

Aber jetzt oder? Euch an die Noten berechnen. Und ich habe hier die Punktzahl der ersten Aufgabe und die Punktzahl der zweiten bis 22. Und ich möchte gerne die erste Aufgabe viel, viel, viel, viel stärker gewichtet werden.

Also könnte ich hier vorne beim Anfang sagen: "Zum stärkeren Gewichten werden", dann sage ich 100×1 . Und die zweite, die weniger stark gewichtet werden soll, dann mache ich 20×2 .

Und jetzt hier die Punktzahl, aber bei einer Aufgabe, bei einer 2×1 , dann ist es ja logisch, dass das hier 100-mal meine Punktzahl, jeder Punkt 100-mal gewertet wird und auch noch 20. In diesem Sinne Parameter.

Jetzt kann ich hier eine Zeile schreiben oder ich kann hergehen und sagen: "Ich mache mehrere Zeilen." Ich kann sagen, y_1 ist gleich ax_1, bx_2, y_2, jx , und jetzt muss ich hier irgendwie ein bisschen anders nummerieren.

Dann machen wir es billig, ohne hier große Indizes schlecht zu machen, sagen wir x_1, x_2, x_3 und so weiter. Und da kann ich eben auch 1, 2, 3 schreiben, spielt nicht so eine Rolle. Gut, jetzt schauen wir mal hier an, wie das ausgesehen hat. So soll es sein oder?

Und jetzt kommt der Fuhr Mathematik und sagt: "Jetzt habe ich das mal hergeschrieben." Sondern ich nehme nur a_1, b_2, a_3 und b_3 und schreibe das so hin: a_1, b_2, a_3, b_3 und sage, das ist eine Matrix.

Da sind nur Gewichte drin. Und jetzt muss ich mit diesen x und diesen y ein bisschen etwas machen. Das mache ich vorne, nehme x_1, x_2 raus und schreibe es so hin als Vektor. Das ist das Problem nicht.

Und dann habe ich ja schon einen Vektor, schreibe das so hin: x_1, y_1, y_2 . Wenn ich so schreibe, ist es ein Matrixschreibweise. Wenn ich es so schreibe, ist es eine ganz normale lineare Gleichung. Das Gleichungssystem.

Sag mal zum Verständnis, weil 2k-Vektorprüfung, das heißt, y_1 wäre eine Prüfung, y_2 wäre eine andere Prüfung und dann nachher würden wir jetzt in dieser Matrix alle die Gesamtpunktzahl vor allem Prüfungen zeichnen. Exakt. Und das ist auf das Mal rechnen.

Da gibt es aber nämlich ein Video, die Gamaschine, die GPU, kann ich ganz schnell rechnen, oder? Dort sind wir schnell am Resultat. Also kann ich auf die GPU eingeben, ich gebe hier Punktzahlen ein, das ist der Reststudent, ich habe zwei Prüfungen von diesem Student und dann wird das ausgerechnet und hier gibt es dann auch die gewichteten Punkte.

Jetzt, das ist einfach das Beispiel, oder? Jetzt, was passiert? Der letzte Satz, das werden wir dann zusammen nochmal anschauen. Das hier sind Gewichte. Aber Gewicht in einem neuronalen Netz habe ich auch von Gewicht geredet. Das ist das Gleiche.

Gewicht in einem neuronalen Netz, im Allen, das ist nichts anderes als Zahlen in dieser Matrix.

Wenn ich eine Abfrage mache, ein Prompt, dann ist das hier das Prompt, das ich eingabe, mit der gelernten, trainierten Matrix wird das interferiert und am Schluss komme ich hier die Antwort heraus. That's it. Und darum braucht man so.

Und die Matrix ist viel, viel, viel, viel größer, das sind Billionen oder was für mich. Die Matrix ist riesengroß, darum braucht man große Rechenleistung.

Also wenn ich jetzt das Beispiel mache, ein Hund ist miss das x_1 , dann würde es reingehen in die Matrix und würde dann nochmal schauen, wie viele Hunde waren miss das x_1 , wo mir gekotet ist.

Genau, es würde im Prinzip interferieren und ich würde im Prinzip die entsprechende Antwort bekommen, die hier verkotet ist.

