht mehr als reine Informatikprojekte sehen müssen, sondern das sind ja aber auch die Klassiker, die wir uns seit ein paar Jahren schon bewusst sind, aber trotzdem nicht anwenden. Das bedeutet, es muss interdisziplinär sein, es muss ein begleitendes Change Management geben, es muss eine Fachexpertise vorhanden sein, es muss eine technische Expertise vorhanden sein und es muss eigentlich auch ein Management Commitment geben. Bedeutet, eigentlich bräuchte man sowas wie Technologiestrategien. Und die, glaube ich, gibt es in der breiten Masse zumindest, wie ich das erlebt habe, nur teilweise, weil die sehr klar an den eigenen Geschäftsstrategien verknüpft sind. Und ich sage mal, in der Geschäftsmodellentwicklung oder in der Geschäftsmodellinnovation, also in der BMI-Literatur geht man ja nicht schnell von seinem Kerngeschäftsfeld weg, sondern schaut eher, wie man sich weiterentwickeln kann. Ich denke aber trotzdem, dass, ich sage mal, wenn man jetzt in einem produzierenden Bereich unterwegs ist und man hat ein Produkt, eine Hardware oder ein Gerät, was man verkauft, dann fängt man ja auch nicht an, eine Technologiestrategie sich zu überlegen, ob man jetzt ein Online-Marketing-Unternehmen wird, sondern man schaut ja eher, wo ist man extrem stark und wie kann ich die Effizienz steigern mit den, mit den technologischen Mitteln, die ich jetzt in Zukunft bekomme. Und das wirklich strategisch zu verankern, das Know-how dort zu binden, denke ich, ist schwierig, weil es dort auch durchaus einen Fachkräftemangel gibt. Ich meine jetzt nicht jüngere Informatikerinnen und Informatiker, da gibt es ja nicht nur ein Überangebot, das sieht man ja auch bei Firmen wie Telekom, Salesforce und SAP, die ja wahnsinnig viel wieder freisetzen, sondern eher, ich sage mal, so fünf, sechs, sieben Jahre, die auch vielleicht einen KI-Hintergrund haben, Physik, Mathematik studiert haben, die in diesen Bereichen irgendwie so eine Art Innovationslabor aufbauen und das dann unter einem CX-Bereich anhängen können als Stabstelle. Ich glaube, das fehlt so ein bisschen und ich denke, das wäre eines der Gelingstkriterien für das, was du gerade angesprochen hast als Problem.

A3

Und diese Stabstelle, wenn die quasi mit KI-Innovationen um die Ecke kommt, sag ich mal, und KI-Produkte entwickelt, wie sollten die quasi an alle Stakeholder kommuniziert werden? Welche Arten von Informationen würdest du dir über KI-Produkte wünschen? Ja.

I2

Also das kommt ganz darauf an, ob es jetzt etwas ist, was intern eingesetzt werden soll, ob es personenbezogene Daten verarbeitet, ob wir uns in einem Produktionsbereich befinden, ob es sowas wie Betriebsratszustimmungspflichten gibt, sind wir in einer Spielewiese, sind wir in einer Sandbox. Ich glaube, die Technological Readiness Steps oder diese Scale, diese sieben Stufen sind ja relativ interessant. Vielleicht könnte man sowas analog dazu machen. Ich meine, es gibt ja den Bonner Katalog von Fraunhofer-IAIS zu den Bereichen, was vertrauenswürdige KI ist. Man könnte ja sowas einfach dann entsprechend erfüllen. Wir haben ja aber auch das EU-Regelwerk, das man abfrühstücken(?) muss mit diesem Ampelsystem oder den Kredit-kreditätsstufen(?). Das ist aber alles etwas, wo ich sagen würde, das ist für große Technologieanbieter gut, die sich damit beschäftigen, ob man möglichst eine breitflächige Zertifizierung hat, die dann 300 Seiten EU-Veröffentlichungen durcharbeiten. Wenn ich aber jetzt in einem Unternehmen bin und eigentlich ein bisschen in einer Sandbox oder in einer kleinen Spielewiese Dinge ausprobieren möchte, ist das ja wieder genau das, was wir im Anfang gesprochen haben. Du hast sehr viele Hürden, du hast sehr viele Barrieren, du hast sehr viel Bürokratie. Davon will ich eigentlich ja genau wegkommen. Das bedeutet, eigentlich braucht man so ein Mittelding zwischen beiden und vielleicht nochmal so ein Zwischenglied davon, wo man sagt, wie es eigentlich auch in der Medizin ist. Dann wird immer gesagt, wir machen Use Cases, aber es ist not for clinical use. Und genauso wäre es ja dann auch in anderen Bereichen. Und wenn es da Vereinfachung gibt, wie ein Ampelsystem oder ganz pragmatisch ein Fragebogen-Tool, das interaktiv ist. Also ich habe ein Use Case, ich will den ausrollen intern und ich habe 50 Fragen, die klicke ich durch, bekomme vielleicht einen Extract, eine Auswertung und dann kann ich sagen, ist problematisch, nicht problematisch. Welches Gremium muss mit einbinden? Ich glaube, solche Tools, ich weiß nicht, ob es die schon auf dem Markt gibt, aber ich kann mir vorstellen, dass das vielen Unternehmen Arbeit abnimmt in der Technologievereinbarung.

