A1

Mal auf, also... Genau, perfekt. Und dann würde ich dich... Also genau, vielleicht noch kurz vorab, es gibt bei dieser ganzen Studie keine falschen und richtigen Antworten. Auch wenn das vielleicht teilweise so ein bisschen klingt wie eine Prüfungsfrage, sind wir halt einfach an einem offenen Gespräch interessiert. Und insofern einfach frei raus erzählen, was einem so durch den Kopf geht, ist das Beste, was irgendwie gemacht werden kann für uns. Genau, dann würde ich dich bitten, dich einmal kurz vorzustellen.

I10

Ja, ich mache es mal ein bisschen länger, dann kann man vielleicht nochmal die ersten paar Fragen beantworten. Ja, XXXXX XXXXX XXXXXX. Beruflich beschäftige ich mich, oder beruflich bin ich in zwei Bereichen. Was heißt in zwei Bereichen? Also beruflich mache ich zwei Sachen, nämlich zum einen mache ich Softwareentwicklung im Consulting. Das ist so der Job, der mir mein Essen finanziert und meine Wohnung. Und dann gibt es noch den Job, den ich eigentlich machen will, aber der sich nicht selber finanziert. Und zwar entwickle ich einen interaktiven, computergestützten Vokabeltrainer für Gebärden. Oder um es etwas einfacher und weniger kryptisch auszudrücken, dass auch Nicht-Informatiker verstehen. Ich versuche ein Computerprogramm zu entwickeln, mit dem man Menschen, die hören und sprechen können, die Sprache beibringen kann von den Leuten, die das leider nicht können. Das war ursprünglich mein XXXXXXXXXXXXXXXXXXXXX. Also vor 13 Jahren habe ich da mal ein Konzept entwickelt im Rahmen von XXXXXX XXXXXXX, wo es halt darum ging, dass man, wenn man Gebärden lernt, wirklich auch ein Feedback dazu bekommt, ist das richtig, ist das falsch. Und im Rahmen dieser XXXXXXXXXXXXXXXXXXXX hat man sich halt mit der Thematik beschäftigt. Da wir aus einem ganz anderen Bereich hierhingekommen sind, und eigentlich nur Gebärden erkennen so als Beispielanwendung für was anderes haben wollte und dann festgestellt hat, okay, Gebärdenerkennung oder Gebärdenübersetzung funktionieren noch gar nicht, hat man es dann halt runtergebrochen im Rahmen von XXXXXXXXXXXXX, um halt diesen Anwendungsfall zu machen auf Einzelgebärdenerkennung, weil das damals wie heute der technische Stand ist und war. Und um dann einen Anwendungsfall zu machen, haben wir halt gesagt, okay, brauchen wir Vokabeltrainer. Und haben uns damals dann weiter mit der Thematik beschäftigen festgestellt, okay, den gibt es ja auch nicht. Und so ist dieses Projekt entstanden. Das habe ich jetzt 13 Jahre entwickelt und im Mai auf der XXXXXXXXX davon einen Prototypen mal vorgestellt. Ich nehme gerade im XXXXXXXXXX XXXX teil. Und über diese Arbeit daran bin ich halt auch in die XXXXXXXXXXX gekommen, diese XXXXXXXXXXXXXXXXX XXX XXXXXXXX. Und darüber bin ich dann nach hier gekommen.

A1

Genau, sehr cool. Auf jeden Fall ein super spannendes Projekt auch. Also wieder so ein Projekt, wo ich mich dann irgendwie sehr begeistern dafür kann, was man auch mit Informatik irgendwie machen kann und was für neue Technologien das halt irgendwie ermöglicht. Weil sowas, glaube ich, super wichtig ist und natürlich irgendwie im Kapitalismus ein bisschen unter die Räder gerät, weil es halt irgendwie nur wenig Personen betrifft, sage ich mal, und sich damit halt nicht viel Geld verdienen lässt.

I10

Ich würde eigentlich, naja, man müsste eigentlich einen Paradigmenwechsel machen. Also im Kapitalismus lässt sich damit kein Geld verdienen, ist halt so ja für die Person, die das Programm macht, wahrscheinlich richtig. Also zumindest nicht viel. Weil grundlegend hat man ja im Kapitalismus verschiedene Player, zum Beispiel auch den Staat. Und wenn wir uns überlegen, wir haben in Deutschland 180.000 Gehörlose und Schwerhörige, die auf Gebärdensprache angewiesen sind. Wenn keiner von uns Gebärdensprache kann, sind das alles Menschen, die nichts tun können. Die irgendwo rumsitzen und die Decke anstarren. Und das kostet uns gesellschaftlich ja schon eine Menge.

A1

Das stimmt, ja, absolut.

I10

In Bangladesch sind wir sogar bei 1,2 Millionen von 160 Millionen. Also die lassen 1,2 Millionen menschliches Kapital in der Ecke verschimmeln, um es mal ganz böse auszudrücken. Also dementsprechend, ich würde es gar nicht auf Kapitalismus, ich würde es eher auf Gesellschaft machen, dass wir als Gesellschaft oder politisch einfach keinen Willen, obwohl das sieht in Bangladesch sogar besser aus als in Deutschland, keinen Willen haben. Also...

A1

Ja, aber genau, also auf jeden Fall ist da halt menschliches Potenzial vorhanden und man kann halt mit Technologie da versuchen, Brücken zu schlagen und das ist ein sehr schöner Faktor im Bereich Informatik. Deswegen finde ich es ein sehr cooles Projekt. Genau, über die XXXXXXX XXXX, wir hatten bei uns über das XXX den Kontakt dahin auch gehabt und so und die haben das dann in ihrem Verteiler wieder geteilt und so. Es ist für uns ganz spannend, dass wir so die Interviewpartner von überall her kriegen. Also wir sind hier alle irgendwie angesiedelt an der XX XXXXXXXX, beziehungsweise in meinem Fall und in XXXXXX Fall am XXXXXXXXXXXXXX XXX XXXXXXXXXXXX XXXXXX XXX XXXXXXXXXX XXXXXXXXXXX. Das ist in erster Linie ein buntländergefördertes Forschungsprojekt, das vielleicht noch kurz, das habe ich vorhin bei der Vorstellung so ein bisschen unter den Tisch fallen lassen. Insofern Grundlagenforschung ist unser Daily Business, aber natürlich machen wir auch Industrietransfer, XXXXXXXXXXXX XXX XXX XXXXXXXXXXXXXXXXXXXXXX XXX XXXXXX XXX XXXXXX XXXXXXX XXXXX XXX XXXX XXXXX XXXXXXXX XXX XXXXXXXX XXXXX XXXXXXXX XXX XXX XXXXXXXXXX, aber jetzt über die XXXXXXX XXXX nochmal ein paar Leute, zu denen wir bisher den Draht nicht hatten. Das ist ganz cool. Genau, damit wären wir jetzt aber auch schon beim Thema. Also wir haben halt irgendwie ein Forschungsinstitut für maschinelles Lernen und künstliche Intelligenz. Inwiefern spielt das denn in deinen alltäglichen Prozessen und in deiner Arbeit als Softwareentwickler und auch im Rahmen dieser, ich weiß jetzt nicht, wie ich es genau formuliere, aber im Rahmen der Entwicklung von diesem Prototypen für Gebärdensprache und ein interaktives System für Gebärdensprache, inwiefern spielt KI und Machine Learning da eine Rolle?

I10

Ja, eine sehr große, weil es eigentlich einer der wichtigsten Elemente in diesem Projekt einfach ist. Ich hole mal was aus, damit es vielleicht besser zu verstehen ist. Gebärden lernt man eigentlich heute in zwei Szenarien nämlich. Das eine ist, man hat jemanden um sich herum, der Gebärdensprache kann. Das bedeutet, man sagt dem, mach mal Bus. Und dann stellt er sich hin und macht einen Bus. Und dann sitzt der Gehörlose da und sagt, ja, das ist richtig. Die zweite Möglichkeit, das ist die, die vor allen Dingen verwendet wird und die hochgradig gefährlich ist, dass Video-CDs verteilt werden und die Leute vor ihrem Spiegel stehen. Das bedeutet, die Leute stehen vor ihrem Spiegel, gucken sich das Video an und machen das nach. Das ist, für viele Sachen oder für einige Gebärden wahrscheinlich auch ganz praktikabel. Es gibt allerdings Gebärden, die sind sehr ähnlich und haben eine völlig unterschiedliche Bedeutung. Da das hier wie ein Video aufgezeichnet werden will, will ich das jetzt einmal nicht nachmachen. Aber wenn man zum Beispiel beim Fußball, das ist die Gebärde für Fußball, die Hand 90 Grad dreht, fragt man was sehr Obszönes. Und ich bin auch mal zum Gehörlosen hingegangen und gefragt, ob wir abends Fußball gucken wollen und habe ihn latent was anderes gefragt. Gut, er kannte mich. Er wusste, was ich von ihm wollte. Er hat das dann einmal kurz mir gezeigt, wie es richtig geht. War aber jetzt schon hochgradig peinlich und hätte mir in einem anderen Kontext auf jeden Fall nicht passieren müssen. Und was wir in dem Projekt drin haben, ist halt, dass wir eine KI da hinzufügen, die diese Kontrolle der Gebärden wirklich macht. Ist das, was er gebärdet, wirklich die Gebärde, die es sein soll? Und dementsprechend eine sehr große.

A1

Nur um das einmal jetzt irgendwie schon mal auszuholen, also wir werden uns jetzt in dem Gespräch wahrscheinlich hauptsächlich um diese Gebärdenspracherkennung, Übersetzung, KI-Prototypen unterhalten im Rahmen der Softwareentwicklung, also deiner Haupttätigkeit. Ist KI und Machine Learning eher im Hintergrund oder gar nicht präsent oder...

I10

Naja, es gibt Projekte, wo da durchaus Kontaktpunkte da sind. Allerdings sind das dann A deutlich kleinere und weniger aufwendigere. Da geht es dann so um Prognostiken oder dass man mal, da ich sie gar nicht selber schreibe, sondern vor allen Dingen das Labeling vorher, also diese Datenerzeugung vorher übernehme. Also das ist schon das Projekt, wo ich am meisten mit KI zu tun habe.

A1

Okay, aber auch im Rahmen der Softwareentwicklung, dann können wir das auch ruhig mit drin lassen. Ich wollte nur einmal verstehen, ob das wirklich so ein Motor nur in diesem Prototypen ist sozusagen oder ob das auch bei der, also ich meine, wir sind jetzt an dem Punkt, dass irgendwie jeder zweite Entwickler auch eh GitHub Copilot benutzt und dann hat man da natürlich auch wieder KI in der Softwareentwicklung, aber man arbeitet halt nicht selber.

I10

Das habe ich jetzt gar nicht als KI-Entwicklung definiert. Also ich finde, da sollte man einen Unterschied machen. Also KI-Entwicklung ist für mich, ich weiß, da gibt es viele Leute, die vor allen Dingen aus marketingtechnischen Gründen was andere machen. Also für mich ist KI-Entwicklung wirklich, ich entwickle ein Modell und ich trainiere da was rum. Und alles andere ist für mich auch nicht, ich setze künstliche oder ich entwickle KI. Also ich setze es vielleicht ein, wenn ich ChatGPT nutze, aber alles andere hier so. Ja, wir machen auch ganz viel mit künstlicher Intelligenz und eigentlich, wenn man nur vortrainierte Modelle verwendet, ist jetzt für mich Anwendung und nicht Entwicklung.

A1

Ja, auf jeden Fall. Also die Unterschiede muss man auf jeden Fall klar haben. Das ist auch was, was wir jetzt bei der Interviewstudie halt gemerkt haben. Wie gesagt, wir als Machine Learner natürlich trainieren wir auch eigene Modelle und so. Ich habe mich aber auch viel damit beschäftigt, praktisch vortrainierte Modelle zu evaluieren, weil auch das kann man ja forschungstechnisch machen. Weil wir sind halt irgendwie jetzt an der Zeit, wo es sehr, sehr komplexe Modelle gibt, die man auch nicht lokal trainieren kann. Also so Sprachmodelle oder sowas. Nicht jede Firma hat irgendwie überhaupt die Daten oder die Infrastruktur, um das mal eben zu machen. Das heißt, gerade transformiert sich dieser Markt so ein bisschen dahin, dass intelligente Systeme auch im großen Stile eingekauft werden, um intern Business Opportunities irgendwie zu lüften. Was auch okay ist. Und auch die haben ja eine Meinung zu KI und auch die haben eine Schnittstelle oder haben eine Form von Interaktion mit KI. Also wollen wir natürlich auch mit denen sprechen. Aber ja, da muss man ganz klar unterscheiden. Also auf jeden Fall cool, dass du auch selber KIs entwickelst. Das ist gut. Genau. Was sind denn in diesen alltäglichen Prozessen so die größten Hürden? Also wo siehst du die größten Probleme? Was kostet dich am meisten Zeit in deiner Arbeit mit KI oder Entwicklung von KI?

I10

Das große Problem an dem Projekt ist, dass man nicht wirklich ein Datenset hat. Also was ich entwickelt habe, ist eine kontrolliert gelehrte KI und dafür braucht man natürlich ein Datenset. Und das größte Problem ist einfach, dass es dieses Datenset nicht gibt und Gebärdensprache ist nichtmal international. Das bedeutet, wir haben auch noch lokal unterschiedliche, also es gibt so 400 Sprachen auf der Welt und so circa 200 Gebärdensprache. Davon sind vielleicht auch nur 75, 80 relevant, weil jetzt wenn man die Flämisch-Belgische nennt, da sprechen 5000 Menschen. Also ich glaube, das kann man wirklich mal ignorieren. Es gibt einfach kein Datenset. Und dieses Datenset ist natürlich eine Grundlage für das, was wir machen. Es ist aber auch eine Grundlage für Grundlagenforschung, die du, XXXXXXX, vielleicht machst, im Bereich Gebärdenspracherkennung. Wie willst du Gebärdenspracherkennung, Grundlagenforschung machen oder irgendjemand anders auf der Welt, wenn es kein Datenset gibt? Und das ist so ein bisschen die Diskussion, oder das, was eigentlich hier seit Jahren die meiste Zeit frisst, dass man versucht, halt Leute zu überzeugen, dass so ein Datenset einfach die Grundlage auch sein kann für eine technisch gestützte Inklusion dieser Menschen. Und das ist halt einfach mit Menschen zu tun haben, die A das Wort Vokabeltrainer schon überfordert, die mir dann die ganze Zeit was von einem Übersetzungstool schreiben. Und dann aufgrund, wenn man das schon als Basis hat, überhaupt nicht verstehen oder in der Lage sind zu verstehen, warum so ein Datenset so wichtig wäre. Und der zweite Punkt ist halt einfach, dass vor allem dieser Gebärdenerkennung, es gibt eine Menge oder ganz viele so Forschungsarbeiten im Rahmen von Bachelor- und Masterarbeiten. Aber es gibt ganz wenige wirkliche Dokumentationen darüber, woran man sich mal irgendwie anknüpfen kann. Das bedeutet, man ist hier gefühlt schon auf so einer Grundlagenforschungsebene ganz häufig unterwegs, weil es halt einfach nicht so richtig viele Menschen gibt, über die sich damit beschäftigen. Und mit dem Datenset, die es gibt... Ich meine wenn wir das ABC nehmen, dann haben wir hier zum Beispiel in A, dann haben wir hier ein E und dann haben wir hier ein I. Das sind für uns klar erkennbare, wenn ich das jetzt so zeige, Unterschiede, ob das jetzt so oder so ist. Aber andererseits sind es nur zwei Finger, die vom A ausgehend mal verwendet werden. Das ist jetzt für eine künstliche Intelligenz ja natürlich nicht das Einfachste. Und diese ganzen Beispielprojekte, die gemacht sind, sind vor allen Dingen halt mit so 25 Beispielgebärden, die dann doch sehr unterschiedlich sind. Mal die rechte Hand, mal beide Hände, mal oben, mal unten. Das macht es natürlich alles… Ja das bedeutet, man hat nicht wirklich eine Grundlage, auf der man mal aufbauen kann.

A1

Leuchtet total ein. Ist, glaube ich, immer das Problem, wenn man in einem Bereich irgendwie versucht, KI und Machine Learning einzusetzen, wo es auch einfach so ein bisschen sich um eine marginalisierte Gruppe handelt, weil da halt automatisch dann statistisch weniger Daten generell vorhanden sind. Also geht jetzt natürlich ein bisschen weg von unserem Gespräch, also von unserem eigentlichen Kerngespräch, aber trotzdem interessiert mich das, gibt es nicht, also ganz oft jetzt so, ich denke da gerade jetzt an Olympia und irgendwie in ganz vielen Fernsehübertragungen und so Live-Übersetzer, die das in Gebärdensprache auch gestikulieren, kann man da nicht einen Link herstellen und da auch Daten abgreifen? Oder, klar, da ist dann auch das Problem, da gibt es dann wahrscheinlich so deutschlandweit irgendwie fünf Leute, die das irgendwie übersetzen live und das heißt, man hat halt nur sehr wenig verschiedene Menschen, die Gebärdensprache in dem Moment sprechen und wo man dann Übersetzung hat, aber kommt man an solche Daten dran?

I10

Ich würde mal ein anderes Beispiel reinbringen, vor allem im deutschen Bereich, wäre XXXXXXX XXXXXXXXXXX XXXXXXX XXXXXXXXXX wird meistens von XXXXXXXX übersetzt. Unfassbar nette Leute. Also falls ihr mal Gebärdensprachen-Dolmetscher braucht, XXXXXXXX XXX XXXX, kann ich sehr empfehlen. Auch im Umgang sehr, sehr locker und so. Richtig nett. Das sind glaube ich so zwölf. Also man hätte da schon eine gewisse Auswahl an Menschen. Das Problem ist halt, dass zum einen ist ein Dolmetscher, der ist nie zeitgleich. Das bedeutet, selbst wenn ich da einfach einen Untertitel drunter lege und die Daten nehme, dann ist es halt versetzt. Es gibt, also wir, wir arbeiten auch mit der XX XXXXXXXXXXXX zusammen, wo eine Studiengruppe sich mit der Thematik beschäftigt. Was die einfach mal probieren in ihrem Studiengang ist, also heutzutage ist es ja so, dass wenn man Spracherkennung macht, dass man nicht mehr die einzelnen Phonome trainiert, sondern man haut dem Ding den ganzen Satz rein und sagt ihm, was es ist. Und das probieren die gerade auf Basis dieser Daten einmal hinzubekommen. Ob da jetzt ein Ergebnis rauskommt oder nicht, kann ich überhaupt nicht sagen. Das weiß ich einfach noch nicht. Aber für uns ist es halt, die Gebärden sind halt nicht klar abgegrenzt. Also man müsste halt wahrscheinlich noch mehr Sachen oder noch viel mehr Daten haben und sie noch aufwendiger aussortieren. Das zweite Problem ist Gebärden. Also wenn ich ein Wort sage, dann sage ich ein Wort und dazwischen ist eine Pause. Meistens kürzer, aber man kann es ja auch lange machen. Und bei Gebärden ist das anders. Also wenn ich zum Beispiel ABC buchstabiere, dann geht das praktisch ineinander über. Das bedeutet, es macht es noch schwieriger, da klar definierte Trainingsdaten zu haben. Komme ich von einem Wort zu einem anderen, ist das Wort auch irgendwo nicht mehr sauber gebärdet. Also schon für den Gehörlosen klar verständlich und das ist auch richtig, das so zu machen. Aber aufgrund dieser Vermischung dieser Worte ist es halt ja nicht mehr klar unterschieden. Und das grundlegende Problem, was so etwas wie XXXXXXXXXX hat, es fehlen halt einfach grundlegende Wörter, weil sie halt je nach Thematik, also ich glaube das Wort Klimaschutz und Klimawandel könnten wir bis zum Erbrechen darum trainieren, weil das einfach sehr häufig in der XXXXXXXXXX verwandt wird. Ukraine, Russland, Putin, auch kein Problem. Ich glaube, sowas wie ein Bus ist aber schon das erste Problem, weil ich weiß nicht, wann das, also okay, vielleicht gab es schon einen Busunfall, aber sonst wird das Wort Bus ja einfach nicht verwendet. Und wir versuchen uns auch so ein bisschen am europäischen Referenzniveau für Sprachen zu orientieren, an dem, was wir beibringen. Und dann wird es nämlich schon sehr, sehr schwierig. Also schlafen, Bus sind ja alles Wörter, die dann irgendwie fehlen würden.

A1

Ja, man hat natürlich, klar, wenn man Daten irgendwo her hat, hat man immer den Kontext, so nennen wir das irgendwie bei uns im Institut immer. Also Daten ist das eine, aber alle Daten kommen ja aus einem gewissen Anwendungskontext heraus und da hat man natürlich potenziell immer ein Bias drin, wo dann irgendwie zum Beispiel bei Sprache bestimmte Wörter unterrepräsentiert sind. Und was man dann teilweise bei Sprachmodellen auch wieder merkt, dass die halt einfach so, die werden auf dem Internet trainiert und im Internet wird aber halt nicht alles diskutiert. Also schon irgendwo, aber es gibt halt Sachen, die tauchen häufiger auf, die tauchen weniger auf und dadurch entwickeln dann auch die oder haben dann auch Sprachmodelle, zum Beispiel die Angewohnheit, bestimmte Formulierungen oft zu wählen, die dann so nach klassischem Internet-Sprech oder also danach werden zum Beispiel Forschungstexte, die über Sprachmodelle generiert werden, auch so ein bisschen ausgefiltert. Es gibt halt so Formulierungen, die so fast schon klischeehaft für Paper sind, die halt irgendwie sehr oft auftauchen und so Wörter wie Intricate zum Beispiel, das benutzt ja kaum jemand im normalen Sprachgebrauch, ist aber was, was man total schön so irgendwie so in so einem Paper schreiben kann so und ja, da gibt es dann immer Bias drin.

I10

Ich finde die Satzlänge bei diesen Dingern nervtötend kurz. Also ich bin jemand, der unfassbar gerne Bandwurmsätze macht, also halbe Seite, Schriftgröße 10 gedruckt. Ein Satz ist voll mein Ding. Ich habe auch überhaupt kein Problem damit zu lesen. Und diese Sprachmodelle machen ja immer eher so, wie wenn so ein Zehnjähriger schreibt.

A1

Ja, weil das auch wieder, so wird halt auch im Internet oft geschrieben, so kurze Sätze und da haben wir auch gerade so ein bisschen so ein Paradigmen-Shift. Ich habe letztens gelesen, dass irgendwie die XXXXX XXXXXXXXXX XXXXXXXXXXX oder so rausgegeben hat, dass Newsbeiträge zukünftig idealerweise nicht mehr als 500 Zeichen haben. So einfach, weil die Aufmerksamkeitsspanne der Menschen durch Social Media so krass irgendwie sinkt und irgendwie die durchschnittliche Zuschauerlänge bei YouTube-Videos lag mal bei einer Minute 40 und ist zwischendurch nur noch bei 40 Sekunden oder so, bis die meisten abschalten wieder. Also das verändert gerade die ganze Gesellschaft auch, ne?

I10

Na gut, das wir jetzt nicht die hellste Spezies sind, die Vermutung hab ich schon ein paar Mal gehabt.

A1

Ja, naja, okay, aber gut, also Labeling ist auf jeden Fall, also gelabelte Daten da an der Stelle zu bekommen, ist ein großes Problem. Das leuchtet komplett ein. Abgesehen davon, dass wir natürlich irgendwie mal einen Datensatz zusammenstellen müssten, fällt dir trotzdem irgendwas ein, wo du sagst, da müssten KI-Forscher, Machine Learning-Forscher einfach besser werden, das wären Schritte, die das Einsetzen von künstlicher Intelligenz oder die Entwicklung von künstlicher Intelligenz in der Praxis extremst begünstigen würden.

I10

Naja, eigentlich kann ich auch bei dem Satz weitermachen, den ich gerade gemacht habe, unsere Spezies habe ich nicht das Gefühl, ist die hellste. Das ist jetzt wahrscheinlich weniger ein Informatiker-Job als für Psychologen und so. Ich glaube, wir müssen da Bildungstechnisch sehr aufholen, vor allen Dingen bei den Leuten, die Entscheidungen treffen. Also ich glaube, da ist ein großes Thema. Ich weiß, dass es jetzt hier mit Deutschland, also ich bin ja auch noch andere Länder gewöhnt, da ich halt in XXX XXXXXXXXXXXX arbeite und damit auch im Grenzgebiet. Also ich habe gute Kontakte nach XXXXXXXX XXXXXXXXXXXX XXX XXXXXXXXXXX. Weiß, dass ich hier natürlich in dem Land bin, wo das jetzt besonders heikel ist, da zu sagen, weil nämlich hier auch das Zitat gefallen ist, warum muss man lesen und schreiben können für den Bundestag. Grundlegend bin ich aber schon der Meinung auch, dass da einfach große Bildungslücken sind, was diese KI-Thematik angeht. Und ich finde, es hat auch ein bisschen gegipfelt jetzt erstmal in diesem neuen EU-AI-Act. Ich denke mal, ihr habt den auch gelesen. Ich glaube, Paragraph 64, 1 und 2 waren das, wo dann das glorreiche drin stand, dass wir sämtliche Trainingsdaten immer dem Land... Also dass wir in dem Land, wo wir das so entwickeln, die Trainingsdaten, also den Behörden einen Zugang zu den Trainingsdaten geben müssen. Also es steht in diesem Gesetz drin, in dem Land, wo wir das entwickeln, müssen wir den Behörden einen Zugang zu den Trainingsdaten geben. Was im ersten Moment jetzt gar nicht so schlimm klingt, wenn man sich überlegt, okay, hey, wir haben da sowas wie medizinische Produkte. Und wir sollen jetzt einfach mal einer Behörde sämtliche Zugang zu Trainingsdaten geben. Ist das aus meiner Sicht hochgradig heikel, weil ja, die werden wahrscheinlich nicht mit Klarnamen drinstehen. Je nachdem, worüber wir aber reden, ist ja trotzdem zu erkennen, wer es ist. Und ich frage mich auch, wie man auf so eine Idee kommt, das im Gesetz zu schreiben, weil das ja eigentlich für mich mit Datenschutz zum einen überhaupt nichts mehr zu tun hat. Es hat auch für mich mit dem Schutz von künstlicher Intelligenz bei Unternehmen oder mit Unternehmenswerten und Unternehmenseigentum ziemlich wenig zu tun. Ich habe auch nicht unbedingt das Gefühl, dass der deutsche Staat besonders gut aufgestellt ist, was so Security und Ähnliches angeht. Also das sehe ich schon sehr, sehr kritisch. Aber ich meine, wir reden über Leute, die ein Gesetz feiern und glaube ich überhaupt nicht wissen, was sie da gerade reingeschrieben haben oder gerade abgesehen haben. Und das ist für mich so der Punkt, wo man am meisten tun muss, dass man vor allen Dingen erstmal dient und dann vielleicht gesamtgesellschaftlich mal beibringt, okay, was ist eine künstliche Intelligenz? Was kann eine künstliche Intelligenz? Was nicht? Worauf muss man achten? Also das, wenn ich ein Modell schreibe und dann die erste Datei generiere, dass ich eine 50 Megabyte große Datei habe, die nichts kann. Die kann ich mit 100 Katzen füllen, mit Tulpen und Nelken. Das ist dasselbe, aber ich muss sie vor allen Dingen füllen. Das sind glaube ich Schritte, die ja nicht angekommen sind. Und ich weiß, ich glaube, es ist jetzt so sieben, acht Jahre her, wo dann irgendwelche Forscher aus Amerika eine KI gemacht haben und sie mit Passbildern von Homo- und Heterosexuellen gefüllt haben. Und dieser Shitstorm, der dann kam, so Gott, die haben eine rassistische KI gemacht. Nee, ob ich da jetzt Hunde und Katzen reinwerfe oder die, ist ja völlig Wumpe. Ich meine, man hätte es mal das Ergebnis, betrachten sollen und vielleicht auch mal überlegen können, okay, was können wir mit dieser Erkenntnis jetzt einfach mal anfangen? Kann man sie auch in andere Sachen, also ich komme aus der katholischen Jugend, wir haben über Pädophilie mehr gehört, als jeder in seinem Leben hören will. Man hätte es ja auch mal für so eine Forschung verwenden können, dieses Ergebnis. Ich meine, es war jetzt erstmal nur ein spannendes Ergebnis, aber nee, wir machen es einfach mit, die haben da rassistische KI gemacht und dann ist dies Ganze… der Ansatz, der vielleicht mal hätte helfen können, anderen Menschen zu helfen, die da Probleme haben, die wir gesellschaftlich nicht wollen, um es mal möglichst neutral auszudrücken, zu helfen haben wir einfach geschreddert weil wir zu blöd sind. Also das finde ich...

A1

Ja, ich muss da ganz kurz eine kleine Lanze für die Politik brechen, weil die natürlich schon in Grundlagenforschung auch investieren. Ansonsten wäre ich ja jetzt nicht hier. Das heißt, da wird schon geguckt, dass man halt irgendwie Geld in die Hand nimmt, um halt irgendwie die Experten in dem Feld auch irgendwie auszubilden. Ich bin da aber völlig bei dir, dass da teilweise die Kommunikationslücken auch extrem sind und wir haben auch immer wieder Termine mit einzelnen Politikern, die dann zu Besuch hier sind als Geldgeber, wo man dann auch merkt, das ist schon wild, dass diese Person in einem Expertenentscheidungsgremium sitzt, was mit Geldern für KI-Forschung passieren soll. Und man merkt, okay, der Hintergrund dieser Person ist halt so weit weg von der technischen Ebene. Das ist ja, Politik ist bei uns, es sitzen nicht, also Politiker sind nicht unbedingt geschult für das, was sie entscheiden. Und dieses Zusammenstellen von Expertengremien, um die zu beraten und zu Beratungsgremien, das ist auf jeden Fall ausbaufähig. Das unterschreibe ich auf jeden Fall.

I10

Ich bin auch der Meinung, dass sie viel mehr verdienen sollten, weil hier einfach jetzt die Idioten dahingespült werden. Also der, der nirgendwo eine Chance hat, und nie die Chance hat, das große Geld zu machen, der geht in die Politik, weil es die letzte Chance, weil er damit noch das Maximum aus seinem Leben machen kann. Also ich meine, ich kenne das jetzt aus dem Bereich Künstliche Intelligenz, habe ich jetzt nicht so die Kontakte, die richtig viel Geld haben. Ich kenne das auch aus dem medizinischen Bereich. Da ist eine Frau, die Großraum Möchen Gladbach hat, eine Firma, die kann so wunderbare Konzepte machen, die dafür sorgen, dass Leute nicht so an Krankenhauskeimen sterben und ähnlichem. Und die stellt sich halt auch dahin und sagt, warum soll ich jetzt in einem Monat das verdienen, was ich in zwei Tagen verdiene und mich auch noch zum Gespött der Leute machen.

A1

Ja, das kommt ja auch noch dazu. Das ist auf jeden Fall auch kein übermäßig attraktiver Job so. Es ist auch nicht ganz einfach in der Politik halt trotzdem unterzukommen und ich glaube auch, da sind eine Menge Leute, die auch überbezahlt werden, aber das führt glaube ich ein bisschen weg vom Thema. Aber ja, also Bildung ist auf jeden Fall ein großer Punkt, da gehe ich auch voll mit. Also Pädagogik an der Stelle irgendwie, das Wissen, wir hatten gestern noch ein Interview, da haben wir da auch schon drüber gesprochen. Ist ja aber was, was wir jetzt als KI-Forscher im Kern nicht gut machen können. Also ich bin ja kein Pädagoge, ich bin halt in erster Linie Informatiker und natürlich mache ich mir Gedanken darüber, irgendwie Wissenschaftskommunikation und so, aber ich bin dafür halt auch nicht ausgebildet und da sehe ich dann halt irgendwie eher die Pädagogik und Leute, die das Schulsystem irgendwie gestalten und so in der Verantwortung. Da passiert auch was. Fällt dir noch irgendwas ein, wo du sagst, da könnten wir vielleicht als Informatiker auch die Brücke besser schlagen? Oder bauen.

I10

Das einzige was mir da jetzt gerade spontan zu einfällt, ich muss aber auch zugeben, da hatte ich jetzt nicht so wirklich viel Zeit lange nachzudenken weil ich nämlich eher zu den Leuten gehöre, die Gegebenheiten hinnehmen und da versuchen was zu machen und nicht so „Oh Gott das System ist so scheiße“ Gedanken mir mache. Das einzige was ich mir mehrfach selber überlegt habe, ich weiß aber nicht ob es das gibt weil so groß war der Leidensdruck noch nicht. Wenn ich so diese Modelle entwickle und vor allem diese Tests, und ich meine, davon mache ich ja ein paar hundert jeden Tag, wenn ich daran arbeite. Was für mich cool wäre, wäre, wenn es was geben würde, was mir die Daten schon mal so ein bisschen auswertet. Also diese Loss und praktisch maschinell, das muss doch nicht intelligent sein, vielleicht sollte doch eine Intelligenz drin sein. Also dieses Loss und Accuracy Data, dass ich die irgendwo reinwerfe und der einfach mir einen Faktor berechnet, so und so gut ist das Ding. Damit man auch mal so Sachen machen kann, wie… ich bin jetzt gerade zum Beispiel hier bei der Großmutter (?), wenn ich dann hier eine Woche nicht bin, dass ich dem Bereich gebe, das sind deine Lernraten hier, das sind deine Gamma-Werte, lauf du mal die Woche und sag mir am Ende, was das beste Ergebnis war. Da weiß ich aber auch nicht, ob es das gibt oder ob es da wirklich Verfahren gibt, womit man das ausrechnen kann. Also ich meine, klar, wenn man diese wunderschönen Kurven hat, dann ist das auch relativ einfach zu programmieren. Aber ich habe hier vor allem dieses Zickzack-Gedöns hier, weil ich einfach viel zu wenig Daten habe, sondern wirklich auch mal als Mensch eine Minute draufgucken und zu gucken, naja, das war jetzt besser. Das war jetzt besser. Das wäre ganz cool für meine Arbeit, also würde es vielleicht einfach mal erleichtern.

A1

Okay, cool. Das geht nämlich so ein bisschen in die Richtung, mit der wir uns auch irgendwie beschäftigen. Also, was du praktisch suchst, ist so eine AutoML-Lösung. Also es gibt halt sehr viel Forschung und Technologie im AutoML-Bereich, die funktioniert gut, wenn man sehr viele Daten hat oder sehr gut kuratierte Datensätze hat. In dem Moment, wo man in so sehr spezialisierten Cases ist, sehr unterschiedliche Daten hat, die auch so sehr domainspezifische Probleme irgendwie mit sich bringen, wie zum Beispiel so komische Losskurven oder so. Und da suffert dann halt irgendwie AutoML, weil das AutoML geht halt immer von so einer also ja, es ist generell bei maschinellem Lernen, braucht man halt viele Daten, damit es gut funktioniert und je mehr Daten man hat, desto weniger Human Interaction braucht man. Also desto weniger Domänexpertise braucht man bei der Entwicklung von dem Modell, wogegen, wenn man in einem sehr spezialisierten Fall ist, dann braucht man einen Entwickler, der sich richtig gut mit den Daten auskennt. Und aus meiner Arbeit irgendwie kenne ich das immer, wenn wir mit Firmen zusammenarbeiten und so sehr spezialisierte Lösungen irgendwie ausarbeiten, ist Data und Business Understanding 80, 90 Prozent des Arbeitsaufwands. Und nur der Rest ist maschinelles Lernen. Also erstmal die Datensätze zu verstehen und zu kuratieren, zu verstehen, wie das in dem Firmenkontext eingebettet ist und so, da geht der Aufwand hin. Und am Ende schmeißt man dann da irgendwie 20 SK Learn oder neuronale Netze oder was weiß ich für Modelle drauf, testet alles aus, guckt sich die Ergebnisse an und wählt dann das, was am besten ist. Aber um das ein bisschen zu vereinfachen haben wir und ein bisschen damit beschäftigt und das ist der Kern, warum wir diese Studie machen, weil wir und fragen: funktioniert das oder hat das auch praktische Relevanz. Dafür würde ich mal meinen Bildschirm teilen. Und ich bitte dich mal direkt eine spontane Reaktion hier drauf zu geben.

I10

eine spontane Reaktion willkommen in Deutschland. grundlegend habe ich jetzt Hunger weil mich dieses ABCDE an den Nutri-Score erinnert

A1

Ist das positiv oder negativ

I10

Da ich noch nicht gefrühstückt habe, gerade negativ. Darf man Fragen stellen?

A1

Ja, auf jeden Fall, bitte. Also dafür sitzen wir hier.

I10

Also wir reden über eine bilderkennung denke ich mal weil da ImageNet drüber steht. Und powerdraw per inference ist dann der Milliwattwert wenn ich ein Bild erkenne?

A1

Genau, also nicht ein Bild, sondern ein Batch von Bildern, weil man ja irgendwie bei Bilderkennung immer Batchweise Daten durchjagt, weil das auf GPUs irgendwie deutlich energieeffizienter ist. Ich glaube, in dem Fall war es eine Batch-Size von 32, also 32 Bilder werden auf die Grafikkarte geladen, klassifiziert mit dem Modell und das verbraucht dann 600 Milliwattsekunden, also ungefähr 0,6 Wattsekunden.

I10

Corrupted Robustness ist, denke ich mal, die Erkennungsrate.

A1

Ja, so in der Art. Das ist was, was wir auch irgendwie in fast jedem Interview haben, aber es ist natürlich für uns ein super gutes Learning. Also die Top One Accuracy unten links ist praktisch die Klassifikationsgüte über die Testdaten. Bei ImageNet haben wir ja tausend Klassen und irgendwie von jedem, von jeder Klasse ist in den Testdatensätzen glaube ich zehn Bilder oder hundert Bilder oder so. Das sind 50.000, 50.000 Bilder, also 50 Bilder pro Klasse und davon werden 63% richtig klassifiziert. Einfach als richtig oder falsch Klassifikation, nicht irgendwie Top Five ist das unter den besten fünf oder so, sondern einfach nur ist das richtig klassifiziert, 60%. Bei der Corrupted Robustness werden diese Bilder zusätzlich mit Filtern überlagert, also zum Beispiel mit so einer Art Schneefilter oder Rauschen oder eine perspektivische Verzerrung wie so ein Kamerazoom, um zu gucken, ob bei Perturbationen von der Eingabe die Vorhersage konstant oder robust bleibt, um einfach ein bisschen die Generalisierung des Modells zu testen. Und da sieht man, dass die Accuracy natürlich schlechter wird.

I10

Und Running time per Inference ist wahrscheinlich der Durchlauf von einem so

A1

genau also genau wie oben bei powerdraw ist es halt 1,3 Sekunden braucht das zum klassifizieren und verbraucht damit 0,6 wattsekunden

I10

Das ist aber jetzt nicht so richtig schnell oder?

A1

1,3 sekunden für 32 bilder ist

I10

Also hätte ich von der Grafikkarte mehr erwartet. Nicht so der CPU-Wert, den ich da so erwartet hätte.

A1

Ja, ich muss dazu sagen, ich würde jetzt auch meine Hand nicht ins Feuer legen dafür, dass die Werte unten 1000%ig korrekt sind.

I10

Das ist wahrscheinlich auch mal ausgedacht.

A1

Nee, die sind nicht ausgedacht, das ist tatsächlich so gemessen worden an irgendeiner Stelle mal, aber das ist auch schon ein, zwei Jahre her, dass ich das Projekt halt irgendwie gemacht habe. In dem Projekt haben wir halt ganz, ganz viele verschiedene Modelle für ImageNet durchgemessen, auf verschiedener Hardware auch. In dem Fall ist das jetzt auf so einer A100-Node, also auf so einem Cluster-Rechner irgendwie mit A100-GPUs drin. Aber ich glaube, das ist normalisiert pro Inferenz, das heißt, ich glaube, das ist tatsächlich

I10

Ein Einzelbild.

A1

Ne, also ein Batch von Bildern, 32 Bilder. Also kann schon sein, ich meine, 32 Bilder in 1,3 Sekunden finde ich jetzt nicht krass langsam so.

I10

Ja, aber ich hätte trotzdem irgendwie einen geringeren Wert erwartet, wenn man da so eine A100-Konstruktion hat, also gut, ich weiß auch nicht, wie groß die Bilder sind, ne?

A1

Ja, das sind, bei MobileNet müssten das 280 mal 280 Pixel sein. Aber dann hat man auch immer so Faktoren, wie ist die Vorverarbeitung da mit drin oder ist die, in der Regel wird die ja auch auf der GPU dann teilweise ausgeführt und so. Also, wie gesagt, auf jeden Fall wurde das...

I10

Wir wollen ja nicht das Modell, sondern das Label da reden.

A1

Genau, also wir brauchen uns jetzt nicht zu sehr in der technischen Tiefe hier verlieren. Erstmal geht es prinzipiell ums Label.

I10

Da sind zwei QR-Codes. Wenn ich den oberen scanne, was bekomme ich?

A1

Dann kommt man auf das Software-Framework und das Paper, in dem ich dieses Konzept vorstelle und da kann man theoretisch dann auch sich eigene Label generieren lassen, wenn man da eigene Benchmarks hochlädt, kann man sich halt irgendwie entsprechend Label generieren lassen. Der untere führt zu dem Paper, in dem MobileNet vorgestellt wird. Also praktisch das Paper zu diesem konkreten Modell.

I10

Welche Faktoren sind in dem A bis E drin?

A1

Die unteren Faktoren und zusätzlich noch ein paar weitere. Also wir haben noch ein paar weitere Metriken gemessen, wie zum Beispiel Anzahl Parametern von dem Modell, Größe, wenn man das Modell auf die Festplatte schreibt, noch eine Top-5-Accuracy, was eine etwas andere Evaluationsmetrik ist als das Top-1. Die ist immer ein bisschen höher, da ist dann die Frage, ist unter den Top-5-Predictions die richtige Klasse? Das heißt, wir haben noch ein paar mehr Metriken ausgewertet und wir haben alle Metriken zusammengefasst und gewichtet, um Gesamtscoring abzuleiten. Und dabei haben wir Unterschieden zwischen Metriken, die sich auf die Qualität des Modells beziehen und die sich auf den Ressourcenverbrauch des Modells beziehen. Und wir haben das so in dem Fall jetzt ausgepegelt, dass beides gleich gewichtet ist für die Gesamtbewertung. Und deswegen sieht man auch unten, dass die Top-1-Accuracy, die ist ja rot, das heißt, die ist relativ schlecht im Vergleich zu anderen Modellen.

I10

Ach das sind Farben? Ich bin farbenblind, deswegen war für mich jetzt alles so eins und nur das Corrupted Robustness, das war… Ich bin übrigens nicht der Einzige, der farbenblind ist. Das sind 9,8 Prozent der Männer in Deutschland. Also das sollte man vielleicht doch mal überdenken, ob das so richtig cool ist, das in Rot und Grün zu machen. Vor allem bei der Größe ist das auch echt dann nicht... Also eigentlich... Also ich bin eigentlich noch farbschwach und nicht farbenblind. Das bedeutet, eigentlich erkenne ich das auf dem Rechner. Und oben bei A bis E habe ich auch erkannt, dass das unterschiedliche Farben sind, rechts und links. Aber bei dem Top One muss man mir das echt sagen. Also wenn das so eng ist, wird das richtig schwer für so. Das sind 9,8 Prozent der Männer und 1,2 Prozent der Frauen allein in Deutschland.

A1

Da muss man an der Stelle natürlich auch sagen, das ist natürlich noch nichts, was irgendwie marktreif ist. So eine Usability Study wäre was, was auf jeden Fall auch sinnvoll wäre. Es ist super wichtiges Feedback, auf jeden Fall. Ich habe irgendwie, ich meine, bei den Energielabels ist das ja auch alles irgendwie in Farb kodiert, aber da sind die Informationen nie, also da steht zusätzlich überall noch A, B, C, D, E irgendwie dran, so wie oben bei der Skala halt auch. Könnte man ja unten ohne Probleme noch hinzufügen, dass man halt irgendwie einen kleinen Buchstaben daneben packt in der richtigen Farbe.

I10

Also für mich nicht ich hab auch nicht gewusst, dass die Dinger unterschiedlich farbig sind

A1

Ja, aber da hat man zusätzlich ja noch, wie bei der Skala oben, sieht man ja, dass das, was praktisch gehighlighted ist, ja auch größer ist. Also man kann ja jetzt bei der oberen Skala unterscheiden, dass das rechte das A und das linke das C ist, weil die Größe halt auch unterschiedlich ist und da halt der Buchstabe noch dabei steht. Und damit kriegt man dann auch eine Reihenfolge. Bei den Energielabels ist das so, da ist es ja eine vertikale Skala von A bis, ich weiß gar nicht, E, glaube ich auch. Und dann hat man da rechts wie so einen Cursor praktisch, der anzeigt, in welcher Zeile man auf der Skala ist. Aber ja, das ist super wichtiges Feedback auf jeden Fall.

I10

jetzt wo da zwei nebeneinander steht fällt mir zumindest bei powerdraw per inference auf dass das unterschiedlich ist

A1

Genau, also im Endeffekt ist die Farbskala unten genauso wie die Farbskala oben, von grün bis rot. Angelehnt natürlich irgendwie an Energielabel und Nutri-Score irgendwie, weil das halt schon geläufige Systeme sind. Aber man könnte halt unten bei den Badges unproblematisch einfach noch einen kleinen Buchstaben daneben machen, um halt darzustellen, auch für Leute, die irgendwie farbenblind sind, welche Wertung sozusagen das jetzt hier hat.

I10

Ich soll jetzt einfach sagen, was mir so in den Sinn kommt.

A1

Genau. Und auch hinsichtlich, ob das vielleicht in irgendeiner Form in deinen alltäglichen Prozessen und deiner Arbeit irgendwie hilfreich sein könnte oder was sein könnte, wo du sagst, ah, das wäre schon praktisch, wenn sowas irgendwie etabliert wäre als System.

I10

Ich glaube, so wie es dasteht, wäre es extrem heikel, das zu etablieren, weil wir nämlich eben schon über das Thema Bildung vor allen Dingen bei Leuten im Entscheiderbereich geredet haben. Was mein persönliches Empfinden an dieser Grafik ist, und das ist auch der Grund, warum ich diesen Nutri-Score bei Rewe überhaupt nicht beachte, ist dieses Zusammengefasste. Also dieses Gesamt-A, es ist Gesamt-C, davon bin ich überhaupt kein Fan. Ich finde es nicht schlecht, dass unten Werte stehen, wie dieses Corrupted Robustness und Top-1-Accuracy und auch dieses Running-Time-Per-Inference. Also die drei Werte label-technisch zu erfassen, fände ich gar nicht mal schlecht. Ob man sie jetzt aber Buchstaben geben muss, weiß ich nicht. Ich meine, dieses Power-Draw per Inference ist für mich eher so eine Sache für Seite 2. Weil ich nämlich wirklich, also ich finde für mich sind die drei wichtigsten Werte für ein Modell einfach diese Top-1, Corrupted Robustness und die Running-Time. Ich finde, das sind eigentlich die drei Werte. Ich weiß, das ist jetzt ein bisschen gegen dein Studium. Was heißt gegen dein Studium? Aber an deinem Studium vorbei, sondern aus meiner Sicht jetzt einfach. Das sind die drei Werte. Es wäre schon praktisch, wenn man so Modelle hat, wenn man auf Kaggle ist, man hätte immer die drei Werte. Dass man die dann auch wirklich vergleichen könnte. So, okay, das wird helfen, diese Modelle zu vergleichen. Für Seite 2, oder Seite 2 im übertragenen Sinne, wäre dann durchaus so etwas wie Power-Draw per Inference wichtig. Dass, wenn man sich drei Modelle ausgesucht hat, man auch anfängt, die zu vergleichen. Aber für mich sind eigentlich die wichtigen Werte und der für mich wichtigste Wert von einem neuronalen Netz ist eigentlich immer diese Top-1 Accuracy. Gefolgt von Corrupted Robustness. Und dann für die Anwendbarkeit halt auch dieses Running-Time per Inference. Da ist ja dann auch wirklich, ich finde es ein bisschen unglücklich, dass hier bei dem Image-Net in Batch gerechnet wird. Ich meine, kann man natürlich machen. Aber ich finde, es sollte auf den Wert vereinzelt berechnet werden. Einfach daher, weil was in Richtung Anwendungsfall geht. Also wenn ich jetzt hier an [XXXX XXXXXXXXXXXXXX] denke, eine Freundin von mir aus XXXXXXXXXX, die Krebs erkennen möchte. Ich meine, die kann ja nicht warten, bis 32 Ärzte mal ein Bild hochgeladen haben und dann rechnen wir mal durch. Sondern da ist dann der andere Wert, glaube ich, interessanter. Ich finde es aber, was mir persönlich, was jetzt meine persönliche Meinung nicht gefällt, ist halt, dass es dieses gesamte A und dieses gesamte C gibt. Weil ich weiß auch nicht, A ist, glaube ich, auch noch besser als C, denke ich mal . Und das finde ich jetzt, wenn ich mir das Ding angucke, überhaupt nicht. Also für mich ist das Linke viel, viel besser als das Rechte. Und wenn ich mir dann vorstelle, da ist ein Idiot im Bundestag und muss irgendwas entscheiden, der nimmt auf jeden Fall das Rechte, weil da steht das A drauf. Dann haben wir irgendeine Kacke mit 63% Erkennungsrate, obwohl es 81 hätten sein können. Bloß weil irgendjemand, der bis drei zählen kann und dann vom Maschinenlernen schon das Ende der Fahnenstange erreicht ist, in seinem Kopf ein A sieht.

A1

Also man muss dazu sagen, das ist auch was, wenn man das Paper liest, was ich darüber geschrieben habe und was auch in meinem Software-Framework unterstützt wird, das sind super anwendungsfallspezifische Fragen. Es gibt Anwendungsfälle, da hat man eine feste Grenze für die Inferenzzeit, weil man weiß, wir müssen so und so viele Gigabytes an Daten hier pro Tag durchjagen. Das heißt, wir müssen eine Response-Time von unter einer Sekunde haben, zum Beispiel. Also jetzt hypothetisch nicht im Falle von diesem konkreten Bild-Klassifikations-Image-Net-Fall, sondern so ganz allgemein, wenn man mit Leuten spricht, was sind deine Prioritäten, was ist wie wichtig für dich am Ende, haben wir auch ganz oft den Fall, dass die Leute sagen, du, 10% Genauigkeit, mehr oder weniger, ist gar nicht so wichtig. Wichtig ist, dass die Response-Time hoch ist, weil am Ende guckt über jede dritte Entscheidung eh nochmal ein Mensch drüber. Andersrum haben wir, jetzt irgendwie 20 Jahre Forschung zum Maschinelles Lernen und Künstlicher Intelligenz, wo die Leute sehr motiviert berichten, dass sie das neue, beste Modell haben. Und dann sieht man, ah, sie sind wirklich 3% besser auf dem und dem Datensatz als alle anderen Modelle. Und dann testet man das Lokal und findet daraus, ja, die verbrauchen auch doppelt so viel Energie. Da muss man dann auch mal kritisch hinterfragen, ist das eine Verbesserung vom State of the Art oder schmeißen wir das Problem mit mehr Ressourcen zu? Und gerade hinsichtlich Klimawandel ist das halt höchst kritisch.

I10

Andererseits, wenn wir jetzt wieder die kleine Chinesin aus XXXXXXXXXX da reinbringen mit ihrer Krebserkennung, boah, die 3% können jetzt aber, es geht ja auch noch um Gebärmutterhalskrebs, also ich glaube XXXXX fände es schon cool, die 3% zu haben, wenn die das nutzt. Das kann ja schon über ihr Leben entscheiden, ne?

A1

Auf jeden Fall. Und deswegen sagen wir ja auch nicht, wir dürfen keine Ressourcen ausgeben, sondern wir müssen ressourcenbewusst sein. Und wenn in einem Anwendungsfall Accuracy und Reliability von dem System oberste Priorität hat, dann ist das so. Dann ist das auch fair. Und dann könnte man halt auch einfach sagen, alle Metriken außer Accuracy und vielleicht noch Robustness oder so sind vollkommen irrelevant. Oder man sagt halt Accuracy und dann gibt es aber auch, gerade für solche Fälle, gibt es dann ja auch Technologien, um so eine Sicherheit der Vorhersage abzuschätzen. Also das ist mit so und so viel prozentiger Wahrscheinlichkeit auch wirklich Gebärmutterhalskrebs oder halt eben nicht. Und dann hat man darüber wieder andere Metriken. Also dieses, welche Metriken sind wie relevant? Das ist eine absolut Use-Case-spezifische Frage. Und man kann in diesem ganzen Framework die Gewichtung von den Metriken unten natürlich auch anpassen. Und halt sagen, ich habe jetzt in dem Fall, habe ich aufgrund meiner eigenen Überzeugung und in ImageNet haben wir ja diesen Fall nicht. Also keiner macht Krebserkennung auf Basis von ImageNet. Weil ImageNet das als Datensatz gar nicht hervorgibt. Da gibt es ja keine Scans oder irgendwas drin. Das ist halt einfach random Bilder of the Internet. So. Da habe ich jetzt halt gesagt, okay, wir gewichten alle Metriken gleich und gucken auf einen fairen Vergleich zwischen Qualität und Ressourcenverbrauch. Aber selbstverständlich muss man das dann im Anwendungsfall immer spezifisch anpassen. Da bin ich tausendprozentig bei dir. Und das ist halt aber auch was, im Moment reporten halt irgendwie sehr viele Leute dann nur Accuracy und lassen alles andere unter den Tisch fallen. Und das finde ich halt auch fragwürdig. Also das kann nicht Idee von Forschung sein, dass wir einfach nur auf Qualität gucken.

I10

das ist ja auch bei Forschung also Forschung sollte er auch ergebnisoffen sein und bei Forschung stimme ich dir zu aber ich denke dass in der freien Wirtschaft die Modelle die da gerade entwickelt wird schon zu 90 Prozent auf der Accuracy der wert ist

A1

Ja, aber es ist dann ziemlich spannend, wenn man mit, also wir hatten das jetzt auch in Interviews und also an sich hast du vollkommen recht, dass die meisten Leute sagen, Hauptsache das Ding funktioniert. Es gibt dann aber auch viele Leute, die im Management dann einlenken, wenn die rausfinden, ach krass, okay, also das ist nur ein bisschen besser. Menschlich merkt man den Unterschied von dem Sprachmodell kaum, das ist zwar supposed ein bisschen besser, aber als Mensch merkt man das nicht. Aber es kostet uns dreimal so viel, das Modell irgendwie zu nutzen, weil OpenAI entsprechend hohe Gebühren dafür fordert, was ja wiederum im Hintergrund gemappt wird auf Stromverbrauch. Also je größer das Modell, desto kompetenter vielleicht auch das Modell, desto mehr Kosten verursacht es aber auch durch Stromverbrauch und Kühlung von der Hardware und, und, und. Und das übersetzt sich natürlich direkt in Lizenzkosten. So, und wenn man Sprachmodelle auswählt, viele sind halt von den Möglichkeiten da auch völlig überfordert. Wir gehen auf Hugging Face, gibt irgendwie 50 verschiedene Modelle, welches wollen wir einsetzen? Und diese Entscheidungsprozesse ein bisschen transparenter zu gestalten, ist halt unsere Hoffnung, dass da so ein Labeling vielleicht helfen könnte.

I10

Aber ich finde an diesem Gesamtwert, vor allem wenn du jetzt auch noch sagst, dass man das unterschiedlich gewichten kann, dann finde ich das Ding da oben, also diese Gesamt Buchstabisierung noch heikler. Also da muss zumindest auch auf der ersten Seite stehen, welche Gewichte haben welche dieser vier Werte. Dass das zumindest klar ist, weil ich meine, ich kann ja dann Idioten beeinflussen in ihren Entscheidungen, indem ich einfach die Gewichtung da ändere.

A1

Ja, aber das tun ja Leute eh. Also, wenn jetzt, stell dir vor, irgendwer entwickelt KIs für die Management-Ebene, stellt das dann irgendwie in so einem Steering-Meeting vor, sagt, hier, wir haben zehn Modelle ausprobiert, das sind die und die Eigenschaften, wir würden empfehlen, das und das zu nehmen oder entscheiden sie frei, aber man kann das ja immer irgendwie so beschönen, wie man es selber haben will und seine eigenen Prioritäten oder Überzeugungen da reinfließen lassen. Ich bin aber bei dir...

I10

das ist ja der richtige Weg ein Chef ich mache das jetzt mal nicht auf Manager sondern mit Chef hat einen Angestellten, der gibt ihm eine Aufgabe er kriegt ein Ergebnis zurück und wenn der Angestellte sagt das ist die Lösung dann sagt der Chef das heißt auch Geschäftsführer er führt den Laden nicht er entscheidet oder er weiß alles er ist nicht Gott, er ist nur Geschäftsführer wenn ich jetzt aber so in Richtung Politik gehe also ich meine ich traue denen halt auch wirklich zu bei dem Bildungsding, das sie haben dass sie dann im laufenden Betrieb dürfen Sachen mit E nicht mehr verwendet werden also das traue ich auch noch unserer Politik zu, dass sie auf solche dämlichen Entscheidungen kommt und ja wenn dann der TÜV sagt okay wir gewichten alles gleich und wir haben dann ein Modell, das unfassbar viel Energie kostet aber leider zu 99% den Gebärmutterhalskrebs erkennt dann ist das plötzlich verboten weil da irgendjemand was sich überlegt hat also ich finde das zu labeln aber dann bitte alle vier labeln und nicht gesamt

A1

Ja, also das kann man auf jeden Fall so annehmen. Das ist auch ein super wichtiges Feedback. Ich habe ja von vornherein gesagt, wir wollen halt Meinungen dazu und das ist ja weit davon entfernt, dass das ein akzeptiertes, vom TÜV verabschiedetes Siegel oder irgendwas ist oder Entscheidung auf der Basis. Im Moment ist das halt irgendwie, das sind drei Papiere, die ich geschrieben habe und ein GitHub repo mit zwei Sternen. Das ist die Marktreife, die wir da im Moment haben.

I10

Wenn du mir das gleich schickst, kriegst du auch einen dritten. Also ich finde es grundlegend nicht uninteressant. Die Idee, das mal zu klassifizieren, nee, jetzt wirklich, die Idee, das mal zu klassifizieren oder diese Werte zu erfassen, also vor allem bei Kaggle, du hast diese Modelle und man hat das als Übersicht, aber ich bin absoluter Fan von den vier unteren Werten. Die Informatiker gucken vor allen Dingen eh auf die vier unteren Werte und die oberen.

A1

Das ist genau der Punkt. Also ich habe ja jetzt irgendwie schon ein paar Jahre zu dem Thema irgendwie Paper geschrieben und geforscht und was ich irgendwann gemerkt habe ist, die Zielgruppe solcher Labels sind nicht Machine Learner, weil die sagen dir direkt, schick mir das Paper und gib mir den Code. Die wollen kein Label, die wollen in der Tiefe verstehen, was passiert da. Das heißt, die Zielgruppe, für wen ist so ein Label gedacht, ist halt komplett unterschiedlich. Auch hinter dem Nutri-Score oder irgendwie hinter den EU- Labels stehen ja komplexe Testverfahren oder Bewertungskriterien. Die kann man auch kritisch finden, das ist auch fein. Aber der Endnutzer, der einfach im Supermarkt eine neue Waschmaschine kauft, der muss nicht verstehen, wie dieses Label zustande gekommen ist. Der will das auch gar nicht verstehen. Der will diese Tiefe gar nicht. Klar, in unserem Fall jetzt oder in deinem Fall hast du das technische Verständnis und bist auch interessiert daran. Sagst, so ein Allgemein-Rating, das gibt mir nicht genug Informationen. Ich weiß nicht, wo das herkommt, ich weiß nicht, wie sich das zusammenstellt. Das lässt Raum für Manipulation. Ist vollkommen richtig. Ist aber was, was an ein paar Stellen halt irgendwie das Einzige ist, was relevant sein könnte. Nicht unbedingt im Fall von KI, weil die Analogie hinkt. Keiner geht ja in den Laden und kauft eine KI. Aber alle können sich irgendwie OpenAI online irgendwie nicht runterladen, aber online ausprobieren oder sich irgendwie vielleicht auch ein Sprachmodell lokal runterladen, wobei das schon wieder ein bisschen mehr Expertise braucht. Das heißt, diese Kommunikationsebene, auf der man das kommuniziert, jetzt sagst du auch, man müsste darstellen, wie die unten gewichtet werden. Ja und nein. Ich verstehe, dass das für Techniker total relevant wäre und die würden das wissen wollen, aber wenn man zu viel Information da drauf packt, dann ist es kein Label mehr, sondern dann wird es zu komplex. Das ist halt ein ganz krasser Trade-off und ich bin persönlich der Überzeugung, dass so Reporting über diese Themen, das muss individualisiert werden. Viel stärker. Und das haben wir im Moment nicht. Im Moment haben wir Paper und GitHub repos und das versteht man nur, wenn man sich da fünf Jahre mit beschäftigt.

I10

Ich meine, was man alternativ machen kann, wenn ich jetzt so mir angucke, dass man es vielleicht in drei Labels aufteilt. Wenn man das untere sieht, dass man dann ein Accuracy Label hat, ein Energie-Label und ein Time-Label hat. Ich glaube, dann wäre es, ich weiß nicht, welche Werte du im Hintergrund noch nimmst,

A1

Das sind auch nicht viel mehr. Insgesamt haben wir zehn Werte untersucht. Wie gesagt, Modellgröße in Anzahl Parametern ist noch mit dabei, weil das halt auch so gerade für Embedded und Edge-Devices irgendwie relevant ist. Wie viel Speicher hat man zur Verfügung? Passt das Modell da überhaupt realistisch drauf?

I10

Ja, aber dann vielleicht Energie und Speichern, Accuracy und Zeit, dass man daraus ein Dreierding macht.

A1

Ja, wäre auch denkbar.

I10

Ich meine, dann kann man die Werte ja auch ohne Icons, sondern einfach unter die Labels. Also das würde mir schon mal deutlich besser gefallen, dass man einfach also ich meine, was ich ja cool finde, ich gucke drauf und ich sehe die einen und 63%. Und ich meine, wenn ich dann eine Sache habe, wo ich mich darauf verlassen kann, dass diese 63% stimmen, dann brauche ich mir auch ein Paper mit 63% nicht mehr durchzulesen, sondern dann schmeiße ich das schon mal weg und gucke weiter und finde dann die 81% und merke mir das. Also das finde ich eigentlich wirklich cool daran. Und wenn das so auf einer Webseite stehen würde, ich werde jetzt hier in diesem Krebsbereich bleiben, weil ich einfach da finde, das ist unfassbar relevant und da würde unten auf der Webseite so das AI-Label stehen oder was weiß ich auch, von mir ist im Impressum. Ich meine, dann würde jetzt schon also ich habe eine Frau, die hat Krebs, ich gucke mir dann vier Dinge an, wo man sowas mal testen kann. Man guckt ja schon aufs Label und sagt, okay, hier 84% kostet das Fünffache, aber okay, das hat die 84% Label Ding, da will die es hinschicken lassen. Also dafür wäre es cool. Oder wenn das dann auch nur so, okay, die haben die Kosten, die drei kosten weniger, die waren so teuer und davon hat einer ein A- und B-Level, ich nenne das A-Level, auch dafür, aber dann ist so, ich finde, dass da Werte zusammengemixt sind, die nicht zusammengemixt gehören, weil sie einfach unterschiedliche Sachen, aber es ist jetzt auch, ich bin auch speziell.

A1

Ja, aber es ist sehr gut, deswegen machen wir die Interviews. Wir wollen ja individuelles Feedback dazu. Nur der Vollständigkeit halber. Du meinst wahrscheinlich sowas in der Art, wo man so einen Vergleich hat und das ist tatsächlich...

I10

Oh das da unten gefällt mir gut die Properties (Unverständlich)

A1

Das ist im Endeffekt das kleine Tool, was ich irgendwie entwickelt habe, so ein kleines Dashboard, wo man dann aus verschiedenen Datensätzen praktisch so Benchmark-Ergebnisse darstellen kann. Das wäre jetzt hier die ImageNet-Variante und dann kann man da über ein Modell hovern und hier hat man jetzt zum Beispiel schön den Accuracy versus Power Draw-Faktor, also auf der X-Achse Stromverbrauch auf der Y-Achse, Genauigkeit…

I10

Das sind verschiedene Modelle. Ja, das hat ein bisschen gebraucht

A1

Genau, das sind verschiedene Modelle und wenn man jetzt über ein Modell hier irgendwie hovert, dann wird halt das entsprechende Label dazu generiert und das ist wie gesagt alles super prototypisch und nichts, was jetzt irgendwie, wo ich jetzt sagen würde, dass es marktreif, aber das habe ich so, für wenn dann mal Leute zu Besuch kommen und sehen wollen, was ich so in meiner Forschung mache, kann ich das halt irgendwie zeigen.

I10

Also oben links für sowas wie Kaggle fände ich ultra geil

A1

Ja, genau.

I10

Wenn ich jetzt hier der Geschäftsführer und du mein Angestellter bist, würde ich auch oben links erwarten.

A1

Und dann sieht man halt auch, dass das beste Modell in der Regel keine richtige, eigentlich ist das eine Pareto-Front, man würde da in der Forschung von Multi-Objective-Optimization sprechen, also es gibt halt nicht das eine Beste, sondern es gibt halt verschiedene Auswahlkriterien und die muss man gegeneinander irgendwie gewichten oder ausbalancieren und den besten Trade-Off, den machen dann halt hier irgendwie in dem Falle Mobile-Nets, aber das hängt halt immer davon ab, wo man die eigenen Prioritäten setzt. Also ich kann dir gerne mal das Repo dazu schicken, falls du dir das nochmal in Ruhe anschauen willst, wir haben tatsächlich noch einen kurzen Abschnitt im Interview, den ich gleich nochmal, den wir jetzt nochmal kurz besprechen müssen, damit wir in der Zeit auch fertig werden.

I10

Ich finde es auch wirklich spannend. Also deswegen, also das Interesse ist jetzt auch wirklich nicht für das Interview, sondern da.

A1

Ja, sehr cool, das freut mich. Also wie gesagt, schreibe gerade meine Diss über das Thema und dann genau, da kommt auf jeden Fall dann die nächsten Monate noch mehr zu. Wenn du KI entwickelst oder dich mit KI beschäftigst, wo kriegst du deine Informationen her? Wir haben jetzt schon ein paar mal Kaggle genannt, das heißt Kaggle ist auf jeden Fall auf der Liste. Gibt es noch andere Seiten oder Repräsentationsformen, die du dir anschaust, wenn du Informationen suchst?

I10

ich würde das einmal aufteilen wenn ich KI verwende und wenn ich KI entwickle also verwenden ist er für mich eher so was wie ChatGPT oder ähnliches bei sowas sind dann auch Zeitungsartikel glaube ich so einer der ersten an Anlaufstellen also einfach mal Googlen was schreiben Zeitungen darüber nicht dass ich jetzt sämtliche Zeitungen immer für voll ernst nehmen aber es ist glücklicherweise so dass die oder vor allem Wikipedia ist ganz cool weil die halt Quellen machen wenn man Kritiken an irgendwelchen Modellen hat dann ist dann meistens so die Quelle hinterlegt oder noch einer Studie von steht meistens in den Zeitungsartikeln und eigentlich gucke ich wenn ich sie verwende hat irgendjemand schon darüber rum geweint und wenn ja warum und guckt mir dann vielleicht mal an warum sie geweint haben und ob mich das irgendwie wirklich interessiert warum sie geweint hat oder ob es darum ging so ja wir erkennen toll krebs aber es verbraucht viel zu viel Strom das wäre dann auch wieder so der punkt wo ich mir sagen ja ich würde gerne überleben und der Strom ist mir gerade egal. Wenn ich entwickle ist es hauptsächlich Kaggle von wem sehr viel halte ist dieses morpheus tutorial also cedric messner (?) und vor allem hat er doch wunderbar so kompletten code immer geschrieben und normalerweise oder der weg jetzt zu dieser KI war wirklich ich nehme mir bei diesen messner das nächste raus und reicher es damit Kaggle an um es mal ein bisschen runterzubrechen wobei ich bei dem Gebärdensprachprojekt auch wirklich sagen muss dass ich bei Kaggle komplett an meine Grenze gestoßen sind also die hatten dazu gar nichts mehr und dann ist es wirklich stupides googeln und pdf files durchlesen und hoffen, dass irgendeiner noch was gemacht hat. Auf Seite 7 findet man dann von irgendwelchen Asiaten mal ein Ergebnis. Was uns in dem Projekt jetzt weitergeholfen hat, war die... Den kennt ihr, glaube ich, auch. XXXXXXX XXXXXXX, der ist in der Genossenschaft drin. Mit dem dürftet ihr auch ein Interview geführt haben, weil er nämlich genauso wie ich einmal böse verlinkt worden ist und gesagt so, macht mal. Der hat mir nämlich sehr geholfen, weil ich mit ihm einfach eine Kommunikationsbasis habe, wo ich ihm dann sagen kann, okay, das ist der Stand. Und dann habe ich mich mit ihm eine halbe Stunde auf Zoom getroffen und er hat mich einfach mit irgendwelchen Ideen zubombardiert. Und die sind mir dann alle durchgegangen. Also auch dieser persönliche Austausch. Ja. Und ehrlich gesagt, die Art, wie wir jetzt diese Gebärden erkennen, war Ostern bei der Familie. Also das war mein Vater, der da irgendwie auf eine Idee gekommen ist. Der ist vor allem auf die Idee gekommen, aufgrund dieser Energieeffizienz, das hier als kleine lustige Anekdote, was hier so gerade reinpasst. Der kommt ja aus einer Zeit, der ist ja noch mal 28 Jahre älter als ich, wo man relativ wenig Ressourcen hat. Und hat halt eigentlich sich überlegt, okay, wie kriegen wir das möglichst effizient hin? Und ist dadurch auf die Idee mit diesen Post-Estimation, dass man da mehr mitmacht oder eigentlich sich komplett auf diese Post-Estimation verlässt, gekommen. Also wir sind ja sogar gerade ziemlich energieeffizient im Vergleich zu den meisten anderen Modellen. Das ist aber eher ein Zufallsprodukt.

A1

Ja, aber es ist umso schöner, wenn man das dann auch noch effizient hinkriegt. Wir haben uns natürlich irgendwie im Rahmen von der Forschung auch damit beschäftigt, was für Alternativen gibt es denn schon am Markt vielleicht, sind schon etabliert, um Informationen zu KI-Modellen irgendwie bereitzustellen. Dann haben wir natürlich erstmal klassisch eine Publikation, ne? PDF, 40 Seiten sind in dem Fall jetzt alles hier für MobileNet. Auf der rechten Seite haben wir eine ModelCard. Das ist was, was jetzt auf HuggingFace auch eingesetzt wird. Also dieses große Repository für Sprachmodelle insbesondere, wo dann so Informationen zu dem Modell at a glance sind, so ein paar Tabellen. Hier sieht man jetzt zum Beispiel auch, dass das hier mal getestet wurde auf einem Samsung Galaxy S23, wo man dann so ein bisschen Benchmarks hat, Inference Time und so, weil MobileNet jetzt auch insbesondere entwickelt wurde für den Einsatz auf Mobiltelefonen, halt Mobile. Dann haben wir Papers with Code. Ich weiß nicht, ob du das kennst. Ist vielleicht auch was, was du dir mal merken könntest, weil es ist ein bisschen ähnlich wie Kaggle mit dem schönen Bonusfaktor. Man sieht halt für alle möglichen Datensätze und Modelle auch direkt die Paper und die GitHub repos, in denen die thematisiert wurden. Es ist allerdings auch komplett User-contributed. Das heißt, jeder kann da alles reinschreiben. Es ist nicht immer super aufgeräumt. Es ist ein bisschen chaotisch, aber für einige Datensätze und so findet man da auch ziemlich gute Übersichten. Rechts ein Blogpost zum MobileNet. Es geht dann so ein bisschen Richtung Zeitung, wobei Blogposts dann natürlich nochmal ein bisschen einfacher sind, weil sie auch keine journalistischen Ansprüche haben, wobei das heutzutage im Internet irgendwie alles sehr verschwimmt. Weil ja irgendwie auch viele journalistische Auftritte dann wieder so User-Stories und so ein Kram irgendwie hochladen. Ich glaube, der Unterschied zwischen Zeitungsartikel und Blogpost, das verschwimmt gerade so ein bisschen im Zeitalter der Digitalität. Links die Dokumentation, so wie sie in Ceras von TensorFlow drinsteht, wenn man das Modell halt in der Praxis einsetzen will. Das ist auch das, was ich benutzt habe, um bei mir die ganzen Modelle mal durchzutesten. Und rechts haben wir jetzt nicht zum MobileNet, sondern ganz allgemein. IBM Factsheet. IBM hat das halt irgendwie lizenziert und da Patent drauf und stellt das nur für IBM-Produkte zur Verfügung. Ist aber sehr ähnlich wie die Model Cards, die ursprünglich mal von Google vorgeschlagen wurden, wo man so ein bisschen so eine ganz, also eine recht einfache Betriebsanleitung für ein Modell hat. Wie ist das einzusetzen? Was macht es? Was sind so die Performance Indicators? Wenn man das jetzt im Vergleich sieht, kannst du da vielleicht auch nochmal ein bisschen zu kommentieren. Was fällt dir dazu ein, auch in Bezug auf unseren Labeling-Ansatz? Wo sind die Vor- und Nachteile?

I10

Also was mich an diesen Artikeln immer unfassbar gestört hat, also hier Mitte unten, ist, dass sie halt aufgrund dieses wissenschaftlichen Anspruchs unfassbar viel Text haben und es unfassbar gedauert hat. Also ich musste teilweise vier, fünf Seiten lesen, um zu sehen, okay, sie machen es nicht so, wie ich es brauche. Also das ist schon wirklich nervig. Model-Card, also sowas wie eine Model-Card als Anfang zu haben, ja, so eine Model-Card also vor allen Dingen auch, dass man die Input-Daten mal klar macht. Also das ist eigentlich so das Wichtigste für die Modelle, also dass man den Input eigentlich am Anfang mal beschreit, dass man dann direkt schon mal aussortieren kann, okay, was kann ich nutzen und was nicht.

A1

Also ein bisschen die Spezifikation der Schnittstelle. Also was geht rein ins Modell, was kommt raus aus dem Modell ...

I10

Ja, also grundlegend finde ich also hier dieses Model-Card, ich gucke es mir gerade mal so ein bisschen nebenbei an oder habe es gerade einfach mal aufgemacht und scrolle gerade runter. Ist das das unten links?

A1

Ne, das In dem Fall ist das hier gerade, also

I10

Ja, genau, das hatte ich auch so im Kopf.

A1

Das ist jetzt hier auf HuggingFace zum Beispiel, da hat man so ein bisschen, wie ist das zu benutzen? Model Stats, Model Checkpoint, Input Resolution, Number of Parameters, Model Size. Das ist klar, es geht um Bildklassifikation, das heißt so ein bisschen diese Spezifikation ist dann mit drin und man hat aber auch direkt eine Betriebsanleitung sozusagen, wie man das benutzen kann.

I10

Das ist wiederum cool. Ich kannte HuggingFace gar nicht, also ich kenne HuggingChat und hab da mal ein bisschen rumgespielt, aber ich wusste jetzt nicht, dass HuggingFace auch die Dinge alle hat.

A1

Ja doch, also Hugging Face ist halt, wenn man guckt, ist auch wieder eine Übersicht, ähnlich wie Kaggle, gibt es halt ganz, ganz viele verschiedene Datensätze und halt entsprechend dann für jeden Datensatz auch wieder ganz, ganz viele verschiedene Modelle irgendwie dafür vorgeschlagen werden und hauptsächlich ist Hugging Face groß geworden in den letzten Jahren, weil die Open Source Sprachmodelle irgendwie bereitstellen und ganz, ganz viel...

I10

[Unverständlich]

A1

Ja, genau. Darüber kennst du die meisten. Bildgenerative Modelle und Sprachgenerative Modelle. Dafür, da ist HuggingFace halt ganz groß ist. Und was ganz lustig ist, die haben halt auch so interaktive Spaces, wo man einfach Modelle auch mal ausprobieren kann. Also, ich habe jetzt keine Ahnung, was das ist. Fake Insects. Start playing. Okay, das ist jetzt ein bisschen so ein Game. You click the real Insects. Achso, da wird noch ein zweites Bild geladen und so. Damit werden jetzt zum Beispiel hier ein bildgeneratives Modell, um erfundene Insekten sozusagen darzustellen. Aber man hat halt auch...

I10

Ich kenne HuggingChat, also das fand ich jetzt nicht so überragend, als sie es mal verwendet haben, aber es war jetzt so mal als Spielerei und vor allem, dass es Open Source ist fand ich es ganz cool

A1

Genau, das ist so, hier Mid-Journey ist ja einer der größten bildgenerativen Modelle, die waren halt auch direkt hier drauf, dann genau, und man kann mit denen auch, also diese ganzen Spaces hier, die sind halt, die werden im Hintergrund, wenn man die viel benutzen will, muss man halt auch wieder Geld bezahlen, weil das halt auf irgendwelche Offshore-Computer-Resources irgendwie linkt, das heißt, die haben auch direkt so ein Cloud-based Infrastruktur-Ding mit dranhängen, also ist ganz cool

I10

Ja, dann wird sich Modelcard, glaube ich, schon, obwohl man vielleicht mal über die Aufteilung reden sollte. Also hier dieses, was da oben steht, diese Model State, Input Resolution und so, das wäre eigentlich für mich das, was ganz oben stehen sollte.

A1

Ja, das ist ein bisschen das Problem, wie du vorhin sagtest, es gibt halt keinen Standard. Jeder macht es halt unterschiedlich und dadurch ist es nicht so einfach, die Informationen immer direkt zu finden, die man gerade sucht. Auf Kaggle sieht es anders aus als auf Hugging Face, da sieht es wieder anders aus als bei, weiß ich nicht, der Keras-Dokumentation und

I10

Ja, ganz schlimm sind dann diese wissenschaftlichen Artikel, da suchst du dich einfach dämlich.

A1

Genau, das ist halt, das ist auch wieder eine andere Zielgruppe. Das sind halt Leute, die in der Tiefe, die sich eine Stunde Zeit nehmen wollen, das in Ruhe zu lesen und wirklich zu verstehen. Das sind dann nicht die Anwender, die das nur mal eben schnell ausprobieren wollen. Und, ähm.

I10

Ich will auch nicht so eine Zeit nehmen, aber ich will wissen, ob ich mich nicht so eine Zeit nehmen kann. Also ich meine, ich gucke ja auch Netflix, gucke mir auch die Zusammenfassung durch und weiß, ja okay, die Romantik-Komödie aus Indien ist jetzt nicht unbedingt meins.

A1

Ja, das wird inzwischen passiert da auch viel, weil ganz viele Leute, wenn sie ein Paper schreiben, dann auch so ein im GitHub ein kurzes Readme haben, was dann wieder so ein bisschen mehr Richtung so eine Model Car geht oder die schreiben noch einen Blogpost dazu und so also da passiert gerade viel auch und wir gliedern uns da halt mit dem Labeling so ein bisschen ein Vielleicht kannst du da nochmal drauf eingehen, siehst du da wo siehst du da Verbindungen zu dem, was wir jetzt hier vorgeschlagen haben

I10

Also vor allem, wenn ich jetzt auf diese Model Card gehe, also man lässt nochmal diese Hugging Face Seite auf, da kann man das auch wunderbar... Einfach so ein Modellding. Also ich meine, wir haben ja da oben… Warum machst du dich selbst so schlecht? Bleib oben. Wir haben da oben Space Shuttle, 86 Rockets und so weiter und so fort. Auf der rechten Seite haben wir ja schon Informationen. Downloads last month. Das waren jetzt nicht so richtig viele. Aber da oben dann drei Scores mit Energieeffizienz, Geschwindigkeit und Erkennung. Das könnte ich mir wunderbar vorstellen. Da wäre es vielleicht nicht in diesem Label, das sowieso aussieht, als ob man es auf irgendeine Maschine klebt. Aber da oben dann diese drei Scores mit ABC und man klickt drauf und kriegt dann die Parameter, die es berechnen. Also dieser Liste, das A leuchtet groß. Und wenn wir jetzt hier bei der Accuracy sind, ich klicke drauf und dann öffnet sich halt so ein Popup mit Top 1, Top 5 und welche Werte waren noch aus dem Bereich Accuracy? Ich kann überhaupt nicht mehr reden. Mit dahinter noch die Gewichtung dieser Dinge. Das könnte ich mir wunderbar vorstellen. Das wäre auch wirklich einfach mal hilfreich. Ich mache das Ding auf, ich sehe C, okay, weg. Oder ich will hier das auf mobil machen, ich mache so ein Ding auf, ich sehe Energie ist E, dann, oh Gott, da brauche ich gar nicht mehr weiter gucken es ist ein Einblick.

A1

Das ist ja dann aber tatsächlich recht ähnlich, als wenn man einfach das Label hier darstellen würde oder den unteren Bereich des Labels hier darstellen würde. Ich meine, das ist natürlich jetzt eingebettet hier in Hugging Face. Die Idee ist ja, dass man das hier einfach auf irgendeine PowerPoint-Folie draufpacken kann und dann sieht man at a glance, okay, es geht um den Datensatz, das ist der Name des Modells. Hier gibt es weitere Informationen und das sind die wichtigen KPIs, die wichtigen Metriken. Also ich sehe da schon eine Ähnlichkeit einfach. Klar, die Darstellungsweise, die kann halt verändert werden.

I10

Nein, aber was ich halt übernehmen oder was ich sehr gut finde, wenn es übernommen wird, sind die Werte, die du erfasst Also das wäre ja wirklich, also ich meine, das fehlt mir auch auf Kaggle permanent Ich meine, da muss ich dann wirklich bis unten durchscrollen und mal gucken was hat er da so gemacht und er macht ja auch vor allen Dingen nur Top 1 und so weiter und so fort. Das wäre aber halt auch diesen Zugang zu Informationen Also wirklich auf dem Paper nur drauf zu haben, finde ich so Oahh… Also weil wir da nur eins sind, so nee, also das ist halt

A1

Ja, das ist...

I10

Also eigentlich sagst du ja damit nichts aus, wenn du nur dieses ABC, weil wenn du die Gewichtung nicht dazu schreibst und so, ist es halt einfach so, ja, wir haben jetzt ein bisschen Tintenfarbe verbraucht. Also ich muss auch ganz ehrlich sagen, ich habe auch noch nie dafür, also ich habe keine Ahnung, was der Nutri-Score ist, es hat mich nie interessiert und es beeinflusst auch nicht meine Kaufentscheidung, ob der Nutri-Score A oder E ist.

A1

Wie ist das mit Energie-Labels?

I10

Beim Kühlschrank. Darauf achte ich. Das finde ich auch ein gutes Label, weil zumindest außer der Name ist falsch habe ich da das Gefühl, es geht rein um die Energie. Das ist ein Wert, der wird verglichen, es ist kein, wir nehmen sieben Werte, schmeißen den Mixer an und haben einen Buchstaben raus.

A1

Ja, aber natürlich, das stimmt so semi. Im Hintergrund steht natürlich, also es steht auf dem Label nicht drauf, aber natürlich testen die verschiedene Sachen im Hintergrund. Und es gibt dann immer so einen Hauptfaktor, wie irgendwie, wie viel Wasser und Strom wird verbraucht, um sieben Liter Wäsche zu waschen. Also jetzt nicht beim Kühlschrank, aber bei der Waschmaschine. Beim Kühlschrank ist es, glaube ich, gegeben, der und der Wohnungstemperatur, wie viel Strom wird verbraucht, um den Kühlschrank bei halber Beladung in der und der Temperatur zu halten.

I10

Ja, aber da geht es ja auch wirklich nur um Strom.

A1

Klar, ja, ja, logisch, ja, logisch, ja.

I10

Und das ist für mich ein vergleichbarer Wert. Und solange ich da fünf Kühlschränke habe, wo ich den Marken vertraue, dass der Kühlschrank auch läuft, dann würde ich auch mal auf das Energieeffizienz-Ding gucken und sagen, ja, die beste, dann nehme ich das auch. Wo ich es zum Beispiel nicht machen würde, ist Bildschirme. Ich meine, ich habe Bildschirme, die haben eine kurze Reaktionszeit, ich habe Bildschirme, die haben eine lange Reaktionszeit, damit mache ich auch einfach teilweise andere Sachen. Oder es ist mir hier auch einfach da, wo ich gerade bin, völlig egal, ob der jetzt schneller oder langsamer, ob der jetzt zwei oder acht Millisekunden braucht, es ist mir völlig wumpe beim Programmieren. Da hilft mir so ein Label nicht, weil ich habe keine Ahnung, welche von denen jetzt die beste Energieeffizienz hat und vor allen Dingen geht es mir da auch um andere Werte. Also wenn es um eine Sache geht, dann finde ich Labelings gut. Aber sobald es gemischt ist, ist es mir auch irgendwo, dann ist es einfach nur so nervtötend, dass man nicht direkt das sieht, was es ist.

A1

ist. Ja, verstehe ich komplett. Das ist, glaube ich, ein gutes Fazit. Also damit kann ich auf jeden Fall gut leben. XXXXX, hast du noch irgendwelche Fragen oder Anmerkungen?

SPEAKER\_01

Die ist sowieso sehr still

A1

Ja, wir haben das tatsächlich so gemacht, dass wir immer sagen, einer führt das Interview und der andere passt immer so ein bisschen darauf auf, dass falls Fragen offen sind, die man selber vielleicht vercheckt, nur dann eingegriffen wird und man nicht so die ganze Zeit durcheinander redet. Aber ...

I10

Also je mehr ich darüber nachdenke, teil die dinger auf und mach drei draus. Oder von mir aus auch 4 wenn dir dann fehlt, dass man halt Accuracy, Geschwindigkeit… Wieso kann ich heute nicht reden? Accuracy, Geschwindigkeit und Energieeffizienz einzeln hat. Dann ist es hilfreich. Also dann guckt man da drauf und denkt sich so, ja, da habe ich weil es ist halt sonst Mixer und Mixer ist doof

A1

Ja, nee, es ist ein gutes Fazit. Wir sind so ein bisschen den Einsatz gefahren, oder ich bin bisher den Einsatz gefahren von, wir wollen einfach transparent unterschiedliche Faktoren beleuchten, aber man könnte natürlich auch wirklich sagen, man macht unterschiedliche Labels. Es ist nur schöner, dass halt irgendwie auf einen Blick zu haben, als wenn man jetzt dann verschiedene Labels irgendwie mit, aber am Ende ist das auch ein bisschen Präferenzfrage. Deswegen machen wir ja auch die Studie und Usability war jetzt gar nicht so der Kern Fokus, sondern eher so ein bisschen allgemein Feedback dazu. Ich würde mich auf jeden Fall an der Stelle für die Aufmerksamkeit und fürs Gespräch bedanken. Ich fand das sehr angenehm und sehr interessante Takeaways . Wir sind jetzt mit der Interviewphase so ziemlich durch. Wir werden jetzt die nächsten Wochen alles in der Tiefe analysieren und kodieren und dann schreiben wir ein Paper darüber, was du dann nicht lesen willst. Aber nein, ich werde natürlich...

I10

Das darf gerne weitergeleitet werden.

A1

Da geht es um ein paar

I10

Da ich jetzt schon weiß, was drinsteht und weiß, das Thema könnte für mich interessant sein, dann nehme ich mal die Stunde

A1

Genau, also wir werden auf jeden Fall auch gucken, dass wir das in Form von einem Blogpost oder so, denke ich mal, auch nochmal aufbereiten, dass man da ein bisschen schneller durchkommt. Uns geht es ja in dem Paper dann vor allem auch um die...