A3

Let's go. Okay, ja, dann ist die erste Frage, ja, stell dich doch mal kurz vor, was ist so dein Background?

I12

Ich habe einen Psychologie-Human-Factors-Background. Also ich habe Psychologie im Bachelor studiert und im Master Human Factors und promoviert im Bereich Mensch-Maschine-Interaktion. Da habe ich mich mit dem Thema beschäftigt, wie verstehen Menschen eigentlich Algorithmen? Und jetzt außerhalb der akademischen Welt beschäftige ich mich mit dem Thema UX-Research, also der Frage danach, wie kann man Software so gestalten, dass sie sinnvoll ist und sogenannten Customer Value hat, also den Kundenbedürfnissen entspricht. Und da habe ich zuvor im Bereich Medizintechnik, diese Art von Arbeit gemacht. Da war ich im strategischen Produktmanagement und im UX-Research. Und jetzt arbeite ich im Bereich Kommunikationssoftware, Telefonie Software und versuche gerade zu verstehen wie man, ja, welche Bedürfnisse, welche sogenannten Jobs to be done, die Kundinnen haben, um mit dieser software irgendwie arbeiten zu können.

A3

Möchtest du dein Unternehmen, wo du gerade arbeitest, disclosen quasi?

I12

Achso, wird das dann auch in der, also in der… nein? okay.

A2

Also es geht glaube ich für uns nur darum, dass wir wissen, ist es mittelständisch, wie groß ist das, was macht ihr so?

I12

Die Firma heißt XXXXXXX, ist in XXXXXXXXXX und ist ein inhabergeführtes mittelständisches Unternehmen mit fast 300 MitarbeiterInnen, ist ein Software-Telefonie-Unternehmen.

A3

Okay, alles klar. Dann kannst du noch mal ein bisschen näher auf deine Rolle genau eingehen und auch vor allem auf deine Aufgaben bezüglich der Arbeit mit KI und ML-Systemen.

I12

Also ich arbeite in der Produktentwicklung, in einem Produktteam. Bis letzte Woche, jetzt wechsel ich mein Produktteam und der Aufgabenschwerpunkt verändert sich so ein kleines bisschen, aber es ist gerade noch in der Mache. Aber ich bin in der Produktentwicklung verortet und arbeite da eng zusammen mit Entwicklern und UX-Designern und auch sogenannten Product Ownern an einem bestimmten Teil der Software. Das war jetzt in der Vergangenheit viel so die Experience der sogenannten IT-Admins, die die Software benutzen. Jetzt verändert sich das so ein bisschen und ich überlege stärker, aber auch mit den gleichen Rollen. Also welche AI-Features im Sinne von welche AI-Feature-Statistiken, Insights brauchen Kundinnen von uns, um produktiver und effizienter zu arbeiten? Also ich glaube, dieses ganze KI-Thema bei uns ist ganz eng mit dem Begriff von Produktivität und Effizienzsteigerung verknüpft. Und da ist meine Rolle herauszufinden, mithilfe von welchen Features. Und man versucht das in meiner Arbeit vor allem auch so, was machen die Leute, was brauchen die Leute, aus eher so einer Nutzerbedürfnisperspektive und gar nicht so sehr aus so einer Lösungsperspektive, da reinzugehen und zu gucken, welche AI-Features und sogenannte AI-Insights, nennen wir das, sinnvoll für Kundinnen sind. Und das ist total abhängig von den Nutzern.

A3

Okay, sehr spannend. Wie ist dein Arbeitsalltag? Gibt es überhaupt dann in so einem diversen Job so einen Arbeitsalltag, einen Ablauf? Ja?

I12

Ja. Also in der Softwareentwicklung arbeiten wir nach Scrum, das ist auch relativ üblich und da gibt es klare sogenannte Scrum-Events, also so bei uns sind das jetzt eigentlich Meetings, aber ich glaube, man kann es auch anders leben, sodass ich da feste Strukturen habe. Wir arbeiten halt in sogenannten Sprints, die sind bei uns zwei Wochen lang und da versucht man innerhalb von diesen zwei Wochen ein bestimmtes Ziel, einen bestimmten Customer Value sozusagen zu shippen, nennt man das, also zu erschaffen in die Welt zu bringen. Und das ist so dieser übliche Arbeitsmodus. Dadurch, dass ich jetzt im UX Research bin, ist es nochmal ein bisschen anders. Ich arbeite jetzt sehr eng mit den sogenannten POs, die halt Produktverantwortung tragen, im Sinne von, die treffen die Entscheidung, in die und die Richtung soll jetzt mein Produkt sich entwickeln oder dieser Teil meines Produkts und ich arbeite ganz eng im sogenannten Sparing mit denen, also das heißt, ich arbeite auf Augenhöhe sozusagen. Ich versuche nicht mit denen so zuzuarbeiten, sondern mit denen zu überlegen, was ist denn jetzt der nächste sinnvolle Schritt und das ist dann so ein hybrides Modell zwischen, ich arbeite schon so im Scrum mit und bin halt mit den Entwicklern und mit den UX Designern an der operativen, ganz ganz vorne sozusagen, wie man das baut, bin aber auch sozusagen, also das nennt man dann Delivery und das andere Fachwort dazu ist dann die Discovery, also dieses vorne weg zu überlegen, was sind mögliche Potenziale, das ist dann die enge Zusammenarbeit mit diesen Product Ownern. Und genau. Ja, Frage?

A2

Ja, genau, ich mache einfach, ich hebe einfach die Hand. Genau, kannst du einmal beschreiben, wo jetzt für dich so die Touchpunkte sind, wo für dich KI ins Spiel kommt, wo das in deinem Alltag, wo dich das berührt?

I12

Also nicht, wie ich das selber benutze, sondern…

A2

Ja doch, doch ja. Genau, wo du es benutzt und wo du, genau, also von dem Prozess, wie entscheidest du dich für welches System und wie benutzt du das und mit welchem Hintergrund?

I12

Okay, also ich glaube, man muss unterscheiden zwischen AI-Tools, die ich selber für meine Arbeit benutze und das andere ist ja, inwiefern wir AI-Tools in unser Produkt reinbauen. Das sind so zwei Sachen und bei dem ersten ist es so, dass ich vor allem nur mit Chat-GTP arbeite, das hatte ich aber, glaube ich, in dem Fragebogen auch angegeben, ich hoffe, das ist in Ordnung. Okay, und da ist bei uns in der Vier…