A3

Okay, super, da hast du schon einige wichtige Punkte reingebracht und du hast natürlich mit dem Ampelsystem auch vielleicht schon darauf angesprochen, was du auf der Webseite gesehen hast. Ich würde mal gerne meinen Bildschirm teilen. Und dann seht ihr hoffentlich das Logo, was ich einmal groß machen kann. Das ist genau ja dieses Ampelsystem, beziehungsweise hier wird es eher an den Nutri-Score, haben schon einige gesagt, erinnern, beziehungsweise auch das Design ist vom Nutri-Score inspiriert. Und ohne dich jetzt noch weiter zu primen, würde ich dich jetzt erstmal fragen, was deine spontane Reaktion auf dieses Label ist. Was siehst du hier? Das kann ich größer machen.

I2

Kannst du, ja genau, ich muss mal ein bisschen größer. Also ich bin selber schon ein bisschen gebiased, weil ich mich mit diesen Labels beschäftige, aus vielerlei Gründen, aber also was mir hier auf jeden Fall auffällt, ganz offen, und XXXXXXXXX, ich nehme dich bei Wort, es gibt keine richtigen und falschen Antworten, also zwei QR-Codes, sage ich, sind zu viele, entweder man referenziert auf einen oder man braucht eine Unterbeschreibung, wofür ich referenziere. Die Codes sagen mir gar nichts. Also hier steht halt, issued 24. Juli, da gehe ich davon aus, dass das Label dann freigegeben wurde und hier kann ich weitere Informationen finden, okay, finde ich nicht schlecht, könnte man vielleicht nochmal anders positionieren. Infer MobileNet3V small, okay, alles fein. TensorFlow, das kann ich noch nachvollziehen, dass es hier steht. Robustness ist eher schlecht, also ein Gelb, warum ist das gefüllt als Icon und warum sind die anderen nicht gefüllt, frage ich mich, also daher würde ich mir eine Vereinheitlichung wünschen des Designs. Dann Top One Accuracy, und dann 60% ist kontraintuitiv, finde ich, also es müsste irgendwie anders sein. Running Time Per Inference, ich glaube, das ist erstmal irrelevant, weil das eher so ein Usability-Thema ist. Draw Power Per Inference, wenn das auf Nachhaltigkeit ansieht, kann ich mit der Angabe nichts anfangen. Also ich sehe eigentlich nur A, also es ist sehr gut und bei einem Nutri-Score finde ich die große Kritik, vergleiche ich jetzt Systeme bei Essen. Ich vergleiche eine Tiefkühlpizzen miteinander und eigentlich sind Tiefkühlpizzen in der Regel schlecht für dich, aber die eine macht ein bisschen was besser als die andere und dann ist es okay. Wenn ich jetzt an einen Nutri-Score von einem AI-Label denke, dann denke ich mir, okay, sind wir jetzt irgendwie in etwas, wo wir kein Social Scoring machen, aber ich sage mal so Kreditfinanzierungsgeschichten, aber die nehmen trotzdem Proxy-Parameter, dann bin ich da sehr gut, weil die anderen noch schlimmere Sachen machen, kann ich nicht beurteilen. Bedeutet, ich hätte mir hier noch ein bisschen mehr Text gewünscht, was der Anwendungsbereich, welche Metriken, warum haben wir diese vier, und die Anordnung der Logos vielleicht nochmal anders. Aber ansonsten finde ich grundsätzlich ein Farbschema intuitiv. Letzter Punkt, den ich noch anführen möchte, den wir auch schon mal diskutiert haben. Es gibt ja auch den Haltungsform-Score, wo das Farbschema disruptiert wurde durch Lobbyarbeit. Bedeutet, ich habe das Beste, das ist grün, dann kommt drei, das ist orange, dann kommen zwei, das ist eigentlich scheiße, das ist blau. Bedeutet, eigentlich, wenn wir solche Labels machen, brauchen wir eine Vergleichbarkeit, eine Transparenz. Und von den Farbschemen müsste es halt genauso sein und nicht irgendwie zäh wird auf einmal blau oder sowas. Das ist was, was ich für schwierig finde, weil ich mir denke, wenn wir so eine freiwillige Selbstverpflichtung haben von AI-Label, wird es schon schwierig. Und letzter Punkt, der mir jetzt noch aufgefallen ist, wir haben ja keinerlei externe Validität durch ein Icon oder durch ein Label oder durch eine Firma, die das nimmt. Vielleicht, wenn ich draufklicke, aber das ist jetzt erstmal ein anonymes Label. Wenn ihr mir jetzt gleich vielleicht eins zeigt von TÜV oder XXXXXX oder sonst wie, kann ich da gerne noch zwei, drei Sätze zu sagen.

A3

Na klar, also auf den Punkt kommen wir nochmal ganz am Ende. Soll das jetzt wirklich eine Selbstverpflichtung sein oder soll das vielleicht eine Stelle machen? Ich würde aber gerne nochmal visuell auf das Label eingehen. Also du hattest ja, ich weiß nicht, ob du Unklarheiten hattest, also Infer, sagt dir wahrscheinlich, es bedeutet einfach, dass es Inferenz ist. Wir haben hier den Namen des KI-Modells, letztendlich ist es hier Bildklassifikator, den wir verwenden, der auf einem ImageNet-Datensatz, ich weiß nicht, ob du das kennst, trainiert wurde.

