A3

Perfekt und wir fangen mit der ersten Frage an. Ja, das ist keine richtige Frage, aber stellen Sie sich, stell du dich doch mal kurz vor.

I3

Ich bin XXXXXX XXXXXXXX, ich bin Student an der XX XXXXXXXX, ich studiere Informatik in meinem dualen Studiengang, bin parallel, also noch Auszubildender als Fachinformatiker bei der XXXXXX XX und ja.

A3

Wie ist dein Werdegang der ist wahrscheinlich noch nicht so lang aber was ist so dein Hintergrund was hast du davor gemacht

I3

Ja, also ich studiere seit vier Semestern Informatik, davor ganz normale Schulinformatik, bisschen Hobbyinformatik, aber jetzt nicht irgendwie anders noch Minijobben oder so.

A3

und wahrscheinlich Abitur gemacht.

I3

Genau, Abitur. Achso darum geht’s, ja Abitur gemacht, ganz normal.

A3

Ok in welchem Unternehmen arbeiten Sie ist jetzt die Frage. Kannst du noch was zu XXXXXX sagen? das habe ich ja gesehen

I3

Genau. XXXXXX ist ja inzwischen ein sehr großes Software-Entwicklungshaus, das sich auf keinen bestimmten Bereich spezialisiert hat, sondern ganz weit in ganz vielen verschiedenen Branchen arbeitet. Ich bin auch im Bereich Cross-Industries momentan eingestellt. Also ich habe auch keinen Fokus auf irgendeine Branche, sondern ich arbeite da, wo gerade Arbeit gebraucht wird.

A3

Okay, Cross Industries, das ist quasi so ein Department in bei XXXXXX, welches wirklich von allen angesprochen wird. So kann ich mir das vorstellen? Okay, gut. Dann beschreib doch mal kurz so, was deine Rolle ist in dem Unternehmen.

I3

Genau, momentan bin ich ein Dualer-Student, das heißt natürlich keine große Führungsrolle oder so. Ich bin halt einfach ein Softwareentwickler, arbeite inzwischen, glaube ich, im vierten Projekt mit, hauptsächlich interne Projekte, eine Verwaltung für uns als Studenten, dann eine Verwaltung für die komplette Mietwagen-Geschichte und alles bei uns. Und jetzt gerade arbeite ich an einer Mitarbeiter-Datenbank, in der drinsteht, was die Mitarbeiter so machen und nutze dafür unter anderem auch KI.

A3

Okay, das wird mich auch schon zur nächsten Frage bringen. Also was sind erstmal so die Technologien, mit denen du arbeitest? Ich sage mal Frameworks, Programmiersprachen und vor allem natürlich das Wichtigste, welche Art von Berührungspunkten hast du mit KI und Machine Learning?

I3

Ja, momentan arbeite ich vor allem mit Spring Boot und Angular, also Frontend, Backend, Fullstack und für die KI-Nutzung momentan LangChain4j, weil das wir dann in Azure OpenAI einbinden, also letzten Endes in ChatGPT, Backend und dann arbeiten wir mit ChatGPT da drin. Sonst aber, wenn möglich, React habe ich auch mal theoretisch gelernt und Java ganz viel über die Uni, aber hauptsächlich eigentlich Spring Boot mit Kotlin.

A3

Okay, kannst du nochmal den Namen sagen von dem Machine Learning Modell oder was auch immer du als vor Azure genannt hast?

I3

LangChain4J ist eine Bibliothek, die die verschiedenen APIs zusammenfasst. Also das ist quasi eine API-Bibliothek, dass man nicht für jedes Modell eine eigene API einbinden muss.

A3

Okay, das kann ich tatsächlich noch gar nicht. Und dann 4J bedeutet für Java?

I3

Für Java, genau.

A3

Genau. Okay.

I3

Ist eigentlich eine JavaScript-Bibliothek, die dann irgendwann geportet wurde. Wobei es nicht genau, also es ist kein echter Port, sondern es ist eigentlich eine Weiterentwicklung. Ich glaube, da sind noch ein paar andere Sachen mit drin, aber benutze ich nur. Habe ich nicht dran entwickelt.

A3

Alles klar, okay, das macht dich zu einem guten Kandidaten für dieses Interview Was ist denn sonst noch deine Vorerfahrung mit der KI? Vielleicht auch, dazu würde auch gehören, ob du vielleicht einen Kurs in der Uni dazu gehabt hast

I3

Ja, ich habe schon vor, keine Ahnung, wie lange ist das her, inzwischen vier Jahre oder so, doch bevor ich Informatik studiert habe, mich ein bisschen mit PyTorch und allen beschäftigt, habe versucht, selber ein paar kleine KIs zu programmieren, so einfach dieses typische Hunde-Katzen-Unterscheiden und so, also nichts Großes, aber ein bisschen Einstieg an der Uni habe ich momentan kein Modul dazu. Allerdings schreibe ich meine Seminararbeit gerade über die Nutzung von KI zur Erstellung von grafischen Benutzeroberflächen, aber da halt nichts entwickelt, sondern auch nur, ich gucke, was es da gerade für Möglichkeiten gibt.

A3

Okay, alles klar. Wie arbeitest du genau mit dem KI? Kannst du vielleicht so einen, mit dem KI, wow, mit der KI, vielleicht kannst du da mal so einen Arbeitsablauf beschreiben, einen prototypischen, den du vielleicht öfter hast, wenn du so eine API ansprichst?

I3

Ja, also die API ist sehr schön sehr einfach, läuft letzten Endes so, ich kann, wie man das bei den normalen Chat-Clients auch kennt, eine System Message reingeben, die einfach mir im Kontext gibt davon, was jetzt die KI machen soll und dann kann ich eine User Message reingeben, auf die dann die KI dementsprechend reagiert, also momentan nutze ich die KI nicht im Chat-Kontext, sondern wirklich in einen, ich gebe dir einen Text rein und die KI soll dann auswerten, ob in diesem Text bestimmte Daten enthalten sind und ja, da muss ich halt wirklich einfach nur zwei Strings reingeben und die werden dann einfach direkt durchgepakt, also sehr angenehm sehr einfach.

A3

Was ist der Use Case? Du musst jetzt nicht irgendwelche Geschäftsgeheimnisse verraten natürlich. Oder was ist so ein Beispiel in deinen aktuellen Projekten, zum Beispiel im Fuhrparkmanagement?

I3

Ja, also für die KI oder insgesamt.

A3

Ne, für die KI. Jetzt schon eher…

I3

Ja, die ist nur für dieses Mitarbeiter-Management da. Die vergleicht verschiedene Projekte, ob da drin gleiche Technologien und so genutzt werden, weil das damals war ein uraltes Projekt wurd damals noch nicht so weit gedacht. Das heißt, anstelle, dass man die Sachen irgendwie wirklich als Klassen anlegt mit einer ordentlichen ID und allen, sind da wirklich einfach nur Strings drin gespeichert. Und dadurch jedes Mal, wenn man ein Projekt bestellt, werden die Technologien wieder neu hinzugefügt und dann schreibt man mal Java, mal Java mit der Version, mal Java mit der Version und einfach da eine Zusammenfassung von allen. Und ja, also hauptsächlich Auswertung, ob Strings zusammenpassen, nicht zusammenpassen, ungefähr das gleiche bedeuten.

A3

Okay, alles klar, interessant. Guter Einblick auf jeden Fall, vielen Dank. Was würdest du sagen, sind die größten Herausforderungen in der Benutzung, die dir aktuell begegnet sind?

I3

Ja, die Konsistenz der Ergebnisse, dass wenn ich die gleichen Sachen reingebe, ich auch bitte immer das Gleiche rauskriege. Und womit ich persönlich nicht direkt Probleme habe, da ich nicht so viel generiere, aber ein paar Kollegen, wo ich das mitkriege, die haben das Problem, dass die KI natürlich auch ganz oft irgendwas halluziniert, dass man da Daten rein gibt, wenn man sagt, ich stelle mir eine Beschreibung aus diesen Daten und denke sich noch irgendwas anderes mit dazu aus. Und dass man das halt verhindert.

A3

Okay, alles klar. Wie würdest du oder kannst du dir vorstellen, wie man diese Herausforderung in irgendeiner Form lösen könnte? Vor allem von der Seite der Forschung auch in KI-Modellen oder auch der Kommunikation von KI-Modellen, was für Infos noch gebraucht sind zum Beispiel.

I3

Ja, wenn ich die Lösung wüsste, dann hätten wir die Probleme nicht mehr. Ja, also für die Terministik sehe ich kaum eine Lösung, weil die KI, so wie sie nun mal arbeitet, zumindest nach meinem Verständnis, die zwar theoretisch deterministisch sein müsste, wenn man halt immer den gleichen Feedback reingibt, aber klappt nicht gerade bei den ganzen LLM-Modellen und so. Und für die, die es nicht halluzinieren, bräuchte man theoretisch nochmal zwei, drei KIs mehr, die nochmal drüber gucken, ob das zusammenpasst, aber habe ich auch keine Ahnung, inwiefern das sinnvoll ist.

A3

Okay, nochmal nachhakend, hast du manchmal während deiner Arbeit die Situation, dass du dich entscheiden musst zwischen den verschiedenen KI-Modellen?

I3

Nicht direkt. Also dadurch, dass ich halt nicht selber bezahle, muss ich das nehmen, was mir bezahlt wird. Und ich habe momentan nur die Auswahl über die OpenAI-Modelle und konnte mich dann entscheiden zwischen GPT-3,5 und GPT-4. Und am Ende hat dann das Geld entschieden, das wurde 3,5. Also theoretisch hätte ich zwar die Wahl, aber es ist jetzt nicht meine Auswahl direkt.

A3

okay ja da würde ich gerne mal darauf eingehen sagen wir mal Geld spielt in dem Fall keine Rolle was hättest du dir denn für Informationen von den beiden Modellen gewünscht die die ja bei der Entscheidung zwischen den Modellen helfen unabhängig vom Geld

I3

Ja, es gibt ja inzwischen relativ viele Benchmarks, die verschiedene Sachen testen, gerade was Wissen angeht bei den Dark Dynamics Models. Da wäre natürlich interessant gewesen, wie die da abschneiden und sonst einfach irgendwie eine Messung der Zuverlässigkeit, der allgemeinen Verständnis, inwiefern die KI das macht, was sie soll. Aber da weiß ich nicht, wie die Benchmarks genau laufen für…

A3

Okay, für welche Benchmarks fallen die denn ein für den ersten Fall?

I3

Die Namen fallen absolut nicht ein. Ich sehe nur immer die schönen Tabellen, wo dann auf einmal steht, hat 80% mehr oder so.

A3

Ja, okay, alles klar. Gut, das ist auch, jetzt habe ich endlich eine gute Überleitung zu dem Teil, wo ich dir auch mal was zeige. So, ich hoffe, man sieht meinen Browser. Ja. Das würde ich gerne groß machen hier. Habe ich das denn das letzte Mal geschafft? Ich glaube hier. Seht ihr das im Vollbild?

I3

Ja. Jetzt sehe ich es im Vollbild.

A3

Okay, perfekt. Und zwar haben wir uns eine Lösung eben zu diesem zugegebenermaßen auch fiktiven Problem von dir, weil du dich ja nicht so häufig entscheidest zwischen Modellen. Zu diesem Problem haben wir uns eine Lösung ausgedacht und zwar nennen wir das eben AI Labels. Und ohne dich jetzt noch weiter irgendwie zu beeinflussen, würde ich dich erstmal gerne fragen, was ist deine spontane Reaktion auf dieses Label? Was siehst du da? Was sind deine Eindrücke?

I3

Ich sehe viele, viele Daten. Jetzt bin ich versucht, die QR-Codes zu scannen, aber am Ende ist da noch irgendein Unsinn hinter. Ne, da sind wir. Ja, viele Informationen. Ich sehe ganz viele verschiedene Infos. Also natürlich erstmal dieses ABCDE-Ranking, was erstmal mir vermittelt, dass das eine sehr gute KI ist, dem Label nach. Unten die verschiedenen Bilder sagen natürlich auch sehr gut irgendwas mit Strom, Stromverbrauch, Powerdraw per Inference, Corrupted Robustness, kann ich mir jetzt nicht direkt was vorstellen darunter, aber auch gut die Accuracy. Und die Running Time, also wie schnell das antwortet. Dann habe ich noch scheinbar eine Angabe, welche GPUs dafür genutzt werden, also worauf die läuft, wenn ich das richtig deute. Ist das A100x8? Genau, und dann das Modell, das dahinter steht, beziehungsweise die Technologie, auf der die KI basiert.

A3

ja okay sonst noch irgendwie eindrücke wie du es findest vielleicht

I3

Also an sich sieht es aus, als würde ich gerade im Baumarkt oder im Mediamarkt nach der Waschmaschine gucken. Sehr verbraucherfreundlich, wenn es genau stimmt, was da draufsteht.

A3

Ja, das stimmt tatsächlich. Da kann ich gleich nochmal drauf eingehen, woher denn jetzt überhaupt dieses Label kommt oder wie wir zu diesen Ergebnissen gekommen sind. Ich sehe gerade, die Welt geht unter draußen, aber wir ignorieren das mal. Siehst du eine Brücke zu den Problemen, die wir vielleicht eben genannt haben?

I3

Ja. definitiv

A3

in welcher Form?

I3

ja ich habe die Top-1 Accuracy die Corrupted Robustness das würde ich… was ist das eigentlich? Kannst du das einmal erklären?

A3

Na klar. Ja, natürlich. Okay, ich kann gerne mal kurz über das AI-Label gehen. Zunächst einmal ist wichtig, diese Ergebnisse, also auch diese Farbcodierung kommt aus verschiedenen Experimenten, die unser Kollege XXXXXXX, der heute leider nicht dabei ist, gemacht hat. Der hat nämlich verschiedene KI-Architekturen, die sieht man hier. Also in dem Fall wäre es zum Beispiel das MobileNet V3 Small. Das ist eine bestimmte Architektur von einem neuronalen Netz. Er hat verschiedene davon eben geprüft und verschiedene Benchmarks eben erstellt, wie du das auch schon eben richtig gemeint hast. Es gibt ja immer viele Benchmarks zu verschiedenen KI-Modellen. Und über den Vergleich von verschiedenen Architekturen kommen wir eben zu diesen Einstufungen hier. Und zwar hat er sich hier in dem Fall Image-Klassifikatoren, das sagt dir wahrscheinlich auch was, der ImageNet. Hast du da schon Erfahrung mit gehabt?

I3

Ja, nicht direkt also ich habe Stable Diffusion am Laufen zu Hause, aber benutze ich nicht aktiv.

A3

Okay, alles klar. Genau, in dem Fall sind es halt eben Bildklassifikatoren, die miteinander verglichen wurden und wir haben uns speziell die Inferenz angesehen und ja, der Power Draw Per Inference hast du schon richtig identifiziert, also wo einmal Daten durch das neuronale Netz durchschieben, wird eben, werden eben so viele Milliwattsekunden an Strom verbraucht. Dann die Accuracy, das meintest du wahrscheinlich eben unter den Benchmarks, einfach die Genauigkeit des besten Ergebnisses des KI-Modells, also wie oft ist die Klassifikation letztendlich richtig. Bei ImageNet reden wir von tausend verschiedenen Klassen, die rauskommen können und kannst dir ja vorstellen, wenn man einfach so random eine Klasse picken würde als Ergebnis, dann hätten wir eine Accuracy von einem Tausendstel. So, ne, weil es dann einfach komplett random wäre. Von daher, so sind diese 63% eben zu lesen. Relativ hoch, aber es ist immer noch einer der schlechtesten Accuracy-Werte im Vergleich. Dann das, was du gerade schon berechtigterweise gefragt hast, die Robustness, das ist eine weitere Benchmark. Da geht es darum, wie robust ist die Entscheidung des KI-Modells gegenüber von Veränderungen des Bildes, was man ihm gibt. Das heißt, du hast irgendwie ein Bild von einem Hund und machst so und so viel Pixelveränderung in dem Bild und schaust dir an, ab wie viel Pixelveränderung ändert das KI-Modell eben seine Prediction. Also ab wie viel Pixelveränderung ändert das KI-Modell seine Prediction. Also ab…

I3

Also irgendwie zufällige Bildveränderung

A3

Genau, also ja. Beziehungsweise, also ja, ich glaube, in dem Fall hat er es mit zufällig gemacht. Man kann es natürlich auch bewusst attacken, das sagt ja vielleicht auch noch was, dass man wirklich bewusst Pixel so ändert, dass zum Beispiel eine Klassifikation von Hunde auf spezifisch Katzen irgendwie herbeigeführt wird. Das sind ja diese Classic Inference Attacks. Genau, und das ist halt eben eine Benchmark dafür, wie robust eben das Modell gegen Attacken ist. Und hier haben wir zuletzt noch die Running Time Per Inference, also auch wie lange das dauert letztendlich. Ja, hier hast du es schon richtig gesagt. Hardware und Framework, was genutzt wurde. Genau. Habe ich irgendwas vergessen? Ich hoffe noch nicht. Ja.

A4

Die QR-Codes vielleicht?

A3

Die QR-Codes. Ja. Die kann man scannen nh?

I3

Ja, habe ich gerade geguckt, irgendwelche eine Studie und ein GitHub glaube ich

A3

Genau, das ist halt die Studie, die letztendlich diese verschiedenen Modelle verglichen hat auch miteinander. Wichtig ist noch anzumerken, dass wir nicht nur vier Benchmarks haben. XXXXXXX hat noch viel mehr verwendet oder einige mehr. Und die Kombination aus allen kleinen Benchmarks führt eben zu diesem Gesamtscore. Okay, gut. Das soweit zur Erläuterung dieses KI-Modells. Jetzt wollen wir doch mal schauen, was wir auf der nächsten Folie haben. Ja, gut. Der Use Case ist hier quasi kurz dargestellt. Also wenn jemand, so wie du, APIs verwendet, dann ist es eventuell interessant, sich sowas vorher anzuschauen, bevor man sich entscheidet für ein gewisses Modell. So, kommen wir zum nächsten Label. Kannst du mal kurz zusammenfassen, was du hier siehst?

I3

Ja, hier sehe ich jetzt gerade zwei Label von verschiedenen Bilderkennungs-KIs. Einmal das EfficientNetB4 und das MobileNetV3 Small, was wir gerade schon hatten. Und jetzt kann ich relativ gut sehen, welche Unterschiede es gibt dazwischen.

A3

Okay, kannst du die Unterschiede interpretieren? Vielleicht einfach mal so sich entlanghangeln?

I3

Ja, also das linke, das neue Modell hat erstmal die schlechtere Bewertung mit dem C im Vergleich zum A, was auf den ersten Blick auffällt. Wenn man jetzt nicht an den Werten entlanghangelt, hat es einen deutlich schlechteren Stromverbrauch, also wirklich deutlich schlechteren Stromverbrauch. Eine bisschen bessere Corrupted Robustness, eine bessere Accuracy, aber eine deutlich schlechtere Antwortzeit. Und jetzt könnte ich halt gewichten, was mir wichtiger ist, dass ich diesen besseren Ergebniswert habe oder ob es mir wichtiger ist, dass ich weniger Strom, weniger Zeit verbrauche.

A3

wie ist da so deiner Einstellung zu? Jetzt nicht generell aber…

I3

Die Frage ist halt, was man braucht. Also wenn das jetzt irgendwelche spezifischen Sachen sind, wo ich keine Ahnung bei mir jetzt die Mitarbeiter danach in Projekten, Empfehele(?) einzuteilen, würde ich mich wahrscheinlich dafür entscheiden, das linke zu nehmen, weil einfach die bessere Accuracy heißt, ich kriege die besseren Ergebnisse. Wenn ich jetzt zu Hause irgendwie keine Ahnung, mein Raspberry Pi, Roboter, Gladys, was auch immer da aufbaue, dann reicht wahrscheinlich auch das Rechte, weil ich habe nicht den Strom, nicht die Zeit, Sachen zu machen, sondern geht es mir darum, schnell die Ergebnisse zu kriegen und wenn sie nicht so gut sind, ist das auch nicht so schlimm. Also da hängt das dann wirklich vom Use Case ab.

A3

Okay, der Use Case. Vielleicht nochmal, um auf XXXXXX wieder einzugehen, kannst du dir vorstellen, in ein paar Jahren bist du der Chef und hast vielleicht öfter mal die Entscheidung, welches KI-Modell letztendlich benutzt werden soll, zum Beispiel in deinem Fall welches Large Language Model letztendlich die Developer nutzen sollen. Kannst du in dem Kontext vielleicht nochmal so ein Szenario aufmachen, wie du dich entscheiden würdest mit den Informationen dieses Labels?

I3

Das wird natürlich heißen, ich muss einen Kosten-Nutzen mir aufbauen. Wie viel würde ich durch falsche Entscheidungen der KI verlieren? Wie viel würde ich Gewinn machen? Kann ich jetzt gerade wirklich nicht sagen, welches davon das besser geeignete wäre. Ich nutze halt gerade auch das 3.5. Ich habe Kollegen, die nutzen 4. Das ist eine Geschmackssache, glaube ich, erst mal.

A3

Alles klar, okay. Gut, jetzt habe ich hier, wie verstehst du die Informationen, die das Label präsentiert? Hast du da Detailfragen vielleicht auch zu noch, jetzt vor allem auch vor dem Hintergrund des Vergleichs?

I3

Ich muss sagen, das ist eigentlich alles sehr gut präsentiert. Also jetzt, nachdem du einmal noch die beiden Benchmarks erklärt hast, absolut kein Problem da irgendwas zu verstehen.

A3

ok alles klar ja gut dann wäre die nächste frage kannst du die Labels vergleichen und

I3

ja war ich ja gerade schon dabei also kann man sehr gut vergleichen

A3

Okay, dann vielleicht anders gefragt, was findest du besonders hilfreich? Also was fällt dir besonders positiv auf hier dran?

I3

Besonders positiv, das, weil, also ich finde es relativ cool, dass die Icons natürlich nicht nur den Wert darunter verändern, sondern auch die Farbe der Icons als solche angepasst wird, das heißt, ich habe nicht nur, ich muss gucken, ist die Zahl jetzt größer als die andere, 1,6 ist ja eigentlich ein Stückchen kleiner als 608, aber dadurch, dass die Batterie gelb ist, sehe ich, oder stimmt irgendwas nicht und sehe, oh, Milliwatt und Watt, also Kunden. Sonst, ja, dieses große ABCDE, sehr präsent, würde mir eigentlich suggerieren, dass ich das Rechte nehmen soll, das ist auf jeden Fall das Bessere. Inwiefern das jetzt am Ende hilfreich ist, weiß ich nicht, weil es kommt auf die Details an.

A3

okay was würdest du denn noch für details wünschen an zusätzlichen

I3

Ne. Die sind ja da gegeben. Also das Top-One -Accuracy und die Corrupted-Robustness sind halt auch da, aber mir geht es halt vor allem um die Ergebnisse und nicht um den Stromverbrauch und die Sekunden. Das ist halt immer die Gewichtung, die man dann da nimmt

A3

ok was findest du besonders schwer zu verstehen was vielleicht die kryptisch auffällt

I3

Wahrscheinlich würde es besser daran fallen, dass man diesen großen Buchstaben, der unterschiedlich gewichtet ist, als das, wie ich es gewichten würde, wahrscheinlich.

A3

Okay, klar. Also hast du das Gefühl, man ist ein bisschen darauf angewiesen, dass der Ersteller dieses Labels, also was der für Prioritäten hat?

I3

Genau, dass der einfach die Bias hat, ich finde das und das wichtiger als andere Werte und dann dementsprechend das Label vergibt, wobei da kann man dann halt auch gucken, was möchte das Label mir aussagen.

A3

ja okay hättest du da konkret vielleicht was sie gerade so einfällt wie man das Label auch anders gestalten kann um vielleicht auch dieses Problem zu lösen

I3

Ja, ich halte mal ein als Beispiel. Man könnte natürlich einfach mehrere von den Rankings direkt da oben hin machen, dass man drei Leisten hat, Stromeffizienz, Antwortrate und noch die ganzen anderen Sachen zusammenfasst nach, wie genau ist das Modell, dass man also direkt auf den ersten Blick zwei, drei verschiedene Faktoren hat und sich dann entscheiden kann, welchen davon nehme ich und nicht erst sieht, okay, ist gut oder schlecht insgesamt und dann direkt auf die kleinen Faktoren gehen muss.

A3

Ah, okay. Da würde ich gerne nochmal nachhaken. Wie gesagt, es gibt keinen richtig oder falsch, das ist wirklich nur dein Eindruck. Aber wie würde sich das dann verhalten zu den kleinen Labels, die hier schon drauf sind? Ja, du meintest, wir haben

I3

Ja, du meintest, ihr habt noch viel mehr Benchmarks dahinter, dass man quasi diese vier Labels wirklich als einzeln gewichtete Labels anstelle des einen Balkens macht, dass man da vier Balken hat und dann unten die ganz, ganz vielen, ganz vielen kleinen Labels hat, ist immer die Frage, inwiefern das für den Endnutzer von Vorteil ist, weil da haben wir ja ganz viele Daten, mit denen man arbeiten muss.

A3

Okay, ja, verstehe. Cool, vielen Dank. Also gute Anregung auf jeden Fall. Dann würde ich jetzt so einen Minisprung machen und dich fragen, welche andere Kommunikationsformen von KI-Modellen fallen dir denn so ein? Also letztendlich irgendwelche Displays von KI-Modellen, die du dir durchsehen kannst, zum Beispiel bevor du das KI-Modell benutzt.

I3

Ja, damit habe ich mich noch nie so wirklich beschäftigt. Also jetzt gerade bei meinem Projekt hatte ich einfach im Azure-Portal die Liste der verschiedenen KIs und dahinter die Preise stehen. Und dann kann man natürlich noch googeln nach LLM-Ranking und dann kriegt man relativ viele verschiedene Studien oder keine Ahnung, was für Werbungswebseiten, die sagen, nimmst du mein Programm bitte. Also da ist momentan relativ viel gegeben und bei keinem weiß man, ob man dem wirklich vertrauen kann und jeder sagt, seine KI ist aber in dem Benchmark besser als die anderen.

A3

Verstehe, okay. Es gibt einige, die wir herausgefunden haben. Jetzt muss ich mal kurz hier draufklicken. Und zwar haben wir diese sechs hier identifiziert. Die würde ich dir gerne mal ganz kurz vorstellen. Das sind einfach wirklich Kommunikationsformen von KI-Modellen. Zunächst einmal haben wir natürlich das klassische Publikation. Also viele KI-Modelle sind im Rahmen von irgendeiner wissenschaftlichen Arbeit entstanden. Und da kann man sich natürlich immer gut ein Paper durchlesen, wenn man da eben Lust drauf hat. In dem Fall ist es halt eben für dieses MobileNetV3, wo wir eben auch schon das Label für hatten, was von den Google-WissenschaftlerInnen erstellt wurde. Ja, also das wäre eine Form der Kommunikation. Das enthält auch Benchmarks zum Beispiel. Dann haben wir die ModelCard. In dem Fall ist es, glaube ich, die ModelCard von Google, wo auch eben verschiedene Informationen zu den Modellen dargestellt sind. Zum Beispiel ImageNet, was wir auch als Information drauf hatten. Oder auch, welche Arten von Inputs verwendet werden von dem Modell. Und auch Benchmarks sieht man hier unten. Auf verschiedenen. Hardware-Umgebungen. Oder, nee, in dem Fall. Verschiedenen Runtimes. Das ist hier der Unterschied. Okay. Dann haben wir hier Papers with Code. Ich weiß nicht, ob du das schon mal gehört hast.

I3

Habe ich noch nicht.

A3

Okay. Vielleicht bei der nächsten Seminararbeit oder so ist es ganz hilfreich. Und zwar sieht man hier oben zum Beispiel dann eben ein KI-Modell. In dem Fall auch wieder MobileNetV3. Und hier unten. Müsste man jetzt runterscrollen. Ist aufgelistet, welche Paper dieses Modell alle zitieren. Also wirklich eine intensive Auflistung von allen, die irgendwie MobileNetV3 entweder verwenden in der Arbeit oder irgendwie als Inspiration aufgreifen. Dann haben wir Blogposts. Also die werden wahrscheinlich medium.com oder towards data science vielleicht schon mal untergekommen sein. Auch beliebt. Wird von, ja. Meistens auch WissenschaftlerInnen und ja, auf jeden Fall Profis werden die geschrieben. Und enthalten auch öfter Fließt ext, aber auch Benchmarks. Dann haben wir natürlich die Library Documentations. Zum Beispiel hier in der Python Library bzw. TensorFlow. Und ja, dann haben wir noch Fact Sheets. Das ist von IBM. Das sind spezielle... Ja, Seiten, die gewisse Kategorien, sag ich mal, abgreifen, die man hier links sieht. Also zum Beispiel Bias. Wie fair ist dieses Modell? Was für Inputs, was für Outputs kriegt dieses Modell? Also sehr aufgeschlüsselt, sag ich mal. Das ist aber nur für IBM eigene KI-Modelle entwickelt. So, das haben wir rausgefunden. Jetzt würde ich dich fragen dazu. Welche Vor- und Nachteile siehst du? Hier so spontan. Was ist dir aufgefallen? Gerade im Vergleich zu dem KI-Label, was wir dir gezeigt haben.

I3

Bei allen muss man sehr, sehr viel lesen. Also alles ist irgendwie deutlich mehr Text, deutlich mehr Daten und das Label präsentiert das natürlich deutlich praktischer, deutlich handlicher. Klar, bei manchen hat man dann eine Tabelle immer mit drin, wo was drin steht, aber nichts von denen ist so einfach zu verstehen wie das Label.

A3

das ist natürlich sehr positiv du kannst auch gerne sehr kritisch sein siehst du nicht Nachteile vielleicht irgendwas was das KI-Label nicht liefert im Vergleich zu den anderen Darstellungsformen

I3

Ja, gerade das Paper und die Papers with Code und alles bieten natürlich deutlich mehr Informationen an als das Label. Also wenn ich so ein Paper lese über eine KI, kann ich die wahrscheinlich deutlich besser verstehen, als wenn ich mir nur das Label angucke. Und auch so ein Blogpost, wo immer die Gefahr ist, dass der Autor noch seine Meinung da irgendwie mit reinpackt, wird auch deutlich mehr Informationen vermitteln als das Label, wenn es ordentlich geschrieben ist. Ja.

A3

war. Was für Informationen würdest du dir aus so einem Blogpost im Spezifischen nochmal dazu erhoffen? Also was schätzt du an dieser persönlichen Meinung?

I3

Da kann man dann nochmal gucken, wofür diese KI überhaupt gebaut ist. Also klar, Image-Classification ist schön, aber mit welchen Sachen hatten Sie jetzt da im Kopf, ob Sie was für Ihr selbstfahrendes Auto gebaut haben oder für irgendwas anderes, wird man da mehr Informationen rauskriegen und dann kann man gucken, okay, stimmt das mit meinem Use-Case überein oder möchte ich was anderes damit machen.

A3

ok nochmal so du bist ja auch Developer wenn du wirklich sagen wir mal irgendwann mal die Situation kommt dich zwischen zwei KI Modellen zum Beispiel für eine Image classification kann ja sein zu entscheiden was siehst du in Bezug auf das Programmieren als Developer für Vor und Nachteile? Fällt dir da vielleicht noch was zusätzlich ein?

I3

Ja, da muss ich mir auf jeden Fall die Dokumentation auch angucken, damit ich auch sehe, wie kann ich das einbinden, kann ich das überhaupt verwenden mit dem, was ich mache. Ja, das Paper wird man sich wahrscheinlich auch nochmal drüber gucken müssen, auf jeden Fall. Da wäre wahrscheinlich auch gerade das von IBM sehr praktisch, ist natürlich nur für deren Modelle, aber nachdem, was ich da gerade gesehen habe, waren da einige sehr interessante Sachen mit dabei, die ich auf jeden Fall brauche. Also ja, eigentlich hat alles durchaus die Berechtigung da zu sein. Jetzt ist halt nur die Frage, wo möchte man drauf schauen?

A3

alles klar okay dann wenn du alle deine Eindrücke dazu geteilt hast würde ich zur frage übergehen dass kleiner Sprung wer sollte solche Label ausstellen deiner Meinung nach?

I3

Das ist eine sehr gute Frage. Auf keinen Fall der Publisher selbst. Im Idealfall wird man das machen, wie bei den meisten Labels, die man in Deutschland kriegt, dass man irgendein unabhängiges Institut dafür hat. Wobei da auch immer die Frage ist, wie unabhängig sind die natürlich Stiftung Warentest haben wir hier eigentlich schon ein sehr eingebürgertes Unternehmen. Ob man sowas verstaatlichen sollte, ist die andere Frage. Bei der momentanen Einstellung der Regierung, also nicht der Regierung, aber der politischen Landschaft in Bezug auf IT. Theoretisch könnte man dafür bestimmt auch irgendeine Standardkommission gründen, die das entscheidet. Aber dafür müssten sich ja die ganzen Unternehmen einig werden. Also es muss auf jeden Fall irgendwas Zentrales sein und nicht jeder darf sein eigenes Label sich ausdenken.

A3

kannst du noch mal drauf eingehen, was genau für ein Vorteil dieses Zentralität hätte in Bezug auf die Transparenz vielleicht

I3

Ja, alles wird nach gleichen Arten bewertet und nicht, dass halt der eine sagt, ich setze den Fokus darauf, dass ich möglichst schnell die Antworten gebe. Der andere sagt, ich setze den Fokus darauf, dass ich möglichst wenig Strom verbrauche. Und dann könnte jeder das so anpassen, dass jeder überall die große grüne Zahl da hat, weil er halt die ganzen Sachen so anpasst, dass die Sachen, die wichtig sind für ihn oben stehen und die schlecht laufen, halt irgendwie so klein untermultipliziert werden, dass man sie nicht merkt.

A3

könnte das nicht auch kritisch gefragt mit einer mit einem zentralen mit einer zentralen vergabe stellt zum Beispiel passieren dass man wie beim Nutri-score zum Beispiel ja auch dass man das quasi trotzdem aus

I3

Das ist natürlich möglich. Deswegen müsste man halt dafür sorgen, dass diese Label so aufgestellt sind, dass, wenn der Wert gut ist, auch auf jeden Fall alles gut ist und dass man nicht irgendwie die Sachen umgehen kann, indem man manche Sachen anpasst. Bei Nutri-Score hat man ja insgesamt sieben Gruppen. Das heißt, bei der KI müsste man auch aufpassen, dass man halt nicht Äpfel mit Birnen vergleicht.

A3

ja okay absolut okay wie findest du denn allgemein die vielleicht auch die Pflicht zur Zertifizierung von KI-Modellen

I3

das ist relativ schwierig weil ich kann nicht als irgendein Developer der sich da sein fein getuntes Stable Diffusion auf Hugging Face hoch lädt mit keine Ahnung was für Zertifikaten rumschlagen das ist ein bisschen aufwendig aber natürlich sobald man anfängt die Sachen anzubieten wäre schön wenn man das machen kann wobei ich da eher denke das Label muss ich durchsetzen dass es anerkannt wird als die Leute müssen verpflichtet werden das Label zu nehmen da kann man halt schön damit Werbung machen aber wir haben hier das Label A und die anderen das Label C

A3

Ah, okay.

I3

Testsieger bei Stiftung Warentest und so, sind ja genau die gleichen Werbeversprechen.

A3

Also freiwilliges Selbstkontrolle, heißt das ja.

I3

Genau, und wenn man es nicht hat, dann haben die wohl was zu verbergen.

A3

Okay, ja, interessant. Also du fändest es hilfreich, jetzt auch auf deinen Job vielleicht wieder bezogen, ich weiß, das ist ein bisschen schwierig, weil du aktuell ja noch nicht diese Entscheidung triffst. Generell, fändest du es hilfreich?

I3

Ja, auf jeden Fall.

A3

Okay, gut. Müsste denn das Label in irgendeiner Form noch angepasst werden, zum Beispiel an deinen gewissen Wissenshintergrund? Vor dem Hintergrund habe ich das noch nicht gefragt. Ja, sollte es verschiedene Formen von Labeln vielleicht geben?

I3

gibt. Also das Label muss in der Form auf jeden Fall angepasst werden, da fehlt oben beim QR-Code die Hälfte, aber nein, theoretisch müsste man nochmal unten eine Tabelle haben oder auf der Rückseite eine Tabelle haben, wo dann wirklich die ganzen feinen Ergebnisse nochmal aufgelistet sind, dass man dann quasi auch direkt über diese zentrale Prüfstelle sich nochmal alle zentralen Werte angucken kann und dann nicht erst ins Paper gehen muss dafür. Aber ansonsten nach dem, wonach ich mich entscheiden würde, ist da erstmal alles gegeben momentan.

A3

Okay, also so eine Art Opt-in, dass man einfach mehr Informationen bekommt. Kannst du dir vorstellen, dass die andere Richtung, dass man weniger Informationen erstmal haben möchte, kannst du dir das vorstellen, vielleicht bei deinen Kollegen oder so, dass es auch gewünscht ist? Oder vielleicht bei Leuten, die gar keine Ahnung von Informatik haben?

I3

Ich meine, einfaches Ranking-System, wirklich nur A bis E, ist natürlich deutlich schöner und einfacher. Und die werden wahrscheinlich auch nicht wirklich was damit anfangen können, was jetzt diese Accuracy, die Robustness und der Stromverbrauch bedeutet, ist jetzt 600 Milliwatt Sekunden viel, ist das wenig. Da braucht man das wahrscheinlich nicht. Und auch auf welchen Grafikkarten das jetzt läuft, ist dem wahrscheinlich ziemlich egal. Aber gerade wenn man in Richtung Entwickler geht, ist es doch relativ interessant.

A3

wunderbar also ich werde durch mit dem Interview Leitfaden tatsächlich. XXXXX habe ich irgendwas vergessen?

A4

Ich glaube tatsächlich nicht.

A3

Ich glaube auch, wir sind relativ smooth durchgewieselt.

SPEAKER\_00

Das Einzige, was mir noch einfallen würde, da habe ich leider vergessen, an der richtigen Stelle zu fragen, du hast ja deinen Fokus quasi sehr auf diese Top-One Accuracy gelegt. Wir haben zum Beispiel auch die Top-Five getestet. Würde dir das, also wird das deine Entscheidung verändern, quasi, wenn du das in Relation zum Power-Draw siehst?

I3

Also da habe ich jetzt nicht direkt genug Ahnung von Image Classification, inwiefern das sinnvoll wäre. Ich brauche einfach irgendeinen Wert, der mir sagt, wie oft trifft die KI die richtige Entscheidung, die wir wollen. Ob das jetzt dieser 1% ist oder 5%, da müsste ich dann nochmal vorher gucken, was das überhaupt bedeutet

A3

Also die Top-5-Accuracy bedeutet eigentlich nur, dass die besten fünf Ergebnisse oder die besten fünf Ergebnisse, die die KI am wahrscheinlichsten findet, ob darunter unter diesen fünf das richtige Ergebnis, also die richtige Klasse drin ist. Das heißt, die Top-5-Accuracy ist immer höher als die Top-1-Accuracy, weil die Top-1-Accuracy ist wirklich nur, ob das beste Ergebnis auch das wahre Ergebnis ist.

I3

Okay, das macht Sinn. Da ist auch wieder Use Case und alles, aber dann ist doch die Top 1 Accuuracy deutlich besser geeignet, gerade da das dann wirklich der goldene Wert ist, den man haben möchte.

A3

Ja. Okay. Ja, vielen Dank. Wunderbar.

A4

Dann würde ich die Aufnahme stoppen.

A3

ja ich glaube das können wir laut Leitfaden