A2

Du bist da auf jeden Fall auch nicht die Einzige

I12

Genau, bei uns in der Firma sind die meisten Leute so technikaffin und so early adopters, also sind sehr neugierig, was so Technologien angeht und es gibt ein ganz großes Bewusstsein dafür, dass es jetzt so neuartig ist und dann man das auf jeden Fall erstmal ausprobieren muss und wir hatten dann so spezielle Veranstaltungen dazu, wo wir das selbst organisiert haben und immer wieder gefragt haben, hey, wie könnte man das jetzt benutzen? Wofür könntet ihr das denn benutzen? Und ich glaube, so ein Tipp war damals von dem einen, der Entwickler, der auch total sehr begeisterter ChatGPT-Fan ist, der meinte, naja, lasst es doch einfach immer auf und guckt mal in eurem Alltag, wo ihr das verwenden könnt, also reflektiert da einfach mal ein bisschen drüber und das versuche ich jetzt gerade auch, also immer wenn es darum geht, wenn ich irgendwie Texte verfassen soll oder Gedanken irgendwie zusammenfassen soll, dann versuche ich das schon mal zu benutzen, also eher so im wie läuft mein Arbeitsalltag so? Kann ich das jetzt hier an der Stelle eigentlich mal verwenden? Ich glaube, so arbeite ich gerade damit und auf der anderen Seite, wenn es um die Produktentwicklung geht, also inwiefern setzen wir AI-Tools jetzt ein? Ist es so eine Mischung aus strategischer Initiative, also dass irgendwie dem Unternehmen klar ist, durch Marktforschung, durch Beobachtung und so, dass das jetzt so der nächste Trend ist und man diesen Trend nicht verpassen darf, weil sonst machen das ja die anderen und man ist dann dann hinten dran so und deswegen springt man auch auf. Und das wabert dann sozusagen in dieser Gemengelage im Unternehmen so rum bei vielen Menschen und dann gibt es sozusagen so Ideen, die darum flitzen, so man könnte das doch dafür verwenden, die anderen Mitbewerber benutzen das doch dafür, könnten wir das nicht auch machen? Also es ist, glaube ich, ein Viel ein nach außen schauen und sich vergleichen und gleichzeitig sehr viele Köpfe im Unternehmen selber haben, die sofort an Lösungen denken, wie man das verwenden könnte. Und ich glaube, da ist die Herausforderung für mich in meinem Alltag herauszufinden, was sind denn die wirklich wertvollen Dinge und nicht nur so Lösungen, die im Raum schwirren und die dann sozusagen einzufangen, zu strukturieren, zu gucken, welches Bedürfnis kann das denn überhaupt befriedigen? Das ist quasi meine Aufgabe. Insofern komme ich da eher so konzeptuell in Berührung, aber auch, wenn ich dann eher so ich kann auch sozusagen proaktiv anstoßen, so hey, wir haben ja dieses Bedürfnis, können wir das mit AI lösen? Dann gehe ich zu meinem Entwickler und sage, was können wir denn heute noch? Was können wir heute schon? Woran arbeitet ihr? Woran experimentiert ihr denn alle? Also es ist eine sehr komplexe Situation, wo viele Akteure drin sind, wo es keinen festen Prozess gibt, wo viele Leute viel miteinander agieren. Und dann der Trick ist sozusagen dabei, die wertvollen Sachen zu identifizieren.

A3

Und das wäre dieser Discovery-Teil deines Jobs.

I12

Ja, im weitesten Sinne ja.

A3

Wenn man die beiden überhaupt so klar trennen kann, wahrscheinlich wäre das nicht der Fall.

I12

Doch, also man unterscheidet zwischen Delivery und Discovery und ich glaube, das, was ich gerade so erklärt habe, dieses Interne, das geht noch ein bisschen darüber hinaus. Product Discovery würde sich eigentlich nur auf den Value für die Customers sozusagen beziehen. Aber es ist auch eigentlich ein, es ist ein schwer greifbares Konzept, aber es ist ein Konzept, das sich auch auf das, so ein Mindset, was sich auf das ganze Unternehmen dann beziehen kann.

A3

Okay, alles klar. Du hast gerade schon gesagt, du kommst konzeptionell mit verschiedenen KI-Technologien in Berührung, aber auch wenn es dann an die Entwickler geht, wird es ein bisschen konkreter. Kannst du nochmal die konkreten Berührpunkte vielleicht mit den KI-Technologien nochmal beschreiben, die du hast und vor allem, was du da für Herausforderungen wahrnimmst in der Benutzung von KI-Technologien, in der Auswahl vielleicht auch von KI-Technologien?

I12

Ich glaube, wir arbeiten ausschließlich mit Chat-GTP und ich glaube Gemini, aber auch nur, weil das günstiger ist. Das heißt, so viel Auswahl gibt es da nicht. Ich glaube, das passiert sozusagen bei den Entwicklern, die wählen das technisch aus. Damit habe ich gar nicht so viel zu tun und wie das bei uns so funktioniert, es gibt irgendwie so eine Gruppe an Menschen, die das wild experimentiert, das ist aber so voll technisch, das ist null mit Designern, das ist auch nicht mit PO, also die machen irgendwie so allein in ihrem Kämmerlein, experimentieren die und sagen dann, ja, wir können jetzt Summaries, automatisierte Summaries mit Hilfe von AI und ja, wir haben das jetzt um dieses und jenes Feature verbessert oder das ist jetzt noch schneller geworden, also solche Sachen machen die und ich glaube, dann gibt es, also und die machen das sozusagen aus so einer technischen Perspektive und ich stelle mir das so vor, die graben sozusagen von der einen Seite und dann grabe ich aber von der anderen Seite und sage, naja, also für wen ist denn das? Und da ist es mir, glaube ich, gar nicht so wichtig, welche technologischen Ansätze die da haben, sondern mir ist dann nur wichtig, funktioniert jetzt dieses Summaries wirklich so gut, wie es notwendig wäre, um Vertrauen an dieses Tool hinansetzen können, sodass wir auch wirklich Produktivität, dieses Wertversprechen von Produktivität einlösen können oder ist das Quark? So, ne? Und ich glaube, dann gräbt man so von der einen Seite, so hat das Wert für den Kunden und die anderen graben von der Seite, ja, ist es technisch möglich und theoretisch eines Tages, wenn jetzt noch von der anderen Seite gegraben wird, das sind dann die Designer, wie können wir das denn gut visualisieren im Interface? Und so graben wir uns da alle irgendwie an diese KI-Lösung sozusagen ran.

A3

Sehr spannend. Ja, also ich glaube, da in diesem Zusammengraben, da setzen wir auch so ein bisschen an mit unserer Vision. Aber zunächst einmal, bevor ich dir unseren Lösungsansatz zeige, fällt dir spontan ein, wie wir diese Arbeit noch ein bisschen einfacher machen können? Also gerade dieses so, ich sag mal, das auch aufeinander zugraben. Zum Beispiel aus der Perspektive der KI-Forschung. Was es dazu geben könnte. Oder auch allgemein?

I12

Man müsste ja erst mal überlegen, was ist eigentlich das Problem? So, wenn wir jetzt so alle unsere Gräben graben und ich weiß gar nicht, ob es ein Problem ist oder einfach nicht Teil des Prozesses ist, dass wir da uns alle sozusagen alignen müssen, also irgendwie klar sein müssen, auf welches Ziel arbeiten wir jetzt gerade und diese Menschen alle irgendwie an dem Gleichen arbeiten zu lassen, sonst kommt am Ende irgendwas raus, was nicht benutzbar ist und sehr teuer war in der Entwicklung oder zu langsam ist oder auch immer. Also insofern wäre ja die Herausforderung etwa, wie können wir sicherstellen, dass wir alle an dem Gleichen, also alle das geteilte Verständnis haben und alle an dem Gleichen auch arbeiten.

A2

Ja, da direkt eine Anschlussfrage zu. Sagen wir, die haben jetzt hier irgendwie die Entwickler irgendwie fünf bis zehn Produkte gefunden und sowas. Welche Informationen brauchst du von denen? Welche, was ist wichtig, was sie dir geben? Also was das Produkt kann und wie gut das das kann zum Beispiel oder was sind diese, wie sagt man denn, Features, die für dich in deinem Alltag sind?

I12

Also die Voraussetzungen sozusagen?

A2

Ja, ja.

I12

Technische Performance, also wie gut funktioniert das? Im Sinne von, funktioniert das schnell genug? Man befindet sich ja im Telefonie-Bereich, das sind ja Telefongespräche. Wenn man jetzt beim Beispiel der Summaries bleiben möchte, darf das ja auch nicht zu lange dauern, bis diese Summary nach dem Anruf erstellt wird. Also Performance im Sinne von wie schnell ist diese Lösung? Aber auch qualitativ im Sinne von, wie gut kann die technische Lösung das Gespräch zusammenfassen? Wenn wir jetzt einfach bei diesem Feature bleiben, es gibt aber noch andere Ideen, aber ich glaube, daran kann man…

A2

Du kannst auch ganz gerne ganz wild überlegen. Das ist natürlich ein schönes Beispiel, das kann ich auch gut nachvollziehen. Aber du kannst wirklich ganz, keine Ahnung, wenn du dir wünschen könntest, dass… Dir auch mitteilen, was wäre das? Vielleicht ist es auch technisch nicht möglich, aber was sind so, wir sind interessiert an deinen Ideen.

I12

Also was müsste ich über die technische Lösung wissen, um damit arbeiten zu können? Das ist die Frage, ne?

A2

Weil die können dir ja jetzt einfach den Code schicken und damit kann man ja wahrscheinlich nicht so viel anfangen. Deswegen gibt es ja bestimmt irgendwelche Features darüber, die für dich in deiner Arbeit hilfreicher wären, als jetzt irgendwie der Quellcode.

I12

Und das Feature ist aktuell der Typ, der mir das gerade erklärt, was das tut. Und er sagt einfach die ganze Zeit, ja, es ist so einfach, ich schäme mich eigentlich dafür, dass ich dafür bezahlt wäre, weil ich einfach nur die ganze Zeit prompte und irgendwas, weiß ich nicht, mache und völlig überbezahlt bin. Da musste ich mal kurz überlegen, was ich da brauche. Weil ich glaube, ich bin selten in dieser Position, dass ich was von denen will, sondern also andersrum, ich was von denen kriege, sondern ich sage denen ja, was ich will. Und die sollen es dann sozusagen bauen. Also so rum ist es. Wenn wir jetzt in der technischen Entwicklung sind, ist es ja so, wir geben Anforderungen rein und die sollen das bauen, letzten Endes ausführen. Deswegen ist es gar nicht so einfach, das zu benennen.

A2

Aber das kann ja auch, von der Seite kann es ja auch sein. Also wenn du sagst, es soll irgendwie, du hast gesagt, Texte, Gespräche zusammenfassen.

I12

Mal überlegen. Also schnell, möglichst datenschutzkonform. Also ich weiß nicht, wie das funktionieren kann, aber ich glaube, vielen Menschen ist es aber wichtig.

A2

Ja, das ist ein guter Punkt. Den hören wir nicht so selten auch.

I12

Das kann ich mir vorstellen. Der Datenschutz, ich glaube tatsächlich so aus meiner Praxis und das, was ich in der akademischen Welt vorher gemacht habe, dieses woher kommt das, ist nicht so wichtig. Also wie funktioniert das Modell oder dieses, aus welchen Quellen speist sich jetzt etwas, haben jetzt nicht, also ich glaube, das ist nicht in meiner praktischen Lebensrealität.

A3

Darf ich dazu kurz anfangen? Was meinst du mit, woher kommt das Modell? Also für mich würde dazu auch gehören, zum Beispiel Google hat dieses Modell trainiert oder erstellt oder Amazon. Meinst du das als, woher kommt das Modell? EntwicklerInnen quasi.

I12

Wenn man jetzt weg geht von summaries sondern wenn ich jetzt zum Beispiel denke und die Frage gib mir mal die zehn wichtigsten X aus so dann steht da ja keine Quelle oder so und dann

A2

Also im Sinne von XAI, also so eine Erklärung

I12

Genau, im Sinne von, ich weiß nicht genau, woher du diese Information hast, aber ich glaube, es spielt auch in meiner praktischen Welt keine Rolle, woher das kommt. Sondern es ist eher so ein, da kommt ein Vorschlag und ich gucke mal, wie ich damit weiterarbeite. Es ist anders als eine Google-Anfrage. Aber jetzt zurück zu dem anderen Thema. Ich finde das gar nicht so einfach, darüber nachzudenken. Oder darüber zu spinnen, weil ich glaube, effektiv geht es um Performance und dass es schnell ist und dass es reliabel funktioniert und dass bei den Summaries die Namen richtig geschrieben sind und dass sozusagen dieses Produktivitätsversprechen eingehalten werden kann. Und dieses Versprechen entsteht sozusagen dadurch, dass entweder verspricht man sozusagen indirekt, du musst selber nicht mehr diese Summary schreiben. Oder sowas wie, du hast in Zukunft, kannst du immer nachvollziehen, was du besprochen hast und so. Insofern, das ja gegeben ist, ist das meine einzige Anforderung an die technische Lösung. Und außer Datenschutz, was halt vielen Menschen schon wichtig ist, sehe ich da glaube ich jetzt gerade nichts.

A2

Das passt aber, also wir haben ja am Anfang gesagt, es gibt kein richtig und falsch. Und ich glaube, was du bisher gesagt hast, passt das sehr in das, was wir bisher auch gehört haben. Aber wir gehen glaube ich in den nächsten Teil.

A3

Absolut. Genau. Also, wir haben uns natürlich dazu auch Gedanken gemacht. Ich teile mal ganz kurz meinen Bildschirm und haben uns auch überlegt, ja, wie kann man denn jetzt überhaupt Informationen über KI-Modelle kommunizieren und welche Informationen über die KI-Modelle sollte man kommunizieren? Ich gehe mal auf diesen Button. Ich hoffe, ihr seht es gleich groß, das Label. Ja, ne?

I12

Jetzt, jetzt, ja.

A3

Okay, jetzt ist es zu sehen. Und ohne dich jetzt erstmal irgendwie noch weiter zu primen, würde ich dich gerne fragen, was siehst du denn auf diesem Label? Was ist deine spontane Reaktion?

I12

Das sieht aus wie ein Energiehaushaltssiegel oder so für Gebäude. Das ist meine erste Assoziation. Ich weiß gar nicht, woher das kommt. Ich glaube, es kommt von der Form, dass es so wirkt und auch von der Farbwelt sozusagen. Also dieses hell, ich weiß nicht, Türkis oder so, wie man das nennt, die Farbe vom Rand und so. Und dass man ja auch bei Gebäudenergie so von A bis E wahrscheinlich auch spricht. Und dass es so einen technischen Touch hat hier, was man hier sieht. Ansonsten habe ich das Gefühl, geht es jetzt um den visuellen Eindruck?

A3

Alles, was dir einfällt, ja.

I12

Also visueller Eindruck, technisch, irgendwas mit Energie und anscheinend ist das hier, A ist wohl ganz gut und irgendwie sparsam und darunter sehe ich viel zu bunte Icons für so ein Energie-Label. Finde ich ja schon mal nicht so toll. Und was sehe ich da? Ich sehe eine Batterie, Powerdraw per Inference, also den Verbrauch von Energie und die Batterie ist grün, also ist es wahrscheinlich wenig. Corrupted Robustness, kann ich nichts damit anfangen. Auch eine sehr interessante Prozentzahl bei vielen Nachkommastellen. Genauso wie bei den Top 1 Accuracy und Running Time Per Inference. Also ich nehme mal an, Inference sind wahrscheinlich die Anfragen, die man zum Beispiel an ChatGTP senden würde. Und dann wie viel Energie, wie akkurat und wie schnell das ist, aber die Robustness ist vielleicht sowas wie die Relativität, also wie oft ich mit der gleichen Anfrage ähnliche Antworten bekomme. Und die Farbe der Icons indiziert wahrscheinlich gut oder schlecht. Und wieso ist es dann bei A, wenn zwei? Weil zwei Icons grün sind und eins rot und eins orange, weiß ich nicht.

A2

Gute Frage, gute Frage.

A3

Sehr gute Frage. Da bist du auch nicht die Einzige, die sich diese Frage gestellt hat. Ich kann ja mal ein bisschen Licht ins Dunkle versuchen zu bringen und das kurz erläutern, was du hier siehst. Außer dir fällt gerade noch irgendwie spontan noch was ein, wo du eine Frage zu hast oder was dir auffällt. Ja, ich? Meinst du? Ja. Ja, okay. Alles klar. Ja, also erstmal sehen wir hier ganz oben das MobileNetV3 Small. Das ist also das KI-Modell, was wir hier gelabelt haben. Das ist ein Modell, das ist so ein Image-Klassifikator. Sagt dir bestimmt auch was. Also das sagt einfach, was auf einem Bild zu sehen ist letztendlich. Das ist quasi die Aufgabe, die es zu lösen hat. Und das wurde auch dem sogenannten ImageNet-Datensatz trainiert. Das ist ein sehr großer Beispiel-Datensatz, der frei verfügbar ist im Internet. Und ja, das führt... Das führt Inferenz aus. Inferenz ist ja eben genau das, was du schon gesagt hast. Dieser Prozess letztendlich bei Large-Language-Models. Du fragst den irgendwas und der gibt dir was zurück. Das ist Inferenz. Ja, und dann hast du dich ja auch schon richtig gefragt, ja, warum ist denn jetzt überhaupt dieser Gesamtscore oder dieser große Score hier A? Dieser Score speist sich aus insgesamt zehn verschiedenen Untermetriken. Hier sehen wir vier davon und die fließen alle in einen Gesamtscore ein. Das Ganze stammt aus einer Studie, die unser Kollege XXXXXXX XXXXXXX gemacht hat. Der hat insgesamt 30 oder knapp 30 verschiedene neuronale Netze miteinander verglichen und eben diese Metriken immer gemessen pro neuronalen Netz und hat daran eben festmachen können, eine Rangfolge letztendlich. Und daran orientieren sich diese einzelnen Farben. Das heißt, wir haben hier eine grüne Batterie und das hast du auch schon richtig gedeutet. Das bedeutet, dass unter diesen 30 dieses MobileNetV3 tendenziell sehr energieeffizient ist und wenig Energie zieht pro Inferenz. Ja, dagegen ist natürlich die Accuracy, also die Genauigkeit mit 63 Prozent relativ schlecht. Man muss sich aber, man muss dabei bedenken, die ImageNet, das ImageNet-Daten, der ImageNet-Datensatz, sorry, der hat insgesamt 1000 Klassen. Das heißt, wenn man quasi random eine Klasse sich raussuchen würde, dann würde man eine eintausendste Accuracy haben, dass die wahre Klasse eben herauskommt. Von daher sind 63 Prozent dann, wenn man das beachtet, doch nicht so schlecht. Ja, Robustness, wie du auch schon richtig gesagt hast, geht in die Richtung Reliabilität. In dem Fall ist es einfach, dass wir die Bilder, die das MobileNetV3 zu klassifizieren hat, da werden so hin und wieder mal Pixel geändert. Und einfach mal so gewisse Noise, also Störungen werden da eingefügt. Und dieser Robustness-Score sagt dir halt, ab wann, ab welchem Ausmaß von Störungen die Klassifikationsergebnisse des MobileNets eben gestört sind. Das heißt, letztendlich wie robust es eben ist, wie reliabel es ist. Ja, hier sieht man zwei QR-Codes. Einmal zur Studie von XXXXXXX und einmal zu den Ergebnissen. Nein, nein. Ich glaube, eins war zu unserer Label-Webseite. Ich weiß es gar nicht mehr. Ups. Und hier sieht man nochmal die Hardware, auf der die ganzen Tests gelaufen sind und das Framework. Okay. Hast du irgendwelche Impulse auf dieser ausführlichen Erklärung und irgendwelche Fragen?

I12

Ja, ich habe eine Frage. Also du hast erst gesagt, das sind irgendwie 30 Icons im Grunde genommen gewesen oder 30 Kriterien und jetzt sind hier nur vier.

A3

Das sind 30 verschiedene Netze, die waren im Vergleich und die Metriken noch weitere, das sind 10 und von denen sind nur 4 hier angezeigt.

I12

Mein Impuls ist, glaube ich, die Frage danach, für wen ist das gemacht? Nicht für mich, glaube ich. Und ich hatte überlegt, okay, wen kann ich denn für den das gemacht wäre? Mit wem würde ich da zusammen, also für wen in meiner Firma ist das denn eigentlich gedacht? Und ich würde sagen, das sind eher noch Entwickler, die zwischen den Modellen halt aussuchen. Und da frage ich mich aber, ob wirklich diese Kriterien, die vier hier jetzt oder die, was waren das jetzt nochmal? Zehn, ne?

A3

Noch sechs weitere, ja.

I12

die relevanten sind, weil ich doch irgendwie die Perspektive habe, dass es viel um Kosten geht.

A2

Meinst du Kosten, wie teuer das ist, einzukaufen? Also wirklich eine Lizenz zu bekommen?

I12

Zum Beispiel, ja. Und zum Beispiel das mit der Energie, dieses Power Draw, ob das wirklich eine Rolle spielt oder ob da nicht was anderes steht. Zum Beispiel die Accuracy oder Robustness oder Running Time, das sehe ich, weil das alles auch die Geschwindigkeit und die Qualität sozusagen einzahlt. Und dann gibt es aber, glaube ich, noch eine Perspektive, eine Anfrage soll nicht so viel kosten. Ja. Und ich weiß nicht, wenn ich jetzt da so jemanden, einen Entwickler, und das werden dann ja wahrscheinlich noch andere Menschen dabei sein, die das mit auswählen, so, nehmen wir das oder nehmen wir das, dann wird das Thema Geld, Kosten auch eine Rolle. Ich weiß nur nicht, im Vergleich zu den anderen, wie viel jetzt ähm

A3

Ja, also gut, in dem Fall, wir haben jetzt, wie gesagt, Bild Klassifikatoren genommen, die sind alle kostenlos verfügbar zum Beispiel über dieses TensorFlow-Framework. Du redest wahrscheinlich gerade über Gemini und Chat-GBT, für die ihr ja Lizenzgebühren zahlt, ne? Ja, okay, interessant. Ja, ansonsten haben wir ja die Punkte, die du auch genannt hast, bevor du das Label gesehen hast, die hätten wir zum Beispiel hier drin, ne? Also die Accuracy, wie genau ist das Ganze, also die Qualität hast du es genannt und hier die Running Time, die Performance hast du das eben genannt, also wie schnell funktioniert das eben. Ja, und es ist sicherlich eher technisch oder an EntwicklerInnen gerichtet. Ja, ist aber interessant die Frage, siehst du da irgendwie einen Kommunikations-Case den EntwicklerInnen quasi mit diesem Label vielleicht, dass die mit dem Label auf dich zukommen können oder siehst du es sehr als unrelevant an?

I12

Nice to know, kein must have ich glaube eher für technisch affine Menschen die da wirklich Interesse haben aber für meine Arbeit sozusagen für meinen Teil der Produktentwicklung ist es nicht so relevant

A2

Weil du praktisch nicht die Entscheidung trifft also du bist du… bei dir wählt jemand anders aus beziehungsweise du sagst das ist jetzt Top oder Flop bitte zeig mir was anderes

I12

weil ich glaube ich… Nee nicht ganz. Weil nehmen wir mal an ich kriege jetzt das Label und da steht jetzt einfach Gemini so. Und dann haben noch ein anderes Label das ist vielleicht B oder so und da steht ChatGPT drauf so und dann ist aber genau das gleiche Kriterium. Dann sagt mir das nicht ob das diesen kundenwert liefert den ich brauche also weil mir ist es egal wie viel Powerdraw das zum Beispiel hat ich sehe das noch am ehesten bei diesem running time und korrupted robustness und der Accuracy aber wenn sich das… mir geht es ja um das Erleben der Nutzer das heißt ich kann zum Beispiel gar nicht beurteilen ob eine sehr gute Running time und eine mittel gute top one Accuracy bei dem Nutzer am Ende so ein „das fühlt sich jetzt aber qualitativ hochwertig an“ auslöst oder nicht. So das kann ich glaube ich mit diesen Informationen nicht beurteilen und deswegen wäre mir das auch nicht wichtig und ich glaube um das zu machen brauche ich dann schon den Technik Mensch der das dann baut und den UX der das dann designed und dann würden wir das zum Beispiel austesten okay ist denn jetzt ein Modell das langsamer ist und zwar um fünf Prozent wirklich deutlich in der Wahrnehmung schlechter als jetzt eins das ein bisschen schneller ist so und ich glaube nur dann würde es Sinn machen aber ich glaube auch dafür brauche ich dieses Label nicht um das zu beurteilen

A3

Das ist trotzdem ein sehr guter Punkt, den du da bringst. Da würde ich gerne einfach mal auf das zweite Label, was wir hier entwickelt haben, zu sprechen kommen. Es ist natürlich sehr geeignet aus unserer Sicht, vielleicht ist es sehr gut geeignet für einen Vergleich von verschiedenen Netzen. In dem Fall haben wir hier das MobileNetV3, was ich gerade gezeigt habe, auf der rechten Seite und das EfficientNetB4 auf der linken Seite. Da würde ich dich, auch wenn es jetzt wieder sehr technisch ist, auch wieder fragen, was sind deine Impulse hierzu, wenn du einen Vergleich hast von verschiedenen KI-Modellen.

I12

Ich glaube, der erste Impuls ist, dass ich mir noch nicht darüber Gedanken gemacht habe, was dieses A bis E eigentlich bedeutet. Da steht ja, also, mein erster Gedanke ist, wenn ich jetzt darüber nachdenke, was das bedeuten könnte, ist die Qualität des Modells. Aber auch das ist wahrscheinlich ein schwieriger Begriff, was das sein könnte. Oder es besteht ja in eurer Definition sozusagen aus diesen zehn Kriterien, oder? Weil es ja so zusammengerechnet ist. Und dann müsste man sich anschauen, was sind diese Kriterien eigentlich? Ich glaube, es macht insofern Sinn, ich muss jetzt gerade an Nutri-Score und Pizza denken. Dass ich mich schon sehr auch von diesem Nutri-Score beeinflussen lasse in meiner Kaufentscheidung. Also, ich mache schon so, ah ja, da ist C, da ist A, ich nehme das A. Aber auch nur, wenn es in der gleichen Produktkategorie ist. Und ich frage mich, ob hier eine ähnliche Denkweise angemessen wäre. Also, sozusagen diese Skala von A bis E zum Vergleich schon zu benutzen. Aber mit welchen Einschränkungen geht das einher, wenn ich das jetzt anwende? Also, wie jetzt bei den Lebensmitteln ist ja die Einschränkung, bezieht sich nur auf die gleiche Produktgruppe. Ich kann jetzt nicht Chips mit Schoko vergleichen. Und ich frage mich dann, ob eine ähnliche Denkweise auch bei den AI-Labels zutrifft. Also, ob es einen Unterschied macht, ob ich, wie hast du das jetzt nochmal genannt, diesen Image-Dingenskirchen?

A3

Bildklassifikation.

I12

Genau, Bildklassifikationen, genau das wollte ich sagen. Und in LLM oder so miteinander vergleichst. Und auch innerhalb der LLMs zum Beispiel wiederum Einschränkungen bestehen, die das schwieriger machen, das zu vergleichen. Weiß ich aber nicht, das wäre jetzt ein bisschen wildes Spekulieren.

A3

Ja, also genau, wir haben diese Label aktuell nur für Bildklassifikation. Das ist quasi durch dieses ImageNet. Ich sage mal, wenn man so mit Machine Learning mal gearbeitet hat, dann kennt man das halt eben als so ein Bildklassifikationsdatensatz und darüber soll es halt identifizierbar sein, was für ein Task letztendlich durchgeführt wird. Aber es haben echt schon viele Leute auch über Large Language Models geredet, eigentlich sogar der Großteil von denen, die ich mitbekommen habe. Da gibt es natürlich auch Benchmarks. Das ist aber noch nicht so ausgereift wie hier tatsächlich. Also da gibt es auch Zahlen, an denen man das irgendwie festmachen kann, Qualität messen kann. Aber das ist bei Bildklassifikation, was einfach schon viel länger existiert, natürlich ein bisschen ausgereifter, die Forschung, wie man das messen kann.

I12

Das ist interessant, weil die Large-Language-Models sind ja viel zugänglicher für viele Menschen. Das war so der erste Gedanke. Und der andere Gedanke ist, das zeigt ja noch mehr, dass diese Labels ja eher für Techniker sind, weil du ja auch voraussetzt, dass du die Datensätze kennst, oder? Also dieses Image-Net und so. Und mir wäre das ja zum Beispiel auch Wurst.

A2

Ja, nicht unbedingt. Also ich glaube, im Idealfall sind das wirklich Leute, die irgendwie jetzt keine Informatik studiert haben und irgendwie nicht schon mal mit Machine Learning in Kontakt waren, sondern die jetzt irgendwie, keine Ahnung, Engineering unterwegs sind und dann aber für ihr Produkt auch noch eine KI-Losung einsetzen wollen. Und die brauchen dann ein bestimmtes, die brauchen zum Beispiel, genau, die brauchen zum Beispiel Bilderkennung und die können jetzt bei, keine Ahnung, auf welche Library geht man, XXXX? GitHub? Weiß nicht mehr.

SPEAKER\_01

Tens of Flow zum Beispiel.

A2

Und da könnt die dann eine Auswahl treffen von wie viel, wie viel so und so viel, Tausenden von Modellen und dann das geeignete raussuchen. Aber ich fand es sehr interessant, was du gesagt hast. Es war sehr abschlussreich, die Vergleichbarkeit. Das ist was, glaube ich, auch bei uns in den Gesprächen noch nicht viel diskutiert wurde. Also die Vergleichbarkeit unterschiedlicher Kategorien von Systemen.

I12

Ich hatte mich halt noch gefragt, ob... Mir kam dieser Gedanke auf, als du gesagt hast, als Entwickler geht man dann ja irgendwo hin und zieht sich das dann ja. Aber ich glaube, so wie ich das kenne, ist es viel Experimentieren und Ausprobieren. Und die Frage ist halt, ob sowas beim Experimentieren hilft oder ob ich nicht des Experimentieren willen schon alleine nicht auf das Label gucke, sondern halt sage, ich probiere mal einfach irgendeinen... Oder vielleicht ja doch wiederum. Ich nehme mal irgendeinen aus E und ich nehme mal irgendeinen aus B und gucke mal, wie sich das verhält. Also ich hatte erst so den Impuls, ja, klar willst du immer das, was A hat, aber vielleicht ja auch nicht. Vielleicht willst du ja. Ein Salat, so ungefähr. Also mal hier, mal da, weil in meiner Softwareentwicklung möchte ich Sachen ausprobieren. Ich möchte hier schauen, explorieren, welche technischen Möglichkeiten es gibt und so. Vielleicht ist das auch noch ein Aspekt, der da mit reinspielt für den Zweck dieser Labels.

A3

Ja, also quasi als Hilfe zur Vorauswahl von Modellen, wenn ich das richtig verstanden habe.

I12

Ja, wobei ja auch wieder die Frage ist, also ich würde ja auch ohne Label ausprobieren und was raussuchen und dann ist die Frage, was ist denn der Unterschied, wenn ich ein Label dann dazu habe?

A2

Hoffentlich kann man dann informierter vorauswählen.

I12

Ja, aber ich frage mich, ob das im Experimentieren selber sozusagen eminent drin ist, dass es dir egal ist. Also muss ich auch noch darüber nachdenken, weiß ich nicht.

A3

Okay, ja, vielen Dank. Das ist auf jeden Fall ein sehr guter, ja, neuer Aspekt, den wir noch gar nicht hatten als Einsatzzweck für dieses KI-Label. Ich hätte dennoch die Frage, hast du denn hier irgendwas, was dich stört, was dir so wirklich negativ aufstößt oder was du auch einfach sehr schwer zu verstehen findest?

I12

mich stört das… also ich weiß es nicht genau mich stört, dass das Label… Gibt es ein Ablauf Datum des Labels habe ich mich gefragt also die ChatGPT und die LLMs zum Beispiel verändern sich ja auch. Inwiefern wird das Label sozusagen geupdatet wie schnell passiert das also wenn ich nämlich ein Label hab dann habe ich nämlich standardisierte Prozesse also dahinter steht irgendwie nur eine bestimmte bürokratische Struktur und ich frage mich ob die Entwicklung von diesen AI Modellen genau das Gegenteil ist von ihrer Dynamik her und ich glaube das stört mich wie ich diese technische bürokratische perspektive auf etwas abbilden kann was wenig bürokratisch ist so fällt jetzt kein anderes Wort ein um das zu schreiben aber

A3

Fast schon anarchisch teilweise die Entwicklung.

I12

Ja, genau. Es entsteht erst und dann wird es reguliert. Es passiert auch teilweise andersherum, aber so dynamisch diese Softwareentwicklung ist, ist es in der Geschwindigkeit zu schnell, um vorzuregulieren sozusagen oder nur vorzuregulieren.

A3

Sehr, sehr guter Punkt. Also ich meine, wir haben dieses Issued July 24, das soll quasi darstellen, dann wurde dieser Test durchgeführt oder ja, dann wurde diese Klassifikation wirklich entschieden. Ja, haben wir uns dazu Gedanken gemacht, wie wir das updaten wollen? Ich weiß nicht, XXXXX, eigentlich kam das noch nicht so zur Sprache, auch in den vergangenen Interviews nicht.

A2

Nee, es gab, ich glaube, es kam einmal vor, dass jemand auch wissen wollte, ja, von wann ist das Label, aber so ein Updating-Prozess war bisher noch nicht so.

A3

Ja, aber da kommen wir gleich nochmal drauf zu sprechen, wenn es auch darum geht, ja, wer sollte das denn überhaupt ausstellen? Ich würde gerne nochmal…

I12

Achso, nur noch eine kurze Anmerkung zu diesem Ablaufdatum. Wenn es zu kurz ist nämlich, dann ist mir die Information ja egal. Und wenn es zu lang ist, also ich glaube, es gibt da so einen Sweet-Spot zwischen es ist lang genug, dass ich darauf vertrauen kann, dass es reliable, eine interessante Information für mich ist, aber wenn das Label zum Beispiel jeden Monat upgedatet wird, dann gucke ich da auch nicht rein. Also dann verliert es an Glaubwürdigkeit, genauso wie wenn der Zeitraum sozusagen zu lang ist. Also ich glaube, da gibt es ein Spannungsfeld.

A3

Guter Punkt, ja. Okay. Vielleicht wäre an der Stelle nochmal interessant, also hast du noch was anderes, was dir negativ aufstößt an dem Label?

I12

Ich glaube, es bezieht sich alles auf die Frage, für wen soll das sein?

A3

Ja, okay, verstehe. Dann wäre die Frage gerade aus UX-Perspektive, naja, was könnte man denn jetzt nochmal, was könnte man an dem Label anders gestalten, was würdest du verbessern? Vielleicht auch im Hinblick auf diese Frage, an wen ist das gerichtet? Vielleicht aufgeschlüsselt nach Zielgruppe, was würdest du denn hier verändern?

I12

Ich bin kein Designer, ne?

A3

Ja, das stimmt, ja.

SPEAKER\_01

Hast aber schon ein gutes Design-Auge. Das kann ich dir schon mal sagen.

I12

Ja, weiß nicht. Kommt drauf an, mit wem du dich vergleichst. Ich glaube, ich würde anders herangehen. Also mein erster Gedanke ist, dass Fokus wichtig wäre. Also ich glaube nicht, dass man mit einer Art von Label alle abholen kann, sondern verschiedene Zielgruppen haben Bedürfnisse nach unterschiedlichen Informationen. Und so wie das jetzt hier dargestellt ist, ist mein educated guess, dass es eher für die Software-Entwickler ist, also Software-Developer ist. Vermutlich, ich bin ja auch selber keiner, aber vielleicht ist es denen auch wurscht, wie viel Power-Draw das hat. Und wenn man jetzt so, ich weiß es nicht, wie das bei Designern ist, ob das denen wichtig ist. Also aus meiner Perspektive ist halt wichtig Performance und Geschwindigkeit, aber auch da bräuchte ich eine andere Art von Information. Also ich brauche nicht diese 59% oder 68% Robustness, sondern für mich macht diese Information nur Sinn im echten Erleben sozusagen. Also in Beziehung des echten Erlebens. Ähm. Weil die... Die Entwickler freuen sich zum Beispiel, wenn sie den Call 0,2 Sekunden schneller aufbauen können. Und ich denke mir so, ist egal. Also es ist egal, ihr freut euch, es ist technisch super aufwendig. Man freut sich riesig, aber at the end of the day telefonieren zwei Menschen miteinander. Und ich glaube, so eine ähnliche Logik ist für meine Rolle auch bei diesen Labels. Also am Ende geht es darum, ein Erleben zu schaffen. Und dafür, ja, im Telefonie Bereich brauche ich eigentlich nur zu wissen, funktioniert die Summary, funktioniert die Sentiment Analysis, wie sie sein soll, funktioniert die Topic Analysis, so wie sie sein soll, ist die Topic Analysis tief genug in ihrer, ähm, in ihrer Semantik, sodass Kundinnen damit was anfangen können? Oder ist es zu oberflächlich? Das sind eigentlich die interessanten Fragen für mich und nicht, wie accurate das alles ist. Auch aber, ja.

A3

Ja, das verstehe ich auf jeden Fall aus deiner NichtentwicklerInnen-Perspektive, dass du da eben andere Anforderungen hast. Kennst du denn Arten der Kommunikation, wie eben solche Informationen schon an dich herangetragen wurden? Oder ist es etwas, was du dir einfach nur wünschst, was noch nicht so im Arbeitsalltag angekommen ist?

I12

Das ist noch sehr neu. Ich würde behaupten, dass die meisten Unternehmen gerade den Umgang mit AI-Tools erlernen und wie man damit entwickelt und wie man das bepreist zum Beispiel. Und ich glaube, für Unternehmen ist immer, wie viel Euro packen wir denn da drauf? Und hat das dann am Ende jedes Unternehmen, sodass wir da kein Euro mehr dran packen können, weil es sozusagen Standard wird und so? Das sind, glaube ich, gerade die Fragen, die sich viele Unternehmen stellen. Aber auch sozusagen intern, wie schaffe ich es, dass meine Mitarbeiter diese Tools auch nutzen? Wie führe ich sie da heran? Wie verliere ich nicht den Anschluss zum Wettbewerb? Und deswegen, zumindest so aus meiner Erfahrung, wird da viel ausprobiert. Ich glaube, es passiert tatsächlich sehr viel zwischenmenschlich sozusagen. Eher im Erklären, eher im Zusammen daran arbeiten als an Artefakten wie das Label.

A3

Okay, verstehe. Das geht jetzt wieder ein bisschen in die technischere Richtung, aber wir haben uns noch ein paar Arten von Kommunikationsformen angesehen, die es so gibt, die quasi Alternativen zu diesen Labels sind. Da würde ich ganz, ganz schnell drüber durchfliegen. Und zwar gibt es ja meistens wissenschaftliche Paper zu den einzelnen KI-Modellen. In dem Fall wäre das auch wieder zu dem MobileNetV3, was von einem Google-Team geschrieben wurde. Dann gibt es sogenannte Model Cards von Google, wo du auch ganz viele Benchmarks, nennt sich das, also letztendlich auch wieder Statistiken über diese KI-Modelle, die finden sich eben hier relativ gut zusammengefasst. Dann gibt es Papers with Code. Das ist eine Seite, die verknüpft quasi alle Paper, die MobileNetV3 zitieren. Und wenn du hier runterscrollst, dann siehst du eben, welche Publikationen in der Wissenschaft alle schon MobileNet verwendet haben in irgendeiner Art und Weise. Dann gibt es Blog-Beiträge. Hier haben wir ein Beispiel von Towards Data Science, aber auch Medium.com wird viel von EntwicklerInnen gelesen. Dann haben wir die technischen Libraries, wirklich der Programmiersprachen. Da sehen wir nochmal technische Details, wie man letztendlich so ein MobileNetV3 implementiert. Das ist kostenlos an dieser Stelle. Und Fact Sheets, das ist quasi wie diese Google Cards, Model Cards, die ich gerade erwähnt habe, wo du auch nochmal Informationen in Textform hast, hier aufgeschlüsselt nach verschiedenen Kategorien zum Beispiel. Bias ist hier ganz interessant. Das führen die auf, also wie fair ist letztendlich das Modell. Da werden auch solche Sachen wie potenzielle Diskriminierung, die von Modellen entstehen kann, aufgegriffen. Und auch hier wieder. Wie robust ist das Modell? Was sind Inputs? Was sind Outputs? Und verschiedene Metriken. Vielleicht nochmal ganz zum Abschluss. Was ist dein Impuls, wenn du jetzt so zugegebenermaßen im Schnellflug diese sechs verschiedenen Darstellungsformen gesehen hast? Was merkst du hier im Vergleich zu unserem Label?

I12

Also ich glaube, es ist, so wie ich die, als Practitionerin sozusagen, so wie ich die Welt in meiner beruflichen, wie ich mein berufliches Umfeld aufteilen würde, ist, es gibt irgendwie die Entwickler und es gibt die sogenannten Product People. Also Menschen, die sich um die technische Lösung kümmern und ich glaube, fast alle Arten von Kommunikation beziehen sich auf diese Menschen, aber es gibt auch noch diese Product People, die ja aus der technischen Lösung ein echtes Produkt machen und dieses Produkt auch erfolgreich machen. Und zu denen zähle ich mich ja, ich bin auch eine Product Person sozusagen. Und ich glaube, in dieser Product Welt spielen Blogartikel und Trends und was machen die anderen. Und wie positioniere ich mich im Vergleich zu den anderen Unternehmen, spielt das eine mega wichtige Rolle. Deswegen ist, glaube ich, so der Blog im Sinne von, hier, ich bin ein Autor mit Fachexpertise und habe mir diese technischen Lösungen angeschaut und sehe da Folgendes, der diese, wie so eine Art Wissenschaftskommunikation da eigentlich betreibt, plus irgendwie die Trends. Und was sagen andere sehr einflussreiche Blogger und so, eine super wichtige Quelle und die auch genutzt wird tatsächlich, als dass jetzt irgendwie, weiß ich nicht, Paper gelesen werden. Und das Label, würde ich sagen, ist eine Zwischenlösung. Also es ist nicht so tief detailliert, dass es an einen Paper rankommt. Es fehlt aber sozusagen die... Also ich weiß ja zum Beispiel nicht, wer dieses Label kreiert. Wie diese, also wie kann ich diese Meinung, die hinter die, also es ist ja eine Beurteilung und diese Meinung kann ich nicht einordnen. Und ich glaube, bei Blogs ist es nochmal anders, weil man über Trends und ist ein besonders häufig geklickter Blog, da gibt es irgendwie eine Community dazu, kann ich das besser einordnen. Wenn ich jetzt zum Beispiel weiß, ah ja, guck mal, der Mensch XY, den habe ich jetzt hier. Den habe ich schon auf Konferenzen gesehen, der macht auch Blogartikel, der macht auch Workshops. Wenn der jetzt irgendwie was sagt, dann ist das eine Meinung, die auf seiner Erfahrung beruht und der ich irgendwie trauen kann. Und dann aber so ein AI-Label, wie jetzt in eurem Entwurf, das ja eher so bürokratisch aussieht, also so eine Art TÜV ist, müsste sich erstmal positionieren. Okay, wer bin ich denn in dieser Welt? Bin ich dieser Filter? Bin ich so eine Art Wissenschaftskommunikator? Und für wen mache ich das eigentlich? Genau, ich glaube, der Impuls ist so ein bisschen diese technische Lösung, die technischen Möglichkeiten explorieren, aber auch diese Product People, die sozusagen da drauf sitzen und aus dieser technischen Lösung ein echtes Produkt bauen.

A3

Okay, interessant. Ja, da hast du dich gerade auch schon selber gefragt, was mich jetzt auch noch interessieren würde. Was denkst du denn, wer sollte ein solches Label ausstellen? Du hast gerade von dieser, die Personifizierung ist ein Vorteil beim Blogartikel, also dass man auch respektable Menschen hat. Könnte man sowas irgendwie analoges auch für diese Label etablieren? Was sind da deine Gedanken zu?

I12

Also meinst du damit, ob es institutionell sein sollte, wie im TÜV, oder ob es sozusagen Professor XXXXXX sein sollte, die diese Labels sozusagen macht oder wie?

A3

Wäre beides irgendwie eine interessante Überlegung auf jeden Fall.

I12

Eine gute Frage, weil ich glaube,

A3

Oder es muss ja nicht ein Individuum sein.

I12

Ist das auch eine Gruppe, ne? Ja, ja.

A3

Ja, eher eine Gruppe, das wäre dann doch realistischer.

I12

Ja, auch witzig, wenn es eine Person wäre, die dann die Übermacht davon hätte. Ich glaube, das hängt, also ja, es gibt dieses Expertenargument, ist es ja, glaube ich, letztlich. Also ich kann dieser Institution, dieser Person sozusagen vertrauen. Wenn der Name draufsteht, dann ist es ein Argument in sich, dass ich dem trauen sollte. Und ich glaube aber gleichzeitig gibt es auch die Prozesse, die dahinterstehen, die mit dieser Person verbunden oder mit dieser Institution verbunden ist. Also beim TÜV ist es so, ich habe da ganz klare, strikte Prozesse, die alle KI-Modelle durchgehen und eine wahrscheinlich maximale Vergleichbarkeit versuchen herzustellen. Und bei einer Person oder einer Gruppe von Menschen ist es ja ein... eher ein argumentativer Prozess, würde ich sagen. Und jetzt müsste man überlegen, welcher dieser beiden Arten sozusagen bei den KI-Modellen besser geeignet wäre, beziehungsweise mit welchen Vor- und Nachteilen das einhergehen würde.

A2

Ja, schöne Gegenüberstellung.

I12

Und die Frage ist jetzt, dafür müsste man jetzt die AI-Modellentwicklung besser verstehen. Also ist es etwas, was... Mein erster Impuls wäre, eher in die Richtung Community-Personen zu gehen, weil es eine so schnelle Entwicklung ist, weil es sehr viel Fachwissen erfordert. Meine Vermutung wäre, dass eine Institution da gar nicht hinterherkommt. Das sind aber alles pragmatische Überlegungen, würde ich sagen. Also... Ja.

A3

Vielleicht die dritte Überlegung wäre natürlich noch eine Selbstverpflichtung und dass die Hersteller der Modelle das übernehmen anstelle einer Community.

I12

Ja, ich glaube, dann ist es aber kein Label mehr. Also ich glaube, dann höht es die Idee des Labels aus, würde ich sagen. Weil das ist ein bisschen so wie... Ich glaube, die Unternehmen, die dahinterstehen... Ich denke jetzt aber vor allem an Unternehmen und Dinge, die verkaufbar sind. Ja, perfekt. Nicht an die wissenschaftlichen...

A2

Ne es soll auch um die Disziplin gehen

I12

Gibt es da einfach noch einen Zugang für Eigenmarketing? Ähnlich wie, weiß ich nicht, ich versuche, mich als Bioprodukt zu positionieren, also schreibe ich da nicht E-Stoffe drauf, sondern die schönen, verständlichen, ausgeschriebenen Namen und eine ähnliche Analogie würde ich hier sehen und eher eine Gefahr darin sehen. Und das Label, also der Zweck des Labels ist ja immer noch die Vergleichbarkeit herzustellen und ich weiß nicht, ob die Unternehmen intrinsisch sozusagen das möchten.

A2

Label-Washing oder so.

I12

Ja, vielleicht.

A3

Ja, ich meine beim Nutri-Score ist es ja auch gerade noch nicht verpflichtend. Das ist ja auch etwas, was die Unternehmen selber entscheiden. Da sind es halt die klaren Regeln, die befolgt werden können, die dann zu einer gewissen Einordnung führen. Da haben die Leute ja auch irgendwie Vertrauen drin, obwohl es die Unternehmen selber quasi...

I12

Ja, ich glaube, da spielt eher die Rolle, dass es eine sehr leicht zugängliche Information ist, also sehr visuell. Es ist sehr einfach in dem Sinne. Also grün ist gut, rot ist schlecht. Es gibt irgendwie fünf Stufen dazwischen. Wenn ich das in der Mitte nehme, mache ich mir nichts Gutes oder tue mir auch nichts Schlechtes. Und dass es aber zum Beispiel nur einschränkbar vergleichbar ist, ist, glaube ich, vielen vielleicht auch gar nicht bewusst oder in dieser Kaufentscheidung auch gar nicht wichtig. Wenn du bei der Tiefkühltruhe stehst, dann denkst du ja nicht dran, jetzt muss ich hier meinen kognitiven Prozess mal hier anstellen, sondern du guckst dir das an, ist was Visuelles. Ja, da ist grün, da ist rot, ich nehme das Grüne. Das ist eine andere Entscheidungssituation.

A3

Da hängt auf jeden Fall der Vergleich, ja. Ja, okay. Dann aber noch die abschließende Frage, wie, ganz allgemein gesprochen, wie sinnvoll wäre eine solche Zertifizierung für dich, gerade im Hinblick auf deinen Arbeitsalltag und vielleicht auch auf die EntwicklerInnen, mit denen du zusammenarbeitest, die diese Label benutzen könnten?

I12

Irrelevant. Ich sehe da verschiedene Gründe. Also ich glaube, für meine eigene Rolle ist es wenig wichtig. Es sei denn, also das einzige Argument, was mir so ein bisschen einfällt, weil ich jetzt so viel über diesen Power-Draw immer wieder gesprochen habe, gibt schon so ähnlich wie beim Datenschutz so eine Art Umweltbewusstsein auch. Und damit können sich Unternehmen ja auch positionieren. Also Stichwort Greenwashing, aber man kann es auch ernst machen. Und vielleicht ist das eine interessante Information, die ich sozusagen an Kunden weitergeben könnte und sagen würde, hey, wir haben hier die grünste AI. Willst du dafür nicht 50 Cent mehr ausgeben? Und ich glaube, jetzt wiederhole ich mich aber ein bisschen, diese anderen Kriterien sozusagen, die da so numerisch da stehen interessieren mich weniger, sondern mich interessiert was… welche Auswirkungen das im echten Leben hat. Also wie ist 59 im Vergleich zu 69? Und diese Subjektivität die durch die Anwendung sozusagen kommt. Das ist glaube ich das, was für mich zählt und das bildet für mich das Label nicht ab. Und jetzt auf der anderen Seite wenn ich mich jetzt in meine Entwicklerkollegen hineinversetze wäre meine Vermutung auch eher, dass es weniger wichtig für die wäre. Vielleicht nicht irrelevant, aber nicht besonders wichtig, weil sie eh so arbeiten, dass sie alles ausprobieren. Und ich könnte mir vorstellen, wenn man jetzt aber in größere Konzerne guckt, wo man vielleicht auch Dinge nachweisen muss, wo man, vielleicht gibt es ja auch eines Tages Regularien so, dass das eine hilfreiche Information sein kann, um Prozesse, die da sind, auch zu unterstützen. Also vielleicht gibt es ja so eine DSGVO für AI irgendwann. Und dann müssen Unternehmen irgendwie beweisen, dass sie irgendwie konform sind. Ich sehe also, dass diese Labels eher als etwas bürokratisch-strukturelles, womit man Prozesse nachweist, als dass es Informationen zumindest für meine Rolle irgendwie gibt.

A3

Okay, vielen Dank. XXXXX, hast du noch irgendwelche Nachfragen?

A2

Boah, bestimmt. Okay. Nee, ich meine, du hast alles sehr gut klargemacht, was die Vorteile und Nachteile sind und warum es für dich jetzt nicht relevant ist. Ich finde das total, wir sind ja an diesen Randbedingungen eben auch interessiert. Also das ist jetzt nicht so, dass das irgendwie eine verlorene Zeit ist oder so. Ich finde, das war sehr aufschlussreich.

A3

Und vor allem dieser Clash zwischen, ja, wir haben ein bürokratisches Label und wir haben diesen komplett dynamischen Herstellungsprozess von KI-Modellen und alles ändert sich immer. Das war ein sehr guter Aspekt auf jeden Fall. Da hast du uns wirklich zu denken gegeben. Okay, gut, dann würde ich sagen, ja.

I12

Ich stopp mal die Aufnahme.