I2

Ah, okay. Mhm.

A3

Genau, das soll das eben sein. Und ja, dieser Score richtet sich, das entstammt aus einem Paper von XXXXXXX XXXXXXX, der richtet sich an einem Vergleich oder der kommt aus einem Vergleich zwischen 30 verschiedenen neuronalen Netzen, die eben alle eine Bildklassifikation sind. Und hier haben wir eben den Score für MobileNetV3. Und der speist sich auch aus verschiedenen Messungen für diese Untermetriken. Also es gibt 10 an der Zahl insgesamt, hier sind nur 4 dargestellt. Und als Compound Score, also zusammen, die werden gewichtet und ergeben dann eben diesen großen Score hier. So läuft das. Genau, das nochmal so als Einordnung. QR-Codes sind einmal Paper und noch weitere Informationen über das Projekt. Und dann haben wir natürlich unsere Grafikkarte und unser TensorFlow Framework. Mit diesen Informationen hast du da nochmal gerade einen Impuls irgendwie nochmal reingekommen, den du dazu loswerden möchtest zu diesem Label?

I2

Ja, auf jeden Fall. Also alles nachvollziehbar und ich glaube, wenn ich nochmal 5 Minuten mehr Zeit hätte, kommt man sich, glaube ich, ruft man sich damit ein. Ich habe gedacht, wenn das für ein Fachpublikum ist, also wirklich, ich sage mal, reinen Entwicklerinnen und Entwickler, würde ich sagen, könnte man mehr damit arbeiten, wenn wir eine Allgemeinzielggruppe oder ich sage mal, ich sage, also wie nennt man das nochmal, niedrigschwellig das Angebot machen möchten. Finde ich, und das soll es ja sein, weil sonst brauchen wir kein Scoring, wenn es für Entwicklerinnen und Entwickler direkt ist, dann gibt es andere Metriken, die man nutzen kann. Bedeutet, es ist zu limitiert, ich brauche mehr Informationen und da sehe ich die Möglichkeit, mit relativ wenig das zu machen, also dann einfach davor zu schreiben, was ist, wie du gesagt hast, das ist das Modell, also Bilderkennungsmodell Doppelpunkt GPU Used Model, was weiß ich, und dann die Icons rausnehmen und vielleicht da einfach fünf Skalen auch in diesem ABCDE und dann nimmt man das einfach Subscaling und dann füllt man die entsprechend aus und dann hat man auch die Gewichte und dann hat man zweimal grün, einmal gelb, einmal rot und dann gibt man vielleicht das Scoring an. Auf der rechten Seite hat man ja immer noch so einen Text, um da Corrupted Robustness, Running Time reinzuschreiben mit einem kleinen Label, aber dann könnte man eben sagen, das ist das gewichtete Scoring und ich habe eine viel bessere Nachvollziehbarkeit, weil hier sehe ich ja auch in der Skala nicht, sind wir dann bei der Running Time bei 1.5 schlechter oder wann ist die Accuracy, im Vergleich zu den 40 anderen besser? Du hast mir das ja gerade erklärt, aber das ist was, das finde ich, 60%, das ist für viele Anwendungen eigentlich minderwertig. Für den ist die Bilderkennung fein, aber das sind die Sachen, die mir da direkt aufgefallen sind.

A3

Ja gut, aber 60% ist ja hier auf jeden Fall eindeutig durch die rote Färbung schon relativ weit unten im Vergleich dieses MobileNetV3, im Vergleich zu allen anderen Netzen. Klar, also dieses ImageNet hat insgesamt 1000 Klassen, das heißt, wenn man wirklich random klassifizieren würde, dann hätte man ein Tausendstel Chance für eine richtige Klassifikation, richtige Accuracy. Von daher sind 63% dann doch irgendwie eine Verbesserung, aber im Vergleich, und da komme ich jetzt auch mal direkt zu, hier haben wir nämlich noch ein zweites Label im Vergleich. Das sieht man eben, gerade wenn man eben zwei Netze miteinander vergleicht und vielleicht auch unterschiedliche Ziele hat, die man verfolgt, soll man hier eben erkennen, welche Vor- und Nachteile zu sehen sind. Jetzt habe ich das zweite Label mal aufgemacht, gleicher Aufbau. Was fällt dir hier im Vergleich zu ein? Kannst du zusammenfassen, was du hier siehst?

I2

Also ich finde das kontraintuitiv, wenn ich die beiden betrachte. Unten rechts die Farbe kann ich nicht erkennen. Wie ist das? Ist das eine Mischfarbe? Also entweder ist es so hellgrün oder so eine Mischung aus grün und gelb. Also das kann ich nicht richtig sehen. Ich glaube, das soll das B sein. Ja, okay, schwierig.

A3

Ja, genau, es ist so dann zwischen gelb und grün, ja.

I2

Wenn wir dann eine ruckgrüne Schwäche haben, dann ist es spätestens vorbei, weil dann hast du grau auf grau. Das ist also dann nicht so inklusiv. Ich finde es kontraintuitiv aus dem Grunde, dass wir vorhin bei A sind und da haben wir zweimal grün, also zweimal A, einmal C, einmal E. Und hier haben wir, wenn ich das sehe, zweimal C, einmal A, einmal B. Wäre für mich in der Gewichtung eigentlich dann nicht so, weil wären wir ja eigentlich bei C, C, A. Ja, nee. Also für mich wäre das EfficientNetB4 müsste im Scoring höher sein als das MobileNetV3.

A3

Ja, genau, das kommt daher, dass wir ja eben noch weitere Metriken haben, die hier jetzt nicht gezeigt sind. Zum Beispiel die Modellgröße ist ein Faktor, wie groß, wie viele Parameter hat das Modell letztendlich, aber auch zum Beispiel sowas wie, es gibt ja die Top 5 Accuracy noch, die Top 10 Accuracy. Also ich weiß nicht, ob du das kennst, aber wie oft eben das Wahrscheinlichste, das von dem Modell am wahrscheinlichsten gehaltene Ergebnis auch die wahre Klasse ist oder ob die wahre Klasse unter den 5 wahrscheinlichsten Ergebnissen ist. Solche Metriken werden noch dazugezogen und die fließen eben auch in diesen Compound Score rein, in diesen großen. Genau, daher kommt wahrscheinlich die Verwirrung.

I2

Ja, das finde ich auch total legitim und ich meine, deswegen gibt es ja Scan for Further Information. Du hast mich ja gefragt, was mein erster Impuls ist bei Sichtung dieser vier und dann würde ich mich auf jeden Fall grundsätzlich als Erstbetrachterin, Erstbetrachter eben fragen, dass es für mich unlogisch ist, wenn du quasi versteckte Informationen hast oder Informationen hast, die erst auf gemeinsamen Informationsstand basieren, den ich abrufen kann über den QR-Code.

A3

Okay, verstehe. Siehst du denn hier dennoch, trotz der Fehler und Verbesserungsmöglichkeiten, eine Brücke zu deiner Arbeit, wie dir das vielleicht helfen könnte oder vielleicht auf jeden Fall die Leute, mit denen du zusammenarbeitest, helfen könnte?

I2

Ja, auf jeden Fall. In einer Use-Case-Evaluierungsphase, wenn wir uns dazu entscheiden, wir brauchen eine Fachsprachesprache-Erkennung oder wir brauchen eine Bilderkennung, dass wirklich alle dann gegenübergelegt werden und dass man darüber dann Auswahl trifft. Also ich sag mal, in einer Use-Case-Phase, in der Vorbereitungsphase, in der Evaluationsphase, wo wir darüber diskutieren, wie jetzt, ich sag mal, ein MVP oder ein Proof-of-Concept stattfinden kann und dann könnte man verschiedene Modelle miteinander vergleichen und schauen, wie gut die abgeschnitten sind. Natürlich, aber die, ich sag mal, so nutzen, dass wir die nehmen, wo die Metrik, die für uns am wichtigsten ist, an erster Stelle ist. Also ich sag mal. Nur als Beispiel, Accuracy und Robustness, so nehmen wir einfach mal und dann würde ich gerne die filtern können und da die Top-Drei nehmen und dann die mal nebeneinander werfen lassen. Das würde ich sagen, wäre ein wahnsinniger Vorteil.

A3

Also eine Interaktion auf jeden Fall, würdest du dir da wünschen, ein interaktives Label. Okay, war das jetzt nur beispielhaft oder wären das wirklich die beiden Metriken, die euch wahrscheinlich am wichtigsten sind, gerade wenn man jetzt mal sich hier den Power Draw und aber auch die Running Time, die auch korreliert ist mit dem Power Draw, wenn man die sich nochmal anschaut.

I2

Das ist ganz unterschiedlich von den Use Cases, die wir brauchen. Also ich sage mal, in der UserInnen-Interaktion mit einer breiten Zielgruppe würde ich sagen, ist Robustness zu vernachlässigen. Da brauche ich eher ein sehr responsives System, das eine okayische Accuracy hat. Wenn ich natürlich in einer Bilderkennung bin für radiologische Anwendungsfelder, möchte ich eine sehr hohe Accuracy haben. Ich würde sagen, ist 81,27 für nicht-explorative Arbeitsbereiche unbefriedigend. Ich denke, dass wir einen viel größeren Fokus auf PowerDraw legen müssen. Also ob jetzt ein Rechensystem mit einer 30 60 läuft oder mit 3 40 80, ist eine Verzwanzigfachung, je nachdem welches System du nimmst oder was noch da drin ist. Das ist das 20-fache an Energie oder an Wattzahl, die dadurch schon daraus geblasen wird. Und das ist jetzt mal für ein Jahr bei einem Modell oder einem System jetzt nicht so wahnsinnig viel. Das sind vielleicht 200 Euro. Wenn wir aber ein großes Upscaling haben, ist das ein hoher Energieverbrauch. Und ich denke, dass wir unter Berücksichtigung von vielen Forschungsbereichen, die jetzt mit SDG-Feldern arbeiten, extrem wichtig, dass wir anfangen, vielleicht Label auch mit einer Energieeffizienzkategorie auszustatten. Und dort auch sowas wie Fugalität der Systemkomplexität reinbringen. Also wie aufwendig sind die Modelle geschrieben? Wie groß sind die Modelle? Und das ist ja etwas, was am Anfang von der Technologie erstmal irrelevant ist. Erstmal muss ein Auto fahren. Wie viel Liter ist erstmal egal. Zu einem späteren Zeitpunkt oder wo wir uns jetzt befinden, denke ich, brauchen wir da eine größere Aufmerksamkeit. Und ich fände das sehr wünschenswert, wenn das gehen könnte.