Man muss sagen, LLMs sind eine Blackbox, ja klar, aus diesen Zahlen kenne ich nicht so viel. Es gibt Techniken, wo man da ein bisschen schauen kann, was ist, aber im Prinzip ist das das Rennen. Wichtig ist, dass es statistisch oder statistisch gesehen wäre. Ja, ja. Aber das kommt viel, viel später.

Einfach nur um zu sagen, was ist eine Matrix und jetzt kann man noch einen Tensor machen, man kann es noch viel, viel komplizierter machen. Vielleicht so das erste Beispiel. Wird nicht gefragt, morgen. Ja.

Gut, nochmals zum letzten Mal. Ich habe gerade das Ding hier gesehen, Alltagssprache. Wir haben letztes Mal was gesagt von Alltagswissen. Wisst ihr noch, was das ist?

Common Sense war das englische Wort. Allgemeinwissen.

Genau, Allgemeinwissen. Wissen, dass wir voraussetzen können, sozusagen, dass jeder das kann. Gesunder Menschenverstand wird es auch manchmal genannt. Okay, in welchem Zusammenhang haben wir das gebracht letztes Mal?

Ah, das war System 1 und System 2.

Was ist das Sprachmodell?

Ja, das ist nicht ganz.

Das sind Expertensysteme, die ein Saugroboter betreiben, zum Beispiel E-Scans, ich glaube, das sind die mit der Technik.

Dann geht es nachher aufs Allgemeinwissen, zum Beispiel ChatGPT oder so, damit man eine klare Aussage hat. Spezifische Aussage zu dieser Firma. Also Allgemeinwissen, das kann man immer wieder gebrauchen.

Und wenn wir eine Anwendung machen, haben wir sehr viel Allgemeinwissen, das wir da auch verwenden. Wenn ich jetzt eine Anwendung mache, müsste ich das jeweils immer neu programmieren. Und da gab es die Idee von Dougland, der das Projekt gemacht hat, wir könnten eine Wissensbasis machen mit diesem Allgemeinwissen.

Und das könnten wir für die verschiedenen Anwendungen verwenden. Wir müssten nicht bei jeder Anwendung dieses Allgemeinwissen nochmal neu in die Wissensbasis bringen. Würde uns Zeit sparen. Außerdem würde es uns helfen beim Lernen, so wie wir lernen, wir verknüpfen das, was wir neu lernen, mit dem, was wir schon kennen.

Und das heißt, je mehr ich weiß, desto einfacher ist es, was Neues zu lernen. Das war so, da haben sie Wissensbasis gebaut, SAIC, der praktisch das Allgemeinwissen enthält.

Und wenn man das jetzt mal überträgt auf die Sprachmodelle, dann kann man auch sagen, die Sprachmodelle sind das Allgemeinwissen. Wenn ich es im Unternehmen anwenden will und wirklich für Unternehmensspezifische Lösungen haben will, muss ich das Wissen von dem Unternehmen noch mit dazu bringen.

Also welche Produkte haben wir, wie ist unsere Abteilung organisiert, wer sind unsere Kunden und so weiter. Das steht ja in dem Sprachmodell nicht drin.

Also man kann sagen, dass das Sprachmodell im Prinzip auch eine Repräsentation von dem Allgemeinwissen ist und wenn wir es im Unternehmen anwenden wollen, müssen wir dort wieder Wissen dazu geben. Und ich habe das, glaube ich, auf der Folie gezeigt.

Hier, das war das mit dem SAIC. So, jetzt muss ich die Folie zurück.

Okay, noch so nicht. Jetzt.

Okay. Das war mit SAIC.

Und hier, das war mit den Unternehmensanwendungen, dass es immer Allgemeinwissen ist und das unternehmensspezifische Wissen. Und das kann man mit SAIC, wir haben im Prinzip die Ontologie oder man nennt das Sprachmodell und kann es ergänzen um Unternehmenswissen. Ein Ansatz ist Retrieval Augmented Generation.

Ich nehme die Dokumente von dem Unternehmen, kann eine Suche machen in diesen Dokumenten, da habe ich mein Unternehmenswissen, gebe ich an das Sprachmodell und mit beidem kann ich dann eine Anfrage beantworten.

Also wenn wir eine Unternehmensanwendung machen wollen, natürlich können wir Copilot nehmen, aber das greift dann auch zum Beispiel auf unsere E-Mails zu oder wir können spezifische Anwendungen machen und das Sprachmodell ergänzen, um Wissen, das in Dokumenten von dem Unternehmen steckt.

Das, was wir mit Sprachmodellen generell kriegen, das ist eher das Allgemeinwissen und nicht das unternehmensspezifische Wissen. Aber wir wollen ja auch unternehmensspezifische Anwendungen bauen. Also müssen wir wissen, dass man es auch ergänzen kann. Da hätte ich schon jetzt Probleme für Sicherheit.

Wenn ich einfach eine Versicherung habe, habe ich glaube ich letztes Mal schon erzählt, wie tue ich sicherstellen, dass das Wissen da nur an spezifische Personen weitergegeben wird. Wenn ich mir vorstelle, der Zugriff auf eine Datenbank kann ich einschränken, sagen wir das, das, das, darf nicht weiter wissen.

Aber hier, das Netz kann irgendwie blöd. Und wenn man geschickt fragt, kann man irgendetwas herausfinden. Also wenn man indirekt fragt, dann kriegt man auch die Findung. Und das ist immer noch ein Softmix-Problem.

Darum sind verschiedene Firmen in der Zusicherung im Speziellen oder jetzt auch die Quintedur, die ich von ihr angesprochen habe, nicht so Fan von dem momentan noch. Und sie machen Folgendes. Die meisten Mitarbeiter, die allgemeine Sachen wissen müssen, haben ein RAG.

Und Leute, die Spezialwissen wissen müssen, haben ein eigenes RAG. Also man tut die RAGs drin. RAG meine ich, dort, wo das Unternehmenswissen im Prinzip abgespeichert wird.

Das ist eine Funktion von einem Agenten, aber ein Agent, den wir anfragen können, der dann noch auf zwei verschiedene Systeme macht. Also der Agententisch schauen wir noch ganz, ganz im Detail, ja. Ja, wir können dem einen Agenten sagen, der dann im Prinzip Information Retrieval macht. Aber immer.

Dann hat man da ja limitiert, du hast da Zugang, du hast da. Ja, genau. Das Problem ist, mit dem Limitieren ist, ich muss ja jede Abfrage, muss ich wissen, ob die gehemmt ist oder nicht. Sonst brauche ich ein System, das weiß, was. Das gibt ein bisschen Probleme. Werden wir noch anschauen.

Gut, ich glaube, so weit sind wir letztes Mal gekommen. Das haben wir ein bisschen Sachen auch wiederholt. Eine Sache, die mich fasziniert hat, als ich mich das erste Mal mit KI beschäftigt habe, ist, dass das, was uns so einfacher scheint, eigentlich für die KI unheimlich schwierig ist.

Also ich habe hier mal ein paar Beispiele, nach etwas zu greifen, einen Gegenstand zu sehen und danach zu greifen, etwas hochzuheben. Für uns relativ einfach, für eine Maschine unheimlich schwierig.

Dass man den Punkt trifft, dass man auch nicht so fest zugreift und was nicht zu zerstören und ja, aber fest genug, um etwas hochheben zu können. Also unheimlich komplex für ein System. Oder Objekte zu erkennen.

Also ja, wir gucken uns irgendwas an und wir wissen, da ist ein Auto drauf und da ist der Manuel und so. Die Sachen zu erkennen, für uns relativ einfach, für ein System unheimlich schwierig. Oder ja, eine Handbewegung, etwas anzusteuern mit der Hand.

Ja, das, was ich gesagt habe, präzise nach dem Gegenstand zu greifen, die Muskeln so anzuspannen, dass man was nicht kaputt macht, aber trotzdem hochheben kann. Und ja, oft geht es, dass etwas schiefgeht. Wir als Menschen machen auch Dinge kaputt oder so, aber wir lernen relativ schnell.

Aber das im System beizubringen, wie es greifen muss, das hat für zum Beispiel ein Ei hochzuheben, hat Jahre gedauert, bis man überhaupt dann eine Idee hatte, wie man das machen kann.

Das passiert auf dem, dass wir ja durch die Evolution diese Sachen praktisch schon irgendwie einprogrammiert haben. Wir müssen das nicht alles nochmal neu lernen. Aber wo geht es jetzt in Sachen, die uns so schwieriger erscheinen? Ganz einfach für einen Computer. Also relativ einfach. Ich habe schon sehr früh Schachcomputer.

Es gab, vorher gab es Dame spielen. Generell rechnen ist für einen Computer einfach. Wir brauchen Jahre, erstes, zweites Schuljahr, um überhaupt mal einigermaßen mit den Grundrechenarten zurechtzukommen.

Also die Sachen, die uns einfacher erscheinen, sind für einen Computer oft schwierig. Das hat was, was mich total fasziniert hat an der KI, dass man so die Vorstellung hat, was geht schnell, was ist einfach, einfach nochmal revidieren muss. Das ging ja auch den Leuten, die Forschung gemacht haben, in den ersten Jahren auch so.

Da haben sie ja gemeint, man kriegt schnell einen Computer, der denken kann, bis Anfang des 17. Jahrhunderts. Gut, okay. Heute ist das Wort Agenten gefallen. Was ist eigentlich ein Agent?

Vielleicht habe ich jetzt einen falschen Begriff davon. Aber was ist ein Agent?

Chatbots, wo ich eine Reihe und da unten entscheidet, wenn ich das aber auflage.

Du würdest Agenten auf Chatbots beschreiben.

Ich würde Chatbots, ja, Chatbots, dann haben wir Agenten und der Agent wird dann nochmal definieren, wenn es nicht geht. Also ein Dispatcher sozusagen. Ja, okay.

Wenn ich jetzt zum Beispiel von einer Firma einen Agenten erstellen will, brauche ich gewisse Angaben, oder? Die wird dann wieder überspringen. Die werden dann weitergeleitet, und mit dem kann ich jetzt zum Beispiel WhatsApp, Kalender und andere Sachen damit dann zufahren. Mit GoHighLevel und andere Sachen.

Was ist jetzt ein Agent, anders als ein Computerprogramm? Der Agent interagiert mit mir und der berechnet mir quasi die beste Lösung. Nicht nur irgendeine.

Und wenn ich jetzt zum Beispiel einen Laufnehmer hätte und Laufpennels verkaufen wollen würde und ich ihm sage, okay, guck mal, ich habe ein Schrägdach, was ist jetzt die beste Lösung dafür? Dann sagt er, keine Ahnung, irgendwie so für Laufdach wäre Schräge geeignet.

Für ein schräges Dach wären die Platten geeignet, weil es sehr schräg ist. Dann kann ich zum Beispiel, wenn ich Preis anfordere oder sonst irgendwie einen Termin mit einem Neubau machen, dann wird das alles weitergeleitet. Wenn ich jetzt zum Beispiel einen Chatbot irgendwie fragen würde, dann würde mir einfach die nächstbeste Antwort geben.

Ja, und dann kann ich nachfragen und oder. Ja, aber da habe ich ja schon Informationen und ich warte trocken drin. Da habe ich vorwiegend schon Informationen, was ich als Antwort beispielsweise geben könnte. Wer hat schon mehr Ahnung als ein Chatbot? Gut.

War hier jemand noch die Hand gehoben hat? Weil sonst gebe ich jetzt, also hinten. Gewissermaßen auf die Umwelt her, die ja also da hat ja gewisse Zugänge, oder?

Was jetzt die meisten Chats nicht haben, dann sind sie halt auch die, die mit reagieren teilen, da ist Kalender und so weiter und können dann immer eine Aufgabe stellen.

Das ist halt nicht nur die Aufgabe so, wie jetzt ein Sprachmodell, sondern berücksichtigt auch alle anderen Informationen, auf die ich Zugriff habe. Also Umwelt spielt eine Rolle. Guter Punkt. Soll man wirklich, kann ich nachfragen.

Das ist immer wieder mit ChatGPT. Da gibt es schon eine Antwort, bis er irgendwann mal dann an der Scharfe sagt, sorry, ich kann es leider nicht. Aber oft versucht es so oft eine Antwort zu geben und verschwendet eigentlich so oft Zeit.

Und bin ein spezialisierterer Keil, würde ich vielleicht ein Resultat kriegen, wo die Antwort eigentlich schon abdecken würde. Also hier geht es auf die Spezialisierung, wenn er von Agenten spricht. Okay, gut.

Ja, er kann nicht Aufgaben erledigen, indem er auf verschiedene Ressourcen zugreifen kann, in verschiedene Schnittstellen. Das heißt, das ist eine externe Ressource, eine Umwelt.

Und das Ganze dann für mich verarbeiten und mir ein Resultat geben. Aber einerseits zum Beispiel auf Sprachmodelle, aber auch auf andere Schnittstellen, die er nutzen kann und mich da auch nicht nutzen kann. Und kann auf andere Informationen zugreifen, die er dann verarbeitet. Okay.

Wo kommt das Wort Agent eigentlich her?

Schönes Wort, ja.

Ja, aber da ist auch Agent drin. Aber was steckt hinter dem Wort Agent?

Niemand Griechisch. Lateinisch. Agieren, ne? Du hast reagieren gesagt. Agieren. Das ist ein Agent. Agent hat nichts mit Spionen zu tun oder sowas, sondern ist ein etwas, was agieren kann, was was tun kann.

Und was bedeutet zu agieren? Wenn wir uns sehen, wie wir agieren, ja, wir können etwas wahrnehmen, was aufnehmen, dann entscheiden wir irgendetwas und dann tun wir etwas.

Also das Agieren, etwas auch zu bewegen, nicht unbedingt physisch zu bewegen, aber etwas zu tun, in gewissem Sinne, das gehört zu einem Agentenwort. Ja.

Ja, ich hätte es jetzt zurückgefragt, aber in welchem Sinne ist es ja agieren? Ja, selbstfahrendes Auto ist ein Agent, spezialisiert darauf, mit dem Auto zu fahren. Und jetzt wollen wir sagen, es reagiert in der Umwelt, hat einen Input, was macht es?

Das ist eine Trauma, also ist eine Umgebung, die Trauma, wie zum Beispiel Rotlicht, ich glaube, die hat Sensoren. Genau, es hat Sensoren. Kameras. Verschiedene Arten von Sensoren, Kamera ist auch ein Sensor, also nimmt die Umgebung wahr und reagiert darauf, berechnet also, wie es fahren kann.

Muss es bremsen, muss es ausweichen, kann es beschleunigen, ja? Und das ist dann der Output. Das Auto bewegt sich dann aufgrund von dem. Also ein selbstfahrendes Auto ist ein Agent. Das ist jedenfalls so eine symbolische Keilung.

Das sind klare Regeln definiert, weil sonst wird interaktiv auf das selber arbeitende Regelmacher gedacht. Das ist ein guter Punkt. Es ist nur symbolisch, es ist ein selbstfahrendes Auto?

Es muss sich selbstfahrend. Wieso? Es ist ein Megakomplex.

Also ja, wenn ich am letzten Mal, glaube ich, Frühstück mache mit dem, ich bin ja das Kind, das Kind, das Kind, vergleichbar mit, ja, ein Taschen oder so, ja, und sieht es um den Schein, ob es jetzt wirklich das Kind ist oder nicht das Zeck, was auch immer.

Ja, eine kleine Zusatzfrage. Ich habe zum Beispiel eine Heizungssteuerung ohne einen Agenten. Eine Heizungssteuerung, also Thermostat. Was?

Ohne Keil, einfach wirklich so Hand, wie Metallmasse. Ja, dann ist du, Temperatur, Sinn, also irgendwo und Anforderungen, die irgendwo verloren.

Ja, und agieren, also reagieren auf Temperaturen, zum Beispiel. Das wäre ganz einfach. Also Agent muss nicht KI sein. Das ist schon mal das, was da rauskommt. Agent reagiert auf die Umwelt und beeinflusst die Umwelt. Ja, und jetzt nochmal zum Symbolischen und Subsymbolischen bei einem Auto.

Also eigentlich gibt es zwei Sachen. Rein symbolisch zu bauen, das hat man versucht in den 80er Jahren. Ich weiß, da hat Mercedes so ein Testfeld gehabt auf dem Flughafen, wo sie mit dem Auto gefahren sind. Straßenrand erkennen und so an der weißen Linie. Und das ist unheimlich komplex.

Bildererkennung, ist das ein Hindernis oder nicht? Und das kann man heutzutage eher mit Subsymbolischen Sachen machen. Also aufgrund von vielen Sensordaten, die verschiedene Sachen abdecken, kann man das steuern. Ist das ein Hindernis oder ist das kein Hindernis? Muss ich ausweichen oder nicht?

Kann ich beschleunigen? Wie stark muss ich beschleunigen? Das muss ich nicht als Regel machen. Vielleicht dort als Beispiel. Wenn ich neben dran hoch kann, neben dem Fahrer. Und ich mit Symbol, muss ich sagen, wenn der Fahrer soll fahren, ist es ziemlich mühsam und komplex, oder?

Sagen wir mal links, links, rechts, also eine Fahrlehrersituation. Wenn ich aber sage, ich fahre. So weit wollte ich nicht gehen. Genau. Stopp. Gar nicht so schnell. Genau, oder? Ich parkiere mit Rehenfahren, wenn wir über das Thema sind, ist es ein Erlebnis.

Das ist relativ schwierig, das zu sagen, oder? Aber wenn ich es analog kann, sagen, also im Sinne von wirklich steuern, ist es viel einfacher. Und jetzt sind wir subsymbolisch. Gibt es auch symbolische Anteile? Ja. Das Regeln, ob es jetzt rot ist oder grün.

Und das sind aber auch Spurenhaltssysteme, ich darf nicht aus der Spur gehen. Ja, bestimmte Sachen sind sehr wahrscheinlich symbolisch oder müssten es eigentlich auch sein, weil die Verkehrsregeln, die muss ich berücksichtigen.

Die Schilder muss ich brandwarnen, nicht nur, okay, bei Rot halte ich zu 99 Prozent an, aber bei 1 Prozent überfahre ich es, weil beim Training habe ich bestimmte Situationen nie gesehen oder so. Also Regeln, die müssen auch von einem Tag auf den nächsten geändert werden können, Verkehrsregeln, wenn sich da was ändert.

Also von daher gibt es auch symbolische Anteile, die das subsymbolische steuern oder den Input von subsymbolischen bearbeiten. Aber dann würde das symbolische, das subsymbolische übersteuern, oder?

Bei Level, zum Beispiel bei einem Unfall, oder der Verhinderung von einem Unfall, jemand läuft durch die Straße, er könnte es verhindern, wenn er nach rechts geht, aber dann würde er eigentlich mit der Verkehrsregel schützen. Ja, deshalb glaube ich nicht, dass eines, wir wissen nicht genau, wie es funktioniert, aber ich glaube nicht, dass eines das andere steuert.

Das ist eine Interaktion zwischen verschiedenen Komponenten. Ihr müsst euch auch vorstellen, wenn ich die verschiedenen Sensoren habe, da habe ich die Abstandsmesser, da habe ich die Kamera, es sind ja verschiedene Sachen, die zusammenspielen und die dann auch wieder beeinflussen, wie das Auto reagiert.

Ich gehe eher davon aus, dass es ein Zusammenspiel zwischen verschiedenen Komponenten ist. Vielleicht noch ein schönes Beispiel, wie Pilatus, wenn ich das Projekt habe, wo es darum gegangen ist, das weiß ich nicht Cockpit einzurichten. Das heißt, nicht einfach nur noch Bildschirme, ich kann zum Teil Overvoice Sachen machen.

Und dort ist es zum Beispiel so, wenn der Autopilot eingeschaltet ist, kann ich sagen, go descent oder irgendwie stall und zurück. Stall bedeutet, wenn die Strömung fast abreißt im Flugzeug. Und das wird immer hin und her übersetzt.

Das heißt, wenn ich sage, go descent, dann gehe ich in Sinkflug, der Pilot weiß es schon, dann gehe ich in seinem Traffic Computer, dann nachschauen, was das bedeutet und werde es dann analog steuern, also subsymbolisch.

Hingegen, wenn er mir zurückmeldet, hey, stall, da sind ganz, ganz viele Sensoren zwischen uns, werden Machine Learning ausgewertet, dann plötzlich ist die Konfiguration so, dass es heißt, ach nein, es ist gefährlich, das Flugzeug ist zu langsam. Und dann geht es zurück, stall.

Das ist ja ziemlich mühsam, das ist ein ganz lautes Signal, wenn ich sage, stall, stall, stall. Dann muss man den Hebel ziehen, sodass man wieder in den Lack kommt.

Und um zu übersteuern, wenn der Pilot nicht reagiert und stall kommt, dann hat er vielleicht 15 Sekunden Zeit, um zu reagieren, dann geht der Autopilot rein und übersteuert. Also, wie du gesagt hast, es ist ein sehr komplexes, ehrliches System.

Weil schon mal eine automatische Spur, so ein Auto gefahren hat, ist halt, wenn man ausschmieg, mühsam, aber der wird nicht mehr so wie ein Pingpongball auf der Spur gehalten, mindestens weiter. Ich kann es auch übersteuern, aber das kommt auf das System drauf.

Also in dem Fall würde es eigentlich zuerst den Menschen fragen, hey, übernimm du und wenn du es nicht übernimmst, dann übernimmst du das System. Ja, das ist ein ganz großes Thema, vor allem im Reinforcement Learning, wenn ein Dispatcher einen Traffic Controller hat oder eben ein Fahrer, ein Pilot, wer steuert wann was? Und jetzt sind wir bei Rehenfragen, oder?

Sie fährt so, ich werde so und umgekehrt, sie hat recht und ich sage irgendetwas anderes oder umgekehrt. Wer hat Kontrolle? Und das ist momentan ein aktuelles Forschungsthema. Ja, das ist aber ein sehr, sorry, ja, sehr trauriges Beispiel.

Kai, ich meine, das ist einfach nur so eine Plastik, die sie dort drin haben und ich glaube, das Kind hat da irgendwie so ein Regenmäntelchen angehabt und ich komme ja, also Plastik kann ich verfahren.

Als Jugendabau. Röntgenbilder.

In einem Röntgenassistenten in einer Nähe, das ist eine Nachbarin, die hat eine riesen Brust. Und zwar, der Automat, der ist für Kai viel, viel, viel besser, für Kardinalen oder für Melanome zu erkennen auf Röntgenbildern. Was sie noch macht, ist, schnell zu überprüfen, ob Kai das richtig gemacht hat.

Und bei den Piloten weiß ich, von Airbus weiß ich, dass, wenn man auswertet, wie gut ein Pilot landet, wie gut die Autopiloten inzwischen landen, kann man Piloten daheim lassen, die stören.

Aber würde ich dort in ein Flugzeug gehen, das automatisch fliegt,

dann kann man noch keine neuen Sachen einbauen oder irgendwie mechanisch etwas machen.

Allein ist schon noch einfach wenig. Ah, was? Das ist ein aktuelles Projekt, wo wir gerade Flight Machine Learning gehen, das ist mit GESAC zusammen, mit der französischen Firma Enaketti.

Wenn ein Pilot, zum Beispiel schon Pilates dabei, der rote Knopf, wenn der Pilot die Übersicht verliert oder merkt, Achtung, Herzinfarkt, drückt er drauf. Und dann wird das Flugzeug schauen, wo die nächsten Flughäfen sind oder Flugfelder oder irgendwie Gras.

Und wird selber entscheiden und landen, absolut automatisch. Wenn der Pilot, so wie ein toter Mann, ich glaube, bei der SBB, der fährt, logisch so, wenn er kein Signal gibt, dass er keine Airbags hat, dann passiert das automatisch. Und dann, das ist sicher, also die Landung ist sicher.

Und das ist lokal, also das ist wirklich ein kleiner Rechner, der ist gut trainiert und da muss ich, wenn ich trainiere, brauche ich viel Rechenzeit und Energie. Sobald ich das habe, brauche ich nicht mehr so viel und dann sind es so eine große Maschinen, da kann man das lokal machen. Ist das dann deadness? Ja.

Aber wie schnell das Fließwind der Verbindung an den Hochpunktumgang, der Punkt kann man nicht so schnell abwandeln, so wie ich das weiß. Genau, darum sind sie lokal. Also die Autopiloten laufen völlig lokal. Das ist so eine kleine Kiste, das ist nicht groß.

Also das könnte irgendwie, wer gamen mit, nimmt irgendwie eine gute Gamerkarte und dann läuft es drauf. Das RTX 2080 ist drauf, die muss auch etwas, das ist wirklich, das ist überschaubar. Nochmal zurück zum Auto, du hast gesagt, dann überlebt der Mensch, also das funktioniert so nicht.

Da gibt es auch Sachen, wo man Mensch, ja, wir haben ja die Regel, man muss die Hand am Steuer haben oder so, wenn man den Autopilot einschaltet, oder? Da gibt es auch Unfälle, weil man dann

doch nicht konzentriert ist, dann doch aufs Handy guckt oder uns schweift die Gedanken ab. Wer ist denn schon mal mit einem selbstfahrenden Auto gefahren?

Ihr habt ja Erfahrung, weil es gerade in San Francisco mit dem Waymo gefahren ist, da ist keiner mehr, der die Hand am Steuer hält. Das ist eine total faszinierende Sache, du fährst mit dem Auto und es ist kein Fahrer. Oder es kommt dir ein Auto entgegen, da ist gar kein Fahrