A1

Genau, sehr cool, dass du dabei bist. Genau. Ja, wir haben eine kurze Vorstellung schon gemacht. Nochmal ein bisschen spezifischer. Also, was ist deine Rolle und was sind so deine Erfahrungen bei der Arbeit mit KI beziehungsweise ML-Systemen?

I8

Genau, ich arbeite ja bei den XXXXXXXXXXX, kümmere mich da an zentraler Stelle um die KI-Plattform. Genau, da sind wir jetzt seit insgesamt gut zwei Jahren dran. Jetzt Ende letzten Jahres haben wir das Projekt, die erste Version abgeschlossen, sind da vor ein paar Herausforderungen gestoßen, wo wir jetzt nochmal in den Review gehen, Refactoring von der ganzen Plattform. Ziel von der Plattform, die ich betreue, ist es, an einer zentralen Stelle die Machine Learning Entwicklung zu ermöglichen, also dass wir quasi Synergieeffekte zwischen Abteilungen heben können, dass wir sagen können, wir haben halt eine Kernkompetenz beim Machine Learning und die arbeitet dann für viele Fachbereiche, das ist dann entweder mit Dienstleistern oder intern. Genau, das Ganze hat sich jetzt nochmal ein bisschen verändert. Wir haben jetzt generative KI dazu bekommen, deswegen gehen wir auch nochmal ein bisschen stärker ins Refactoring, weil wir da halt auch die Verbindung zwischen Machine Learning und Gen AI eigentlich weiter halten möchten oder quasi schaffen möchten. Es werden viele Produkte angeboten und genau das dann jetzt alles so zusammenhalten, die Projekte jetzt begleiten für Machine Learning und das war quasi meine Aufgabe jetzt.

A1

Okay, cool. Ja, Gen AI, das ist was, was wir jetzt natürlich auch irgendwie schon in den ersten Interviews viel mitbekommen haben, dass das gerade das KI-Feld so ein bisschen von hinten aufräumt. Als wir die Studie geplant haben, war zwar ChatGPT schon am Markt, aber das ist jetzt halt wirklich so, im letzten Jahr ist das halt gekommen, dass das immer mehr Firmen natürlich auch machen. Insofern, also gar nicht kritisch gemeint, aber es ist immer spannend irgendwie. Ich glaube, vor zwei Jahren, wenn wir diese Studie durchgeführt hätten, wäre das komplett anders gelaufen. Da wäre KI dann nur Bildklassifikation gewesen oder so als naheliegende Task, sage ich mal. Was sind denn deine konkreten, soweit du das kommunizieren kannst und darfst, deine konkreten alltäglichen Prozesse dabei? Also, was ist so dein Daily Business, was diese Plattform angeht?

I8

Genau, Daily Business ist, wir haben ja verschiedene Projekte, jetzt kann ich am besten daraus mal berichten. Wir haben ein Projekt im Kraftwerksbereich, wo wir Predictive Maintenance auch nutzen möchten, arbeiten damit mit einem externen Dienstleister zusammen und da arbeite ich halt beratend halt zu, dass wir auf der XXXXXXXXXX-Seite einmal das Verständnis quasi auch haben, was da passiert, dass wir nicht eine Blackbox kriegen eigentlich am Ende und wir das auch später überführen können, weil es da auch viel darum geht, dass wir unsere Daten, das ist ja auch, wir sind ja viel im Kritis-Bereich und das wird ja jetzt auch noch weiter aufgesponnen, dass Energieversorger halt mit in dem quasi eigentlich komplett kritische Infrastruktur als kritische Infrastruktur gewertet werden, das kommt da immer mehr hin und dadurch können wir auch Daten nicht abgeben. Also müssen wir irgendwie in die Lage kommen, dass auch Dienstleister bei uns in den Systemen sicher arbeiten können, ohne zu viel einzusehen. Also da ist quasi meine Aufgabe, Verständnis für Machine Learning mitzubringen und dann halt später auch das Onboarding quasi von den Dienstleistern bei uns in den Systemen. Genau, dann haben wir ein Projekt, wo es im Mobilitätsbereich um so Busauslastungsprognosen geht, wo ich operativ die Daten halt auswerte, visualisiere und genau jetzt geht es darum, mit ML-Modellen die, ja quasi vorherzusagen, wie die Auslastung an einer bestimmten Bushaltestelle ist und eben generell, also mit einem relativ KI-Thema geht es quasi darum, die Plattform an sich, also die Infrastruktur so aufzubauen, den ganzen Rahmen zu schaffen, die Machine Learning Operations an der Stelle auch, dass wir die halt auch für Language-Modelle halt auch haben und nutzen können, dass wir die da zur Verfügung stellen, also einmal Rahmenberatung eigentlich ein bisschen operativ dann, wobei der operative Teil aktuell von Werkstudenten übernommen wird und da ich nicht mehr so viel drin bin.

A1

Okay. Superspannend. Also ist ja auch dann einfach sehr projektabhängig, sage ich mal. Und es klingt nach sehr vielen Schnittstellen, die sich teilweise nicht sehr ähneln, die man irgendwie in seiner Arbeit dann versucht abzudecken. Ich kann das ein bisschen nachvollziehen als Hintergrundinfo: Ich arbeite auch in einem Industrietransferprojekt bei XXXX, die bauen Wasserpumpen, sagt ihr vielleicht was. Die sitzen auch in XXXXXXXX und da versuche ich jetzt auch schon im dritten Jahr irgendwie KI sinnvoll in deren Unternehmen unterzubringen, zusammen mit anderen Leuten, da haben wir so eine strategische Partnerschaft. Wir hüpfen da auch immer so alle paar Monate von Projekt zu Projekt und das ist dann mal so ganz viel Data und Business Understanding und Exploration und Prototypen basteln und so. Und das ist sehr unterschiedlich mit welchen Leuten man da zusammenarbeitet, was man dann konkret macht, unterscheidet sich teilweise sehr. Die bauen natürlich auch so eine KI und Datenplattform auf und so und das ist sehr interessant. Das ist auch viel im Umschwung, gerade wenn man so Systeme hat, die es schon lange gibt, in die man versucht, das irgendwie zu integrieren.

I8

Ja, das ist sehr spannend da. Das ist halt gar nicht, weil es gibt ja nicht die eine Antwort. Also natürlich bietet auch jetzt jeder Hersteller auch eigene Lösungen an. Das zeigt halt immer noch, dass die Sprachmodelle, am Ende ist halt eine Schnittstelle. Die sind halt sehr viel einfacher als Machine Learning Sachen. Also ich bin ja auch als quasi jetzt letztes Jahr ja angestellt worden und dann war auch der Plan, dass wir jetzt irgendwann da hinkommen, dass wir auch eine zentrale Datenplattform machen, dass ich die Themen irgendwann übernehme und jetzt halt mit KI werden wir jetzt eine eigene Abteilung schaffen, die mit vier Leuten besetzt wird auch. Also das kriegt einen ganz anderen finanziellen Background auch. Also das hat die Machine Learning, also ich habe ja vorher beim Maschinenbauhersteller gearbeitet, da hat man dann quasi gesagt, hier, wir haben gewisse Algorithmen, die haben gewisse Wirtschaftlichkeit jetzt irgendwie erreicht. Dadurch hat man halt so ein bisschen verkauft das System. Aber das ist halt ein Bruchteil an dem, was jetzt quasi an Investitionen über die Sprachmodelle reinfließt in die Systeme.

A1

Auf jeden Fall, ja. Da ist gerade richtig viel im Umschwung und da ist es sehr spannend mitzubekommen. Was sind denn so die größten Herausforderungen bei diesen alltäglichen Prozessen? Also klar, das hängt immer ein bisschen davon ab, in welchem Projekt du da jetzt konkret gerade aktiv bist, aber kannst du so ein bisschen runterbrechen, was sind so die größten Hürden, denen du so alltäglich begegnest?

I8

Genau, also ich würde ich mal sagen, so die aktuell ist eine große Hürde, dass wir ja XXXXXXXXXX, zum Beispiel Stadtnetze sind. Wir haben viele Fachbereiche und wir haben verschiedene Dienstleister. Also wir haben relativ viele Personen, die eigentlich dann irgendwie auf der Plattform irgendwie arbeiten müssen und da ist halt eine Herausforderung, die halt systemisch so zu trennen, dass sie auch eigentlich voneinander nichts wissen, aber an den passenden Stellen zentrale Komponenten zu öffnen, damit die halt wieder die Synergien halt irgendwie schaffen, also dass man an den passenden Stellen einschränkt, aber dann auch irgendwie an den passenden Stellen offen genug ist, dass man da halt auch ein Argument für eine zentrale Plattform hat. Das wird jetzt nochmal irgendwie spannend, dass wir da auch mit Berechtigung auf Speicherebene, wie sieht es dann mit Dokumentationen aus für die Projekte, wie sieht es mit Dienstleistern aus, aber dann sind wir gerade so ein bisschen in dem Prozess, dass wir halt Erfahrungen sammeln. Datenanbindung ist halt immer auch ein Thema, da geht halt, wenn man so mit dem Machine Learning Teil vergleicht, würde ich fast sagen, dass das zeitlich, also nicht von dem Aufwand her, aber von der Länge bis das umgesetzt ist, glaube ich, hat man schneller ein Modell am Stehen und die Evaluation als wenn man die Systeme vernünftig angebunden und auch den Output dann irgendwo hingepackt hat an der passenden Stelle, weil das halt einfach so viele Absprachen benötigt. Weil das auch wieder andere Softwarehersteller sind, wir haben die Schnittstellen, die gibt es noch nicht, dann räumt man auch viel in den Systemen auf, weil da einfach noch gar nicht so weit gedacht wurde und die Systeme sind dann teilweise zehn Jahre alt, dann macht man meistens noch irgendwie eine Modernisierung in den Systemen und kann dann erst in die Systeme anbinden. Das ist schon irgendwie so die größte Herausforderung, also diese unterschiedlichen Arbeitsbereiche, Arbeitsgruppen und halt die Anbindung eigentlich an die Systeme.

A1

Das sind jetzt alles sehr technische Sachen, also natürlich, klar, dass das so die alltäglichen Herausforderungen sind und auf jeden Fall dieses Data und Business understanding, das ist immer der Löwenaufwand. Das ist auch bei jeder Industriekooperation, in der ich jetzt gearbeitet habe war das so. Bis man zu dem Punkt kommt, dass man wirklich mal ein Modell trainieren kann, das ist irgendwie schon so 75% des Weges. Es ist glaube ich ein bisschen anders, wenn man sich so Monsterprojekte wie LLMs oder auch so Bildklassifikatoren vorstellt, weil da reden wir ja auch von wochenlangem Training dann. Das ist ja nichts, was man dann mal eben trainiert und auswertet, sondern da ist dann dieser Prozess tatsächlich auch nochmal sehr, sehr, sehr umfangreich. Aber ja, gerade bei Industriedaten, wo wir irgendwie von, weiß ich nicht, weniger als 100.000 Samples und weniger als 50, vielleicht 60 Features oder so haben, ist dann am Ende Training nicht der Löwenanteil, sondern überhaupt dahin zu kommen, dass man diesen Datensatz so kuratiert hat oder eine Live-Anbindung hat und so. Weg von diesen technischen Systemen und Infrastruktur-Themen. Wie ist das denn bei dir mit der Kommunikation über KI. Also ich kann mir auch vorstellen, an Schnittstellen hat man ja ganz unterschiedliche Leute sitzen mit ganz unterschiedlichem, fachlichen Hintergrund. Die sind vielleicht auch mehr oder weniger interessiert, oder sogar kritisch gegenüber KI. Ist das auch eine Herausforderung bei euch?

I8

Genau, ja, da gibt es quasi verschiedene Punkte eigentlich. Also es ist so, dass wir jetzt am Anfang vom Jahr gesagt haben, dass eine KI-Abteilung geschafft wird an zentraler Stelle. Das lag auch daran, weil sich so viele Fachbereiche, also wir sind ungefähr 1400 Leute mit ewig vielen verschiedenen Produkten, die quasi eigenständig, relativ eigenständig halt arbeiten und nur mit bedingt zentralen Vorgaben, die haben natürlich irgendwie auch ihre eigenen Budgets und ihre eigenen Ideen. Und da ist natürlich diese Gen-AI-Thematik, ist ja quasi so der Trieb ins Unternehmen reingekommen, dass jeder seine eigenen Sachen machen wollte und das nicht schnell genug gehen konnte. Und dann haben wir irgendwann oder am Ende hat der Geschäftsführer dann entschieden, dass wir dann eine zentrale Stelle machen und alle KI-Projekte in dem Sprachmodellumfeld oder quasi auch im ML-Umfeld, die sind da so ein bisschen mit reingerutscht jetzt, die gab es nämlich vorher schon, dass die von zentraler Stelle verantwortet werden und auch von der zentralen Stelle ins Unternehmen reingehen, weil einfach erstmal einmal quasi von ganz vielen Stellen das Interesse erstmal da war, da was eigenes zu machen. Und um diesen Wildwuchs quasi da zu verhindern. Und dann hat man da erstmal gesagt, hier, Cut, wir kommen quasi auf euch zu, wenn wir was haben, dann an der Stelle. Genau, der zweite Teil mit den Ängsten ist gerade so ein bisschen schwierig, finde ich. Also gerade so eher die Stimmen, die überwiegen, die dann sagen, hier, das ist gut. Wobei ich denke halt immer so, das sind so diese 20 Prozent, die am lautesten schreien. Diese Menge von den Leuten, die das skeptisch sehen, die haben gerade nicht so den Kanal, würde ich behaupten. Ich glaube, da gibt es genug von, die da auch Respekt vor haben, wo es dann am Ende auch irgendwie darum geht, dass da irgendwie sich die Aufgaben halt sehr stark mit verändern werden mit der KI. Genau, da ist am Ende sehen wir da so, also sehe ich da so ein bisschen den Betriebsrat, der da eigentlich der Kanal sein muss, der aktuell aber noch sehr positiv für KI spricht. Weil ich so denke, so, da ist vielleicht eher seine Aufgabe, negativ da auch mal darüber zu sprechen, um den Leuten. Und irgendwie auch ein Sprachrohr zu geben, ne. Wenn jetzt alle die ganzen Führungskräfte alle nur sagen, das ist alles toll, dann ist halt irgendwie schwierig, ne.

A1

Ja, so eine gewisse Technologie-Skepsis ist ja auch was Gesundes. Aber es ist interessant, dass das tatsächlich gerade anscheinend bei euch im Unternehmen nicht so die Aufmerksamkeit irgendwie hat, was diese Kritik daran angeht, sondern alle eher so ein bisschen hyped da drauf sind und sagen, ja „let's go“ Chat-Modelle und so. Wenn man da so in der Bevölkerung einfach mal so rumfragt, kommt man da auf jeden Fall auf andere Ergebnisse. Da gibt es sehr viele, die, ich meine, kein Wunder, 40 Jahre dystopische Sci-Fi-Filme und dann haben die Leute natürlich Angst, dass KI denen alles wegnimmt und alles kaputt macht. Vielleicht auch nicht ganz so unrecht. Fällt dir was ein, was wir als Machine-Learner-Wissenschaftler, die an Grundlagenforschung irgendwie arbeiten, die versuchen, diese Technologie weiterzuentwickeln und natürlich auch nutzbar zu machen, was ein, wo du das Gefühl hast, das könntet ihr besser machen? Das würde uns die Arbeit so krass vereinfachen, den Transfer vereinfachen und so?

I8

Ja, bei dem Machine Learning, also das sind irgendwie diese Labeling (undeutlich), so quasi den Ansatz zu sagen, welches Modell kann man initial jetzt erstmal verwenden, das ist ja der Wissenstransfer, das ist halt spannend, zu sagen, ich habe ein Problem, das ist ungefähr der Rahmen von den Daten und dafür hat das Modell gut funktioniert, das nimmt halt Aufwand auf unserer Seite weg. Oder dann später, jetzt aktuell bin ich so ein Freund von Auto ML und da dann, wo dann halt so ein bisschen was für dich abgenommen wird oder wenn man dann halt so Bibliotheken hat, wo dann halt solche Modelle für eine gewisse Problemstellung irgendwie gesammelt werden, die dann halt, die man dann einfach verwenden kann, finde ich irgendwie spannend. Bei Gen AI ist halt die Erklärbarkeit, das ist so ein bisschen quasi eine Review machen, was ist hinter dem proprietären System eigentlich dahinter. Also weil die sind halt am Markt, sind die halt am stärksten, wir werden so große Sprachmodelle halt nicht selber hosten werden, das ist zu teuer für fast alle und da dann halt so in diese Einkaufssysteme ein bisschen mehr Erklärbarkeit aus wissenschaftlicher Sicht reinzubringen, das fände ich auch unheimlich spannend und ich glaube für so ein Unternehmen ist das auch interessant, dass man da mal hinguckt, wo, wie funktioniert das eigentlich das ganze.

A1

Genau, also Erklärbarkeit ist auf jeden Fall auch in der Forschung natürlich ein Riesenthema und das geht auch auf verschiedenen Ebenen, also sei es so generell das Modell, wie es funktioniert zu erklären, da hilft aber natürlich auch eine Erklärung ja, das sind die und die Neuronen, die verrechnet man miteinander, am Ende ist das alles irgendwie eine riesige Matrix-Multiplikation, da kommt ein Ergebnis bei raus, das ist ja keine Erklärung, die einem Menschen irgendwas nützt, insbesondere wenn er keinen technischen Hintergrund hat, also da passiert tatsächlich viel, dass man versucht, dieses Blackbox-Modell so ein bisschen transparenter zu machen und zu öffnen würde ich jetzt hier tatsächlich ein bisschen ausklammern, weil das ist nicht das, was wir mit den Labels im Kern machen, klar, mit den Labels das ist schon, ja, deswegen machen wir die Studie ja, natürlich. Da muss ich jetzt aber tatsächlich nochmal zurückfragen also ist das was, was dir in deinem alltäglichen Alltag und in deinen alltäglichen Prozessen nützen würde, oder hast du eher das Gefühl, das ist eine gute Idee, aber eigentlich bist du nicht Zielgruppe, da bin ich mir nämlich gerade ein bisschen unsicher

I8

So wie sich das gerade entwickelt, bin ich nicht Zielgruppe mehr. Also da wäre dann eher quasi, quasi die die halt operativ wirklich das umsetzen. Das hat jetzt erstmal quasi mit dem Rahmen, also da bin ich ja gerade irgendwie aktuell halt relativ aktiv. Da wäre das halt dann irgendwie eher dann die Zielgruppe, dass man da sagt halt, hier, ihr habt das Label, da sind halt gute Modelle, man hat einen Katalog, quasi mit den Modellen, die gut funktionieren. Das kann ich mir vorstellen, dass das gut klappt, aber würde ich auch eher sagen, dass ich da nicht die Zielgruppe für bin.

A1

Aber du sagst, es gibt eine Zielgruppe und es gibt irgendwie einen Bedarf dafür und du würdest ja solche Projekte vielleicht verwalten und würdest du ein Reporting von den Ingenieuren haben wollen, die jetzt an Lösungen gearbeitet haben, an konkreten und für dich wäre das dann vielleicht hilfreich, dass du diese Labels zurückgemeldet bekommst und dann vielleicht auf einer höheren Ebene man dann auch Entscheidungen auf der Basis treffen könnte.

I8

Genau, wenn man das halt sieht, dann wäre wieder die Synergieeffekte heben über die Projekte hinweg. Also am Ende ist das ja dann bei den ML-Aufgaben ja auch, der Dienstleister fängt auch ungefähr von einer grünen Wiese an, also mit den neuen Daten und der wird ja auch seine Erfahrungen haben und das ja quasi für mich so ein bisschen, so verstehe ich das zumindest, ein zentraler Katalog, welche Modelle halt gut funktionieren und welche eher nicht so. Das ist halt ein guter Einstieg eigentlich am Ende.

A1

Ja, cool. Ja, das ist so ein bisschen auch genau das, was wir natürlich erhoffen irgendwie. Also vielen lieben Dank für das Feedback. Das ist genau die Zielgruppe, die wir irgendwie haben. Ich würde tatsächlich an der Stelle dann jetzt einmal konkreter auf die Labels irgendwie eingehen und meinen Bildschirm dafür teilen. Dafür muss ich nur mal eben erst die PowerPoint-Präsi starten. Das vergesse ich jedes Mal, dass ich das vor dem Gespräch machen sollte, weil irgendwie braucht PowerPoint auch irgendwie immer ewig. Und dann teile ich mal meinen Bildschirm. Zack, zack, zack. Let's go. So, du solltest meinen Bildschirm sehen.

I8

Jetzt, ja,

A1

Ja, perfekt, sehr gut. Ohne, dass ich jetzt hier direkt irgendwie groß erkläre, was du hier siehst oder ähnliches, kannst du mir mal eine spontane Reaktion auf das Label geben? Was siehst du, was fällt dir auf, was sind deine Gedanken dazu?

I8

Ich hatte mir ja mal die Beschreibung durchgelesen, also das so ein bisschen Energielabel für Waschmaschinen. Von der Darstellung trifft es mich da noch ganz gut.

A1

Ja klar, wir haben natürlich im Vorfeld schon auf der Webseite auch ein Label dargestellt und so, also es ist jetzt nicht die spontane allererste Reaktion, aber vielleicht kannst du dich noch reinversetzen, was so dein erster Gedanke war oder...

I8

Ja, also gut dargestellt, sieht modern aus und unten die Symbole finde ich auch gut gewählt also die sind, man kann das Symbol sehen, wenn man die Zahl direkt drunter auch ohne das durchzulesen, kommt man irgendwie da schon mal also ich habe eine Einschätzung, was ihr damit darstellen wollt, genau. ImageNet, so insgesamt gibt es eine gute Übersicht

A1

ImageNet sagt dir auch was?

I8

Nicht direkt, das geht eher wahrscheinlich Richtung ResNet oder so also ImageNet kenne ich jetzt nicht mehr

A1

Mhm, okay. Das kann ich direkt einmal erklären. ImageNet ist tatsächlich der Datensatz, der in dieser 2012er Challenge vorgestellt wurde. Das konkrete Modell, was wir uns hier anschauen, ist ein MobileNetV3 Small, aber es gibt auch verschiedene ResNets, die praktisch auf ImageNet trainiert wurden. Also die ResNet-Architektur wäre praktisch ein Competitor, also ein Gegenmodell sozusagen zu dem Modell, was jetzt hier gelabelt wurde.

I8

Okay, das wurde mir jetzt nicht klar, finde ich aber nochmal spannender, dass also dass ja ihr den Datensatz dahinter habt, also es gibt ja für Zeitreihe-Daten gibt es ja auch da diesen Test-Datensatz, wo das hingeht Das ist natürlich nochmal irgendwie interessanter halt, also das wird mir jetzt gerade nicht klar, dass da oben der Datensatz halt dann ist.

A1

Ja, ist auch immer ein bisschen die Frage, mit welcher Expertise natürlich die Leute da drauf gucken. Also es ist, ja genau, erzähl erstmal gerne weiter. Also ja, das nur als Erklärung. Es ist MobileNetV3 trainiert und evaluiert auf ImageNet und genauso könnte man das für ResNet machen, ja.

I8

Ja, das ist halt immer so ein bisschen, für mich wäre das jetzt so ein bisschen, man muss halt die Begriffe einmal kennen, also jetzt das ImageNet oder halt das MobileNet jetzt auch da das Modell ist. Genau, das wäre nochmal ganz interessant und dann halt der Vergleich halt über die Daten mit der Accuracy, genau. Genau, die PowerDraw per Inference ist das dann, also pro Ausführung wäre das dann die, wie viel Energie das quasi kostet?

A1

Korrekt, genau.

I8

Auf, ist halt da dann die A100, das ist mir jetzt auch nicht ganz klar, ist das eine Grafikkarte dann?

A1

Ja, genau, die A100 ist von Nvidia, ist eine Grafikkarte und das ist jetzt praktisch ein Cluster-Node bei uns, wo acht GPUs drin sitzen mit einer TensorFlow-Installation. Das heißt, das ist praktisch das, ich bezeichne das immer liebevoll als Environment, in dem dieses Modell halt evaluiert wurde. Hardware und Software ist halt sehr komplex und nicht so einfach auf ein Wort oder zwei Wörter runterzubrechen. Ich habe zum Beispiel auch mal den lustigen Moment gehabt, dass ich mich gewundert habe, es ging auch um Mobile-Net und die Accuracy war halt so 20% schlechter als im Paper-Reported und ich habe wirklich, was ist hier los? Das kann nicht sein. Alle anderen Modelle passen zu den reporteten Ergebnissen. Dann habe ich TensorFlow von 2.3 auf 2.5 oder so geupdatet und auf einmal passte die Accuracy. Das heißt, die pre-trained Gewichte von dem Modell waren faulty oder irgendwo waren Implementierungsfehler, weil das Modell wird auch so von TensorFlow halt bereitgestellt von Keras. Also man hat halt keine, ein Modell ist nicht so und so genauso oder so und so schnell oder so und so Ressourceneffizient, sondern das ist halt immer gegeben der Umgebung. Ja, genau deswegen ist das da mit drauf vertreten. Da haben es Kühlschränke einfacher, die ziehen halt immer gleich viel Strom, egal wo die aufgestellt werden. Wobei, wahrscheinlich schon ein Unterschied, ob die im Keller stehen oder in einer 30 Grad Dachgeschosswohnung wie bei mir.

I8

Das wird ein Unterschied machen. Was mir jetzt an dem Label nicht so ganz klar ist, also dafür, glaube ich, fehlt einfach so ein bisschen die Erfahrung noch mit, mit A, B, C, D, E, also wie kann man das überhaupt einsortieren? Also unten mit den Zahlen, da hat man irgendwie so ein Gefühl für, weil da eine Zahl steht. Das ist natürlich auch erstmal irgendwie, man muss ja irgendwie einen Vergleichswert dafür haben. Aber das mit dem A, B, C ist halt dann erstmal schwierig zu deuten, was dahinter ist.

A1

Ja, kann ich dir kurz erklären. Im Endeffekt wird jede Metrik, die unten reported wird, für sich im relativen Vergleich in so Bins unterteilt, in fünf Bins. Das heißt, von den getesteten Modellen bedeutet Grün, es ist unter den Besten, Rot unter den schlechtesten, Gelb irgendwo im Mittelfeld. Aus allen evaluierten Metriken, und tatsächlich sind das auch noch mehr als die, die unten dargestellt sind. Die haben jetzt nur die wichtigsten dargestellt, weil wenn man alle zehn darstellt, dann sieht man halt gar nichts mehr auf dem Label. Ergibt sich praktisch ein Gesamtscoring. Und das Interessante dabei ist, dass das natürlich auch angepasst werden kann, diese Gewichtung von diesen Metriken auf den konkreten Use Case. Also man könnte halt auch sagen, ganz ehrlich, ob das jetzt 80 oder 70 Prozent Accuracy hat oder so, ist nicht so wichtig, aber wir haben diesen harten Constraint, dass das Modell innerhalb von einer Sekunde ein Ergebnis rausschmeißen muss. Und das ändert dann ja dieses komplette Scoring so ein bisschen. Auch da ist natürlich die Analogie bei Kühlschränken und Co. wieder ein bisschen anders, weil da hat man diese Priorisierung in der Regel nicht. Es ist einfacher zu sagen, was ist das beste Modell? Bei Machine Learning ist es halt immer super abhängig davon, was man für Constraints hat, was man für ein Use Case hat, was für Prioritäten das Unternehmen hat. Wenn das Unternehmen sagt, wir sind ein grünes Unternehmen, dann will man vielleicht da auch ein bisschen auf Ressourceneffizienz achten, zum Beispiel. Klar, auf Management-Ebene wollen die eigentlich unten nur noch sehen, wie viel Geld bringt mir das und wie viel Geld kostet mich das. Das ist aber natürlich sehr schwer in Zahlen zu fassen, das weißt du wahrscheinlich auch. Ja, genau.

I8

Ja, das ist natürlich interessant, dass irgendwie, jetzt hast du jetzt auf Bildverarbeitung irgendwie bezogen, wie es dann eigentlich aussieht, so, ich weiß nicht, für einen Random Forest oder wie ist das dann halt bei so ganz einfachen Geschichten, so ein Label, wie stellt man das dann dar? Das ist jetzt ja nicht, dass wir für jeden Fall halt immer diese großen Modelle haben, die dann irgendwie auch einen Impact haben.

A1

Ja, bei kleineren, also man kann prinzipiell so ein Label natürlich genauso gut auch für einen Random Forest ausstellen, habe ich auch schon gemacht in einem Paper. Also es ist halt immer abhängig davon, welche Aufgabe man hat. Für so einen Datensatz wie ImageNet wird ein Random Forest einfach nicht funktionieren, weil das eine Komplexität, die Lernaufgabe und der Datensatz hat eine Komplexität, dass man Convolutions braucht. Man braucht irgendwie eine effiziente parallele Verarbeitung von Bilddaten, man braucht so spezialisierte Sachen wie Convolution Layers in dem Netz, damit halt da auf visueller Ebene interessante Muster erkannt werden können. Ein Random Forest, der einfach so random Bilder splittet, ist für diesen Einsatzzweck einfach nicht geeignet. Insofern würde da ein Label wahrscheinlich gar keinen Sinn machen. Man könnte das machen, dann steht da irgendwie, weiß ich nicht, vielleicht 5% Accuracy oder so und wahrscheinlich deutlich höherer Stromverbrauch, weil es nicht vernünftig parallelisierbar ist.

I8

Genau, wenn man diesen Standarddatensatz nimmt.

A1

Genau. Aber wenn man jetzt einen tabellarischen Datensatz aus der Industrie hätte, könnte man das genauso gut, man könnte einfach mit AutoSK Learn oder so sich 30 Modelle aus dem Ensemble durchziehen, die einmal durchprofilen und benchmarken und die Ergebnisse genauso visualisieren. Also es ist prinzipiell völlig übertragbar und tatsächlich über die Software, die ich irgendwie so geschrieben habe in meiner Dissertation auch möglich. Da kann man einfach ein Pandas DataFrame reinladen und der spuckt einem dann solche Labels aus. Genau. Ein großer Faktor hast du jetzt aber tatsächlich gerade auch schon so ein bisschen angesprochen. Natürlich, also ermöglicht das erstmal so schwer verständliche Paper oder abstrakte Implementierungsdetails oder so ein bisschen zu abstrahieren und zugänglicher zu machen. Insbesondere ermöglicht das aber auch den Vergleich von verschiedenen Labels. Da können wir auch noch mal ein bisschen drüber sprechen, was dir hier so auffällt.

I8

Genau, ist ja relativ schön dargestellt. Man sieht ja auch unten sofort, dass die Accuracy irgendwie dann höher ist, halt auf Kosten wieder vom Stromverbrauch und der Laufzeit am Ende, bei der Robustheit, was damit gemeint?

A1

Ja, das erkläre ich, glaube ich, auch in fast jedem Interview. Also die Accuracy ist ja einfach gegeben der Testdaten, die Accuracy. Bei der Robustheit nimmt man die Testdaten, verändert die leicht, rechnet da Rauschen drauf, verzerrt die Bilder ein bisschen, versucht irgendwie so Objektiv- oder Belichtungsänderungen künstlich hervorzurufen, um abzuschätzen, wie sehr verändert sich die Genauigkeit, wenn die Daten nicht mehr eins zu eins so sind wie im Testdatensatz. Das ist also so ein bisschen eine Möglichkeit, um die Robustheit des Modells gegen äußere Einflüsse abzutesten. Das ist insbesondere ganz relevant im Bereich autonom fahren und so, weil man natürlich sicherstellen möchte, dass wenn da jetzt gerade irgendwie ein bisschen die Linse von der Kamera oder das Glas vor der Linse von der Kamera beschlägt oder so, trotzdem die Accuracy hält. Und wenn man halt nur Image-Net nimmt, dann sind das alles sehr saubere, ordentliche Bilder, fast alles. Deswegen ist dieser Bereich Robustness-Testing und Robustness-Training auch, also entsprechend versuchen dann Leute im Training auch schon so was zu machen, um das Modell von vornherein robuster zu machen. Halt ein großes Thema. Genau.

I8

Spannend wäre dann noch mal wie so, diese für die Energielabel die kann man ja tatsächlich ausrechnen da gibt es ja quasi eine zertifizierte Formel für da wäre dann interessant ist das quasi euer Ziel da auch so eine zentrale Formel zu haben die dann vorgegeben wird oder ist er dann auch eher so auf Unternehmens oder in der in den Unternehmen schon Label gedacht?

A1

Ich glaube, das ist ein bisschen was von beidem. Also wie gesagt, grundsätzlich ist das erstmal Grundlagenforschung hier und das ist natürlich alles nicht Industriestandard und im Einsatz und komplett durchexerziert, aber das ist genauso ein bisschen Sinn der Studie auszuloten: Ist für sowas überhaupt Bedarf da? Hilft das Leuten? Macht es Sinn, sowas zu machen? Wir haben jetzt den AI-Act, das heißt, die Firmen werden jetzt irgendwie dazu verpflichtet, bei der Entwicklung von KI-Produkten sich irgendwie ein bisschen Gedanken zu machen und nicht einfach frei raus, damit möglichst viel Geld zu machen. Da werden solche Instanzen halt irgendwie kommen und genau da wollen wir halt gucken, ob so ein abstraktes Labeling, wie wir das hier entwickelt haben, vielleicht da hilfreich wäre und vielleicht wie so ein Zertifikat auch irgendwie funktionieren könnte. Und dann müsste man natürlich diesen ganzen Prozess da, ich meine, so wie dieses Label zustande kommt, das steht in meinem Paper. Das kann man schon nachlesen. Wer das wissen will, der findet die Informationen dazu. Genauso wie man sich irgendwie darüber informieren kann, warum ein Impfprodukt an den Markt kommt zum Beispiel. Da gibt es ja, also diese Studien und alles, was da im Hintergrund steht, das ist an sich ja öffentlich. Aber genau, es ist ein großes Fragezeichen eigentlich, wie man sowas am besten auch macht. Also die Frage kann ich eigentlich nur zurückstellen.

I8

Ja, aber genau so grundlegend halt da irgendwie so einen Aufschlag zu haben, dass man da dafür einschätzen kann. Erwähnt natürlich, genau jetzt, also ich glaube da hat sich Einschränkung auf Machine Learning wieder auch so ein bisschen klarer. Ich glaube mit der generativen mit fremden Modellen wird es dann irgendwie nochmal eine Ecke schwieriger, die erklären zu können. Aber es war vielleicht auch dann wieder ein Ausschlusskriterium vielleicht für das ein oder andere Modell dann, ne?

A1

die Eigenschaften, die wir unten reportet haben, die könnte man jetzt nicht eins zu eins, weil man bei Sprachmodellen nicht von Accuracy spricht, aber natürlich gibt es da auch Metriken, um die Güte des Modells praktisch so ein bisschen abzuschätzen. Da passiert auch gerade viel halt in die Richtung, dass man so Qualität von Sprachmodellen nochmal mehr misst als rein empirisch, indem man Leute fragt, ist das plausibel oder nicht. Ich glaube, auf der einen Seite ist so die Benutzbarkeit von ChatGPT und dass es halt so viel benutzt wird, irgendwie auch schon Güte, also das sagt schon aus, dass das anscheinend schon ziemlich gut funktioniert, aber natürlich findet man auch immer wieder Randfälle und Corner Cases, wo es halt gar nicht funktioniert. Und insofern würde ich jetzt sagen, man kann das auch alles labeln, dann ist aber natürlich die Sache, dass zum Beispiel OpenAI halt ihre Modelle auch gar nicht public rausgibt. Das heißt, man kann die nicht auf einer beliebigen Hardware nochmal selber durchprofilen. Die Bezahlmodelle von denen, die man praktisch nur remote nutzen, mit Internetanbindung nutzen und dann praktisch ein Abo abschließen muss, da ist man natürlich ein bisschen darauf angewiesen, dass OpenAI solche Informationen bereitstellt, weil man das nicht durchmessen kann selber. Was sie nicht tun bisher. Also soweit im Großen und Ganzen, bis auf Robustness, sagst du, ist das alles auch recht verständlich, was auf dem Label zu sehen ist? Klar, nicht ganz nachvollziehbar, vielleicht im Detail, wie die Scorings und so zustande kommen. Wobei man das halt, wenn man oben den QR-Code irgendwie folgt und sich unser Paper dazu anliest, könnte man das rausfinden. Hast du denn das Gefühl, dass hier Informationen fehlen, wo du sagst, also es gibt so Messwerte, da gucke ich immer drauf oder das würde mich total interessieren und das fehlt hier komplett.

I8

Ja, fällt mir gerade schwer, weil die so generell geben den guten Abriss, muss ich nochmal irgendwie mal sacken lassen, ob mir da irgendwie noch was einfällt, das wäre dann wieder halt so ein bisschen im Doing halt, was mir da vielleicht so ein bisschen fehlt, dann, dass man da irgendwie öfter mal drauf guckt und dann mal vergleicht, was sind eigentlich so dann die Vergleichswerte, wo es irgendwie hingehen soll. Gut, dann mache ich nochmal ganz...

A1

So ein bisschen hatten wir das vorhin im Bereich Sprachmodelle, wo du gesagt hast, da wären Erklärungen oder so die Erklärbarkeit, die Transparenz des Modells, wenn man das irgendwie messen und kommunizieren könnte. Das hattest du vorhin schon mal angebracht, glaube ich, ne?

I8

Aber die also so das Label gibt ja schon einen guten Abriss wenn jetzt noch so ein bisschen gucken würde wäre so falsch positive also, dass man da immer so die, die precision recall kurven da vielleicht mal angucken kann dass man in dem Bereich hat reinschaut wie sieht es dann irgendwie halt aus. Die Accuracy ist halt immer noch relativ simpel dann ich weiß aber auch nicht ich habe da mit dem imageNet dann zu wenig Erfahrung wie das im Bildbereich hat die Messung der Genauigkeit aussieht. Genau mit dem Datensatz ich würde wieder wenn ich da Label nutzen würde, würde ich irgendwie sagen ich gucke mir den Datensatz an, guck dann was da drin ist kann ich den mit meinem Datensatz vergleichen und würde dann halt schauen welches Modell dann am besten performt hat nach meinen wichtigsten Kriterien dann wäre immer so eine Beschreibung vom Datensatz im Hintergrund hat irgendwie schön also welche Eigenschaften hat dieser Datensatz eigentlich vielleicht auch schneller zu weil am Ende glaube ich würde ich da halt nicht reingucken ich will gucken ob die Beschreibung passt ,ich kenne ja meinen Datensatz und würde das halt irgendwie vergleichen das wäre spannend

A1

Das versteckt sich so ein bisschen hinter dem unteren QR-Code. Also der untere QR-Code linkt praktisch zurück zu dem Papier, in dem das Modell vorgestellt wurde. Und da steht dann natürlich auch drin, welcher Datensatz verwendet wurde und so. Also man kann immer beliebig in die Tiefe gehen. Das ist echt so ein Trade-off, wie abstrakt möchte man halt bleiben. Und in dem Moment, wo man hier jetzt so eine Wall of Text, so eine textuelle Beschreibung draufpackt, ist man wieder an einem anderen, also da ist man wieder auf einer anderen Darstellungsebene. Aber also keine Kritik, das ist nur, aber es ist halt, deswegen sind wir interessiert, was so auf diesem Label jetzt vielleicht auch fehlen würde noch.

I8

Ja genau, wenn man überlegt, wie kann ich mit dem Label jetzt arbeiten, um da jetzt mal eben, wenn ich mir jetzt mit dem ImageNet, dann habe ich ja erstmal eine gute Übersicht. Dann würde ich halt vielleicht mal starten mit dem ersten Modell, würde da halt mir die Bibliothek ziehen, damit mal machen. Und dann halt in der zweiten Stufe würde ich halt einmal genauer reingucken und finde, da wird über die QR-Codes eigentlich ganz gut abgebildet, wenn man da reinschaut. Das ist halt alles dann auch, ja man muss ja irgendwie einen Einstieg liefern und das ist eigentlich so ganz gut dargestellt, finde ich gerade. Dass dann für bestimmte Anwendungsfälle das dann nicht passt, wo du sagst, man muss die Gewichtung dann irgendwie anpassen, das ist ja auch irgendwie vollkommen klar. Und dass die Hardware sich irgendwie dann verändert oder so, aber dass man so eine grobe Richtung irgendwie findet, dass man auch sagt, man hat eine Entscheidungsgrundlage, wohin man überhaupt geht halt. Das heißt, dann kommt man vielleicht so ein bisschen davon weg, zu sagen hier, entweder man hat Erfahrungswerte oder es hat halt Bauchgefühl, wo arbeite ich am liebsten mit. Und dann vergesse ich vielleicht irgendwie nochmal ein Modell, was dann irgendwie weniger Energie verbraucht und besser ist. Also gibt auch vielleicht nochmal eine bessere Übersicht, wenn ich dann das eine Modell steckt halt in der einen Bibliothek, das andere in der anderen, dann, wenn ich dann nicht gerade jetzt jede Zeitschrift lese oder jede Veröffentlichung kriege, das vielleicht auch nicht mit. Gerade in der Industrie ist das ja, dann guckt man vielleicht auch nicht so oft nach links und rechts.

A1

Ja, auf jeden Fall. Bzw. auf jeden Fall nicht. Genau. Nee, auf jeden Fall. Vielen lieben Dank für die Einblicke. Wir haben jetzt gerade schon so ein bisschen drüber gesprochen. Was für andere, also wo holst du dir denn, wo würdest du dir denn Informationen holen? Also so Labels, das ist ja noch nicht so richtig am Markt, aber welche Kommunikations- oder Repräsentationsformen, Informationsquellen zu Machine Learning Modellen und Algorithmen kennst du?

I8

Ich habe viel mit IEEE gearbeitet eigentlich, da direkt nach Papern gesucht und sonst finde ich halt Medium, ich weiß jetzt ja dann quasi offenes, ja, offene Plattformen, irgendwie finde ich die da relativ gut, hat da nicht ganz diese Veröffentlichungscharakter, aber man findet da schon mal irgendwie gut aufbereitete Informationen mit guten Beispielen. Genau da gucke ich eigentlich, das sind eigentlich so die, IEEE jetzt nicht mehr, ja, Medium gucke ich normal rein, aber das ist halt genau die Stelle, wo ich jetzt halt nicht mehr unterwegs bin, aber ja, das ist halt Publikationen sind da glaube ich so immer so der Hauptteil am Ende.

A1

Genau, wir haben eine kleine Liste zusammengestellt, Publikation natürlich dabei und Blogpost hier jetzt towards Data Science, aber liegt ja bei Medium auch mit dabei. Es gibt auch noch Model Cards, die werden zum Beispiel auf so Library-Plattformen wie Hugging-face verwendet, ist ursprünglich so eine Google-Idee. Im Endeffekt so eine kleine Mini-Betriebsanleitung für ein Modell, sehr ähnlich wie die Factsheets von IBM, das ist allerdings auch alles patentiert und wird nur bereitgestellt für IBM-Produkte und Modelle und für IBM-Kunden. Ist nicht publicly available sozusagen, aber Model Cards kann ja jeder einfach so schreiben, das ist einfach wie so ein Readme im Endeffekt zum Modell, die wichtigsten Takeaways. Dann gibt es natürlich auch noch Library-Dokumentation und was vielleicht jetzt nicht so viele kennen, aber was ich tatsächlich auch ganz praktisch finde. Papers with Code, das listet für verfügbare Modelle alle möglichen Paper auf, in dem diese Modelle genutzt wurden, genau dasselbe auch für Datensätze und man hat dann für diese Sachen auch so kleine Leaderboards und hat teilweise so kleine Graphen, wo man dann sieht, okay, dieser Datensatz, der wurde jetzt über die letzten 15 Jahre gemined und die Accuracy-Kurve, die geht halt über die Jahre immer weiter hoch und man kann dann auf jeden Punkt klicken und sich angucken, in welchem Paper das war. Das finde ich auch ganz spannend. Hier jetzt als Beispiel für MobileNetV3 das zugehörige Papier, was auch Arxiv ist, die Model Card, so wie sie auf Hugging Face steht. Auf Papers with Code links sieht man dann halt, sieht man so ein bisschen angesetzt die Liste von Papieren, in denen das vorgestellt und genutzt wurde und auch wie viele Bewertungen das bekommen hat. Man hat direkt einen Link zum Code auch, daher der Name Papers with Code, also sind dann direkt auch GitHubs mit verlinkt. Ist natürlich alles Community-Driven, also das kann man da halt alles eintragen. Das heißt nicht, dass alle das da eintragen. Genau, dann Blogposts, links die Dokumentation wie in Keras. Wenn man sich das alles so in Gesamtkonstellation irgendwie anschaut, dann merkt man ja schon, dass da auch Vor- und Nachteile und Unterschiede und so in diesen ganzen Repräsentationsformen sind. Kannst du da vielleicht ein bisschen was zu sagen?

I8

Ja, erstmal, wenn man jetzt hier schon mal die Übersicht sieht und jetzt nochmal rechts das Label vergleicht und einmal links jetzt schauen in die ganzen Sachen, glaube ich, sind erstmal quasi, muss man sich erstmal in die Begrifflichkeit ja auch links erstmal einarbeiten, bevor man da auch aus einer Grafik jetzt mal eben irgendwie da den Inhalt rausziehen möchte, den man irgendwie gerade haben will. Also ist ja auch immer erstmal schön und gut aus einer Tabelle dann irgendwie darzustellen und dann ist ja noch eine einfache Form, da irgendwie mal reinzugucken. Das sieht halt alles relativ tiefgehend und komplex aus. Also da muss man sich halt einen Artikel durchlesen und man hat halt nicht diesen quasi einen Blick und ich gucke halt, habe da so eine grobe, schon mal eine Übersicht, was da ist. Das ist halt, glaube ich, fällt mir jetzt am meisten auf, halt rechts zu dem Label und links halt zu den ganzen Inhalten, die ich da zusammengestellt habe, dass man da schon sieht, da ist eine ziemlich große Komplexität drin und ich dafür auch erstmal ein ziemlich großes Verständnis da irgendwie haben müsste. Wenn ich jetzt zum Beispiel an meine Rolle denke oder so, würde das Label halt allein schon besser vielleicht sein, um dann halt irgendwie schon mal zu sagen, ja, versucht doch mal mit den drei Modellen, die haben wir hier vom Label, die sind halt irgendwie gut, die könnte man mal einsetzen. Links die Dokumentation ist ja dann gar nicht so interessant halt, das geht da dann auch vielleicht viel zu tief. Dafür muss ich dann schon mehrere Stunden investieren, um da überhaupt da irgendwie mal reinzukommen, ne.

A1

Ja, insbesondere, also das möchte ich auch nochmal klarstellen, wir sagen ja natürlich nicht, es alles Mist und wir müssen jetzt alles einfach nur noch labeln und dann sind die Probleme gelöst. Die Ingenieure und Entwickler, die wollen natürlich Informationen in der Tiefe haben und idealerweise steht hinter so einem Label auch ein detaillierter Report, wie dieses Label zustande gekommen ist und wie man das reproduzieren kann und all das. Das wäre halt total wünschenswert. Aber ja, genau, das ist das, was wir auch gemerkt haben, immer wenn wir mit Firmen zusammenarbeiten. Also für die ist halt, wenn man Machine Learning Results zeigt, ist das halt voll die Magie. Und wir hoffen halt, dass so Labels irgendwie da so ein bisschen die Schnittstelle irgendwie vereinfachen, um sich über Modelle auszutauschen, eben ohne, dass man dafür sagen muss, ja, du musst hier nur in dem Papier nachlesen, in der und der Tabelle, da steht genau drin, warum das besser ist für diesen Einsatzzweck so. Oder im Vergleich mit anderen Modellen hat sich das so und so verhalten, kann man hier ja in der Tabelle mal eben sehen, aber so mal eben ist es halt nicht. Man muss sich dann halt schon in Detail damit beschäftigen. Und da hoffen wir halt, dass wir so ein bisschen in dieser Nische irgendwie die Kommunikation, das Reporting und auch den Austausch über Machine Learning Ergebnisse irgendwie vereinfachen können. Genau. Kommen wir zum Ende des Gesprächs nochmal so ein bisschen auf prinzipiellere Fragen zurück. Nochmal zurück zum Thema Vertrauenswürdigkeit. Auch jetzt vielleicht gerade in diesem Vergleich der Modelle. Hast du da ein Gefühl, wie würdest du diesen Informationsquellen jetzt vertrauen? Welchen mehr, welchen weniger? Oder wie könnte, würde sich vielleicht auch dein Vertrauen oder das Vertrauen von irgendwelchen Mitarbeitern möglicherweise erhöhen lassen?

I8

Ja, da würde ich mal sagen, das ist immer sehr datensatzspezifisch. Also, wenn wir jetzt wirklich auf die Daten im Predictive Maintenance Umfeld gucken oder so, dann habe ich halt nicht so eine große Vergleichbarkeit und ich habe auch nicht diese großen Modelle. Deswegen muss ich mich da um die Erklärbarkeit selber kümmern. Da ist halt wieder die Frage, wie ist das im Image Umfeld? Das kann ich irgendwie schwer einschätzen, sonst würde ich immer sagen, dass wir die Erklärbarkeit aus dem Use case heraus eigentlich halt beschreiben müssen und evaluieren müssen, also wenn man jetzt in diesem Machine Learning Kontext guckt. Es ist halt die Frage bei ImageNet, also bei den Großen Geschichten geht’s dann halt wieder dahin wo dann halt wieder jemand anderes die Erklärbarkeit irgendwie auch darstellen muss. In dem Random Forest muss man sich selber drum kümmern oder wenn wir Controlling Daten oder so haben, dann geht es halt irgendwie, das Beispiel, dann wäre da wieder so das Thema, dass wir das quasi selber machen würden, aber irgendwann kommen ja auch größere Modelle, die dann mehr übernehmen schon und da wird dann halt wieder, da kann man halt nicht so die Erklärbarkeit dann irgendwie auch darstellen im Unternehmen selber. Also da ist so ein bisschen, man muss viel selber erklären, aber ein bisschen Erklärung muss halt auch kommen von den Dokumenten, Papern etc. da.

A1

Ja, tatsächlich ist das Thema Erklärbarkeit, das habe ich letztes Jahr auf der Konferenz so ein bisschen vorgestellt, wo ich XXXXXXXXX kennengelernt habe, dass Erklärbarkeit halt auch so ein sehr komplexer Begriff ist. So viele Leute sagen, dass Random Forests oder Decision Trees insbesondere, dass die Decision Trees inhärent interpretierbar und erklärbar sind, weil man natürlich bei jeder Entscheidung einfach dem Baum folgen kann und gucken kann, warum ist am Ende diese Entscheidung getroffen worden. Das ist aber natürlich auch nur so bedingt richtig. Wenn man einen Decision Tree der Tiefe 500 hat, dann würde ich nicht sagen, dass der inhärent erklärbar ist. Klar, wenn der eine Tiefe von 10 hat, dann kann man noch gut nachvollziehen, warum diese Entscheidung getroffen wurde. Aber wenn da an 500 verschiedenen Featuren irgendwie random gesplittet wird, dann ist da keine Erklärbarkeit mehr gegeben. Man kann natürlich jede Entscheidung konkret zurückführen auf, warum wurde jedes Feature an welcher Stelle gesplittet. Aber trotzdem ist das eine ganz schwierige Frage. Im Vision-Bereich genauso wie im NLP-Bereich wird deswegen ja sehr viel an Post-Talk-Explainability geforscht. Das heißt, man will für eine konkrete Modellentscheidung daraus wieder eine Erklärung generieren und zurückführen, zum Beispiel an welchen Featuren lag denn diese Entscheidung. Und das ist natürlich auch ein sehr schwieriger Bereich. Um diese Brücke halt herzustellen. Weil das Modell so komplex ist und da so viele tausend oder teilweise Millionen Operationen oder Rechenoperationen erfolgt sind, dass man... Man kann auch da rechnerisch... Also natürlich ist das deterministisch und nachvollziehbar, warum welche Layer welchen Wert ausspuckt und welcher Node welchen Wert ausspuckt oder welches Neuron welchen Wert ausspuckt. Aber das ist halt nicht menschlich interpretierbar. Also Erklärbarkeit ist da ein großes Thema. Ich finde es immer noch sehr schwierig, weil... Die Erklärbarkeit... Also man musste praktisch sagen, ab einer Modellgröße X ist dieses Modell nicht mehr inhärent erklärbar. Das kann man irgendwie an der Anzahl Parametern oder so versuchen abzuschätzen. Aber ob das Modell jetzt ein oder zehn Millionen Parameter hat, ist eigentlich egal. An der Stelle ist es nicht mehr inhärent erklärbar und man braucht halt Post-Talk-Explainability-Methoden. Ganz großes Problem da ist halt, dass man keine Ground-Truth hat. Man kriegt halt... Also es gibt keine Ground-Truth für Erklärungen. Ich erkläre dir ja auch Sachen anders als meinen Kollegen auf der Arbeit. Das heißt, auch Erklärbarkeit ist nicht... Da gibt es keine Ground-Truth, kein Label, was man erreichen möchte. Das ist die richtige und das ist die falsche Erklärung. Dadurch, dass man diese Daten gar nicht hat für Bilder... Also warum erkenne ich auf dem Bild ein Auto, dass der eine sagt, ja, liegt an den Reifen. Der andere sagt, ja, das ist so die grobe Form. Der andere sagt, ja, es ist auf einer Straße, es kann nur ein Auto sein. Dadurch, dass man diese Ground-Truth nicht hat, ist es auch sehr schwer, halt Modelle oderMethoden zu entwickeln, um solche Erklärungen auszuspucken. Also da sind viele Fragezeichen auch in der Forschung, was Erklärbarkeit angeht. Ja. Wie sieht das denn aus mit der Autorität, die praktisch das irgendwie am Ende verabschiedet oder bereitstellt? Hast du da Gedanken zu, dass das vielleicht auch ein Thema spielt bei der Vertrauenswürdigkeit?

I8

Wie meinst du das?

A1

Also es ist ja ein Unterschied, ob jetzt einer einen Blogpost schreibt über MobileNet und sagt, ja, ich habe das mal benutzt und das ist so und so. Oder ob das die Entwickler von MobileNet sind, die Informationen bereitstellen. Oder ob das irgendwie eine Third-Party-Instance ist, die das durchprofilt. In anderen Bereichen in der Industrie haben wir ja genau solche Unterschiede. Vertraut man dem Unternehmen, was eigentlich ein Produkt verkaufen will? Oder vertraut man einer unabhängigen Instanz? Ist eine unabhängige Instanz wirklich unabhängig?

I8

Aus der Arbeit heraus würde ich sagen, ist das gar nicht so wichtig, die Quelle. Es ist irgendwie so gefühlt irgendwie das, was gut funktioniert, was gut erklärt ist, zählt da mehr, als wer es veröffentlicht hat. Deswegen habe ich auch Medium so gesagt, weil das oftmals ein praktisches Beispiel gut erklärt ist und man da gut mit arbeiten kann. In so Papern ist es natürlich auch ganz spannend, da mal reinzugucken. Aber die auf den Use Case umzubiegen, die ich dann selber habe, da ist oftmals ein bisschen mehr Arbeit. Dokumentation von den Herstellern direkt ist immer der nächste Schritt. Ich ziehe mir das Beispiel und das Vorgehen aus Medium raus und gucke dann in die Dokumentation rein. Da komme ich auch nicht drum herum. Ich kenne jetzt auch nicht so viele Unternehmen oder Veröffentlicher, die dann ihre Modelle selber in ihre eigene Dokumentation mit anderen Modellen so präsent vergleichen. Das ist mir noch nie so aufgefallen, zumindest.

A1

Wenn dann wäre das im Paper, ja.

I8

Genau da wäre halt so der Einstieg wobei das halt nicht ein Einstieg in die eigene Entwicklung so ist gefühlt

A1

Ja, die Frage kommt daher, XXXXXXXXX hat letztes Jahr eine Studie dazu gemacht, zu so Trust Certificates, also einfach groß angelegte, quantitative Studie, wir haben 20.000, 30.000, 40.000 Leute befragt, wenn ein KI-Produkt ein Trust Certificate hat, ja oder nein, macht das das Produkt vertrauenswürdiger oder nicht und da kam halt auch daraus, im Zweifelsfall misstrauen die Leute KI in hypothetischen Produkten, wobei sich in der Praxis halt zeigt, wenn die Usability und die Benefits da sind, dann ist den Leuten das auf einmal völlig egal, wo das Modell herkommt oder wie das zertifiziert, ob das überhaupt zertifiziert wurde, wo die Daten herkommen, so wenn das einfach ein cooles Produkt ist, was irgendwie einem im Alltag Spaß oder Arbeit, also Spaß macht oder Arbeit abnimmt, dann ist das wichtiger. Gleichzeitig ist natürlich, also es kam halt auch teilweise so das Feedback, dass das wie so ein Zertifikat ist und die Leute vertrauen ja zum Beispiel dem Energy Label auch irgendwo und hinterfragen das nicht, aber es ist halt die Frage, würden sie das, wenn der Hersteller da drauf schreibt, dass das sehr effizient das Modell oder ob das halt eine Third-Party-Instance ist, und wir fragen uns so ein bisschen, wo müsste man das am besten ansiedeln, wer müsste solche Label ausstellen? Hersteller werden es nicht machen, weil die haben ja nichts davon, vor allem weil das halt die Weak-Spots von den Modellen aufzeigt, also müsste man eigentlich eine Third-Party-Instance haben und legt man dann sowas eher beim TÜV oder bei DIN oder ist das ein Regierungsunternehmen oder ist das eine NGO oder, da gibt es auch noch keine klare Antwort drauf, aber wir sind so ein bisschen interessiert an Feedback zu der Frage.

I8

Ich würde das so ein bisschen vergleichen eigentlich mit dem Kühlschrank-Label. Da hat man irgendwie so ein bisschen generelles Vertrauen. Da haben sie sich irgendwie selber ein bisschen verbaut mit dem Plus Plus und das irgendwie nicht dynamisch genug aufgebaut haben. Aber da gehe ich halt erstmal davon aus, ja, das ist ja von der zentralen Stelle, dann ist das auch gut. Wenn man das so sagt hier, man hat irgendwie so ein offenes Label, dann hat man irgendwie so wie bei Bioprodukten oder so, da ist ja immer jedes Label hat ja doch ein bisschen Skepsis da, dass das überhaupt gut ist. Also so eine zentrale Stelle oder wenn es zumindest das gleiche Format hat, schon mal und irgendwie schon mal öffentlich kommuniziert ist, also auch gar nicht, glaube ich, von der Zertifizierungsstelle, ich glaube, dann hat das schon mal so einen relativ offiziellen Charakter und man durch diese Darstellung sieht es dann halt irgendwie auch relativ schnell offiziell aus und man vertraut dem, glaube ich. Also das könnte ich mir bei dem Label doch schon auch so vorstellen, dass man da das Vertrauen auch drin hat.

A1

Okay, cool Sehr gut, ich glaube, wir sind tatsächlich dann noch mit den Fragen durch XXXXX, hast du noch irgendwelche Anmerkungen, irgendwas unklar, irgendwas, wo du jetzt auch aus deinem fachlichen Hintergrund sagst, da müssen wir nochmal drauf gucken?

A4

Nee, ich glaube, wir haben tatsächlich alles abgedeckt. Mir würde jetzt auch nichts mehr einfallen

A1

Dann, XXXX, solange du nicht sagst, nee, da ist mir gerade noch ein Gedanke gekommen das muss auf jeden Fall nicht raus, könnten wir die Aufnahme dann schon mal beenden

I8

gerade nicht vielleicht kommt noch was, aber gerade fällt mir nichts ein

A1

Okay, dann beenden wir auf jeden Fall schon mal die Aufnahme.