A3

Absolut, okay. Ja, da gehe ich total mit. Also KI verbraucht so viel Strom heute schon, KI-Modelle. Deswegen, ich denke auch, dass da auf jeden Fall viel mehr darauf geachtet werden sollte. Das hat XXXXXXX ja auch wirklich in mehreren Studien schon erforscht, gerade dieses Thema Ressourcenverbrauch von KI-Modellen. Um nochmal auf die Label hier zurückzukommen. Ist hier irgendwas, ich meine, du hast es gerade schon ein bisschen angeschnitten, irgendwas, was du gar nicht verstehst, was du gar nicht hilfreich findest?

I2

also nichts von dem was ich nicht vorhin schon gesagt habe tatsächlich

A3

Ja. Okay, gut. Was findest du besonders gut? Was gefällt dir besonders gut an dem, also was ist irgendwie, was das rausgestochen ist vielleicht an dem KI-Label?

I2

Also ich finde es erstmal gut, dass es wie so ein Kartensystem ist, dass es einen Wiedererkennungswert hat. Ich hatte darüber nachgedacht, wenn man, es gibt ja so Design-Sprinting-Ansätze, dass man mit Kartenmodellen, Technologie, Einsatzbereiche diskutiert, evaluiert und dann mit physischen Karten arbeitet. Und ich kann mir vorstellen, dass sowas sehr hilfreich sein könnte, wenn man solche Sachen daneben legt oder zumindest auf der Rückseite dargelegt sind, wenn man dann sagt, wir wollen ein Use-Case X Bilderkennung, ich sag einfach mal Fahrräder, wir wollen Fahrräder erkennen, warum auch immer, weil wir sind die Stadtwerke und wir wollen irgendwie eine neue Fahrradstraße bauen. So und dann haben wir Use-Cases, dann nehmen wir mal ein paar Entscheiderinnen und Entscheider mit, die keine Technologen sind und dann legen wir die Sachen einfach mal daneben. Und dann könnte man vielleicht auf einem Kartenbereich die möglichen Modelle, die zum Einsatz kommen können, daneben legen, dass man einfach auch erklärt, dass es nicht ein Modell gibt, es gibt nicht eine KI oder ein System, das jetzt alles machen kann, sondern dass das im Zweifel auch ein Trade-Off ist und wir da was ausprobieren müssen. Da wäre eben aber die Frage, ist es dann etwas, was eher so ein Data-Literacy-Ansatz ist, dass man Personen aufschlauen möchte und informieren möchte über die Eigenschaften von Systemen? Oder soll es eher als Entscheidungsgrundlage dienen? Ich denke, das ist durchaus fair, dass es beides sein kann.

A3

Okay. Also auf jeden Fall würdest du sagen, die Modelle sind vergleichbar durch Label und auch vor allem für Laien vergleichbar?

I2

Ja, auf jeden Fall. Also das sollte ja auch der Anspruch an das Label sein, aus meiner persönlichen Perspektive. Nur, was du ja eben auch schon mal nochmal erklärt hast, wir haben noch irgendwie Subdimensionen oder Nebendimensionen und mir persönlich fehlt da eine klare Gewichtung. Also hier, finde ich, kann nur sein, diese vier, die da stehen, Power, Draw, Accuracy, Time und Robustness, sind jeweils mit 25% gewichtet und dann kommt eigentlich das Scoring zusammen. Wenn du irgendwas anderes an Metriken dahinter liegen hast, dann muss das hier stehen oder es braucht eine andere Darstellungsform, wie die da unten, wie ich ja vorhin schon angekündigt hatte, vielleicht eine Tabelle oder eine andere Darstellart, weil ich persönlich finde, dass nicht transparent

A3

Okay. Und dass man irgendwie einen Aufklapp-Button hat oder so, kannst du dir sowas vorstellen? Das würde auch wieder in Richtung Interaktion gehen, Interaktivität zumindest.

I2

Ja, kann man machen, wenn man aber das, ich sag mal, als physisches Produkt machen möchte und was ich denke, was gut ist, weil Workshops funktionieren gut physisch auch, da bricht es einfach die Karte ein bisschen höher designen oder ein bisschen weiter setzen und dann schauen, wie viele Parameter man da reinsetzen kann, die dann zu der Bewertung führen oder Rückseite, alles fein.

A3

Okay, sehr gut. Ja, dann hattest du eben schon genannt, kannst dir vorstellen, dass für jede dieser Untermetriken, nenne ich sie mal, auch ein Nutri-Score so dastehen könnte. Habe ich das richtig verstanden? Ja.

I2

Ja, oder zumindest, ich sag mal, das Icon und dann eine kleine Skala von, ich sag mal, 0 bis 5 oder 1 bis 5 oder 1 Prozent, muss ja nicht nochmal ein eigener Nutri-Score sein, aber wie viel erfüllt worden ist auch im Vergleich zu den anderen, also der In-Group-Vergleich deutlicher werden sollte.

A3

also da wäre quasi die Kritik an der Farbskala, die ja genau das eigentlich aussagen sollte.

I2

Ja, ne, die Farbskala ist ja für mich, wir haben A bis E. E ist sehr schlecht, A ist sehr gut. Und wenn wir jetzt da unten ja nochmal die Farbskala haben in dem Fall, ist das ja eigentlich die Gewichtung, die es ergibt. Und ich sehe jetzt zweimal gelb, einmal grün, einmal B. Und dann liegen wir ja eigentlich irgendwo bei B und C bei der linken Grafik. Bei der rechten Grafik haben wir einmal gelb, zweimal grün, einmal rot. Und so, wenn ich das jetzt erstmal sehe, dann haben wir eigentlich zweimal gelb, einmal grün, dann wären wir eigentlich auch irgendwo zwischen B und C. Aber hier ist bei dem MobileNet haben wir das Ergebnis A. Wenn ich aber jetzt unten nicht diese Farben hätte, sondern ein Schema sind auf Platz 7 von 40, Gewichtung N, 0 komma N, wäre das für mich deutlich besser zu verstehen oder es wäre in sich logischer. Weil die Erklärung, die du mir gegeben hast, hat mich das Ganze verstehen lassen, aber dann finde ich, also es muss für sich stehen ohne jegliche Erklärung dann erstmal. Und wenn ich dann interessiert bin, okay, jetzt schaue ich mir mehr Informationen an, aber so ist es für mich erstmal unlogisch. Aber vielleicht habe ich auch einen Denkknacks, aber das ist das, was mir auf jeden Fall als erstes aufgefallen ist. Er kann ja nicht A geben, wenn ich zweimal grün, einmal rot, einmal gelb habe, sozusagen.

A3

Du hast absolut keinen Denknack. Das ist alles legitim. Wie gesagt, keine falschen Antworten gibt es. Das ist eben dein Eindruck. Und vielen Dank auf jeden Fall für die Inspiration und für die Verbesserungsvorschläge. Ich würde gerne dich noch fragen, ob du noch weitere Kommunikationsformen kennst, die quasi den gleichen Zweck wie diese Label erfüllen. Und ob dir sowas mal im Alltag begegnet ist. So dieses Konzept, ein Modell muss irgendwie dargestellt werden.

I2

Neben jetzt dem Nutri-Score?

A3

Ne, also schon spezifisch auf KI-Modelle, ob dir da irgendwelche Kommunikationsformen schon mal aufgefallen sind.

I2

Also was ich sehr viel jetzt lese, ist eben das TÜV und DIN, an denen zumindest Zertifizierungsthemen arbeiten. Aber was ich da sehe, sind sehr viele, also wenig Modelle und sehr viele Konzepte erstmal. Und was ich natürlich viel mehr sehe, sind Dimensionen für Trustworthy oder Responsible AI, die aber nicht in Labeling gehen, sondern eigentlich eher wieder Bewertungsmetriken sind. In der Juristerei auf jeden Fall ja auch. Die sind ja super hinten dran. In der Medizin ist es mir tatsächlich noch fast gar nicht begegnet, weil das alles über dieser Medizinproduktzertifizierung abgeht oder abgegolten wird. Extrem wenig, nein. Also das meiste ist ja wirklich, ich sag mal technisch oder sogar Open Source-mäßig abrufbar. Und so bei AWS oder so ist mir das auch ehrlicherweise nicht untergekommen, nein.

A3

Ja gut, wir haben mal ein bisschen Recherche betrieben und uns angeschaut, ja wie wird es denn im Internet für EntwicklerInnen noch so aufbereitet und da haben wir sechs verschiedene Arten von Darstellungsformen gefunden. Das wären zum einen natürlich die wissenschaftlichen Publikationen, die zum Beispiel in dem Fall wie das MobileNetV3, das ist die ursprüngliche Publikation, eine Google-Sache, in der eben das MobileNetV3 entwickelt wurde, kann man sich natürlich sehr viel dazu durchlesen und da wurden auch zum Beispiel Benchmark-Tests durchgeführt für diese MobileNets. Dann haben wir natürlich sowas wie Modellkarten, das ist jetzt in dem Fall von Google, wenn man da bei TensorFlow sich mal irgendwie durchklickt und Informationen haben möchte zu dem MobileNetV3, dann findet man hier sicherlich zum Beispiel Benchmarks, Informationen dazu, wo wurde es drauf getestet, auf welchen Datensätzen, wie sehen die Daten aus, wie viele Parameter hat das Modell und wie groß ist es letztendlich auch auf der Festplatte. Dann gibt es die schöne Seite Papers with Code, das verknüpft quasi ein Konzept, in dem Fall ein KI-Modell, mit den Papern, die dieses Modell eben zitieren. Also wenn man jetzt hier rumschaut. Wenn man hier runter scrollen würde, dann würde man eben alle wissenschaftlichen Publikationen sehen, die das MobileNetV3 irgendwie zitiert haben. Dann natürlich die typischen Blog-Beiträge, Medium oder auch Towards Data Science sind hier wahrscheinlich schon mal untergekommen. Dann die Documentations, also in dem Fall ist es eine Python-Library, da steht dann auch allerhand Informationen zum Beispiel jetzt wirklich zu den Methoden, die jetzt eben verwendet werden, beim Aufruf dieses MobileNets im Code. Aber auch eben Metriken, also Benchmark-Tests. Und zu guter Letzt haben wir noch die Factsheets, das ist von IBM für IBM-spezifische KI-Modelle, die anhand von verschiedenen Kategorien eben Informationen über deren KI-Modelle liefern, zum Beispiel Bias, wie verzerrt ist eventuell ein Modell, wie fair ist das Modell letztendlich, gerade wenn man über Diskriminierung und sowas redet. Wie robust ist das Modell? Das hatten wir auch schon im Label. Wie sehen die Outputs aus? Und so weiter. So, das haben wir jetzt mal zusammengestellt. Und wenn man jetzt so das Label, was ich dir ja eben gezeigt habe, oder immer noch zeige, daneben packt, welche Vor- und Nachteile siehst du hier im Vergleich? Dieser Kommunikationsform, du musst nicht durch alle gehen, nur das, was dir ins Auge springt.

I2

Also für mich liest sich das nach einem klassischen Disseminationskonzept oder Prozess auch, wie Wissenschaft betrieben wird und Wissenschaft in die Anwendung kommt. Und das ist ja eigentlich der klassische Weg. Ich forsche, ich habe ein Fachpaper, ich habe einen Anwendungscase, der kommt in ein niedrigschwelliges Fachjournal rein. Parallel sind die EntwicklerInnen ja auch ForscherInnen. Das bedeutet ja, nach dem Fachpaper kommt das auf die eigene Webseite der Firma und dann schlussendlich habe ich ein Label. Also das ist, wie auch wir Forschung betreiben. Also erstmal Spiegelwiese, dann Publikationsschmerz. Das dauert ja lange, deswegen schnell mal irgendwo schon mal ein paar Infos veröffentlichen, vielleicht auf eine Website, dann zu einem Newspaper oder so gehen, damit, ich sage mal, die Time-to-Market, also die Zeit, die von der Idee zur Umsetzung bis es funktioniert, kürzer wird, damit ich auch schneller einen Austausch bekomme mit anderen WissenschaftlerInnen. Und ich denke, wir brauchen die alle. Es ist einfach nur der Unterschied, welche Zielgruppen habe ich. Bei dem wissenschaftlichen Paper. Da sind es 99% WissenschaftlerInnen. Bei den Beschreibungen von IBM und Google sind es, ich würde mal sagen, die allermeisten EntwicklerInnen und auch WissenschaftlerInnen. Hier unten rechts, everything you need to know, das ist auch für EntwicklerInnen, AnwenderInnen, vielleicht noch für einige, die so ein bisschen interdisziplinärer sind, aber ja, schon auch eher normal. Technisch orientiert und das Rechte ist ja dann was, was nochmal das Ganze abstrahiert und auch von der Niedrigschwelligkeit auch an Nicht-Techies ausgelegt ist, um dann auch Nachvollziehbarkeit, Vertrauen und Transparenz zu schaffen über das, was ich vorher gesehen habe. Wenn ich aber den Weg zurückführen kann, deswegen hast du ja gesagt, es geht direkt zum Paper, ist das sehr cool. Gedanke wäre ja auch, kann man da Studien nehmen, also Scan for further Information hat man Landing-Dashboard, wie tief möchtest du reingehen? Ich will die Originaldaten haben, ich will aufs GIT-Repository(?) oder ich will die Grafik mir anschauen, vielleicht diese [Ungenau] anschauen. Nee, ich will nur auf die Google-Seite referenziert werden oder ich will auf den Newspaper-Artikel kommen. Das wäre so die Möglichkeit, weil das Label ist ja, ich würde mal sagen, so umfassend für erstmal alle Zielgruppen, wenn ich das erste Mal mit diesem System oder in dem Fall mit dem Bilderkennungssystem, in Berührung komme. Von daher wäre das so in der Gegenüberstellung eher keine Gegenüberstellung, sondern eher etwas, was sich gegenseitig ergänzt.

A3

Okay, interessant. Und da auch wieder über eine Interaktion. Okay. Ja, vielen Dank. Dann würde ich jetzt auf den Punkt, den du eben schon aufgemacht hast, noch zu guter Letzt zurückkommen. Ja, wer sollte denn ein solches Label ausstellen? Punkt. Was sind deine Gedanken dazu?

I2

ja das muss man testen das kann ich jetzt nicht sagen die Institution die bereits in Vertrauensvorschuss haben in technischen Bereichen tun gut daran da rein weiter zu investieren weil das marktvorteil bringt sich da auch festzusetzen ich denke es wird viele geben so was wie Google, Microsoft, IBM, vielleicht auch sogar Salesforce und solche die selber vielleicht Zertifizierung anbieten auf der Seite werden TÜV, DIN und etablierte Unternehmen auch am deutschsprachigen Raum versuchen selber noch mal Zertifizierung anzubieten die vielleicht auch noch andere unternehmen targeten mit weniger, ich sage mal, Entwicklungspower das heißt ich glaube es wird nicht das eine Label geben es wird mehrere Label geben weil der Markt zu heterogen ist. Am Ende des Tages wird auch die Frage sein sind die Label Geber noch diejenigen die die Sache machen die state of the art sind die technologisch führend sind. Und da würde ich persönlich da weiß ich aber nicht wie das jetzt Entscheider in Entscheider in Großunternehmen sehen oder auch in kleineren unternehmen ich meine TÜV ist ein Name DIN ist auch ein Name XXXXXX ist auch Name aber wenn ich jetzt international gehe vielleicht ist das kein Mensch mehr, dass da XXXXXX vor drauf steht und wenn er Google drauf steht auch schön wenn es ein Konsortium ist vielleicht noch eher aber da… Ich kann es wirklich nicht sagen welchen weg das gehen muss ich denke aber dass wir eine Vertrauens Erhöhung massiv bekommen wenn wir starke Partner in der Belabelung haben in der… Von Zertifizierung von Modellen oder von Systemen dass das nicht einfach für sich stehen kann dass man da neue ich sag mal Firmennamen sich ausdenkt oder neue Proxy Namen, sondern da müssen die Logos eigentlich drauf.

A3

Okay, alles klar. Sehr spannend, was du dazu denkst. Entschuldigung. Okay. Findest du denn allgemein die Zertifizierung sinnvoll? Und vor allem auch nochmal übertragen auf deinen Arbeitsalltag. Würde es dir helfen?

I2

Also das ist wieder eine sehr breite Frage, weil wir ja schauen müssen, über welche Anwendungen wir sprechen. Und wir schauen ja bei einem AI-Label auf der einen Seite auf Modelle. Da haben wir aber noch nicht die Anwendung mit drin. Also wenn ich ein Bilderkennungssystem habe, kann es ja in einer unkritischen oder kritischen Anwendung zum Einsatz kommen. Das bedeutet, ich habe ja jetzt technisch erstmal einen Hinweis, das ist einfach ein sehr gutes Modell. Wenn ich aber entwickle, brauche ich ja viel mehr als nur das Wissen, ob ein Modell sehr gut funktioniert oder nicht. Ich muss ja wissen, also vielleicht ein blöder Vergleich, aber wenn wir ein Auto haben und ich weiß, es fährt schnell, dann ist das gut. Aber ich muss ja auch als Autobauer wissen, funktionieren die Airbags gut? Funktionieren die anderen Systeme gut? Habe ich einen Anschnaller? Ja, und genau diese Fragen müssen wir uns ja auch beim Einsatz von, ich sage mal, Bilderkennungssystemen stellen. Also will ich die im Produktionsbereich einsetzen? Was kann ich tracken? Welche Daten bekomme ich? Und da ist es gut, ein AI-Label zu haben zu dem Bereich. Aber eigentlich bräuchten wir ja auch, so wie ja die EU das im EU-AI-Act zumindest jetzt schon mal anteasert oder eigentlich ziemlich klar auch gestellt hat, aber wieder in einem sehr komplexen Bereich oder in einem komplexen Framework, wann es kritisch ist. Und wann es nicht kritisch ist. Von daher denke ich mir, das ist der erste Schritt, diese Labels zu haben. Aber die Kombination aus Modellqualität und Anwendungsbereich wäre für mich der nächste Schritt. Wie das dann aussieht, das ist, glaube ich, nochmal ein komplett neuer Workshop mit ganz vielen Leuten. Aber das ist das, was ich persönlich als das Praktikabelste sehe. Deswegen habe ich am Anfang gesagt, sowas wie ein Fragebogen. Ich habe Fragen, die abgefragt werden. Dann kann ich meine Labels aussuchen. Die werden schon, also kann ich mir meine Modelle aussuchen, die ein Label sind. Und dann gehe ich in die Anwendung rein. Und das ist ja das, was vielleicht auch dann dazu führt, dass die Systeme mehr genutzt werden und vielleicht auch verbessert werden. Und dadurch vielleicht auch nochmal einen Aufschwung gibt, der ja gerade, oder wie wir am Anfang ja, oder ich zumindest behauptet habe, runtergeht, weil viele so ein bisschen frustriert sind darüber, dass es jetzt nicht alles schon schafft, was vielleicht versprochen wurde.

A3

Und dann die letzte Frage wäre, wie müsste das Label angepasst werden an den Wissenshintergrund? Das hast du auch eben ganz kurz mal angeschnitten. Aber gerade wenn wir so über verschiedene Stakeholder, verschiedene Gruppen reden, auch in deinem Arbeitsalltag, wie kannst du dir da so eine Customisability… Hast du da irgendwie eine Vorstellung, wie das...

I2

Eigentlich ziemlich einfach. Also mit einer Rückseite, wo die Metriken, die Bewertungsmetriken oder die Parameter sauber erklärt werden. Mit einer Skala und dann auf den Scan for Further Information auf zwei Leveln, also sehr in einfacher Sprache gehalten und einmal sehr technisch, wenn ich mehr haben möchte, wo vielleicht auch Informationen aus den Papern drinstehen, die erklärt werden, warum das wichtig ist und warum hier so bewertet wurde, wie bewertet wurde. Da kann man, glaube ich, ziemlich viel kopieren an Texten, die Einstiegsantworten für, ich sag mal, AI-Beginner werden ja ziemlich kopierbar und dann müsste man überlegen, ob man die halt wirklich an phosphides Modell über eine neue Grafik reinpackt oder nicht. Vielleicht reicht es auch, wenn man die gleichförmig lässt, aber über Scan for Further Information mehr Informationen bereitstellen, über die Dimensionsausprägung. Ich glaube, das würde schon wahnsinnig helfen.

A3

Okay, alles klar. Vielen, vielen Dank für diesen Input. Ich habe auch tatsächlich meinen Leitfaden hier durchgearbeitet. XXXXXXXXX, habe ich irgendwas vergessen oder hast du noch eine Frage?

A4

Nee, ich glaube, soweit haben wir alles.

A3

Super mit der Zeit. Das ist wirklich immer sehr spannend. Wir können gerne die Aufnahme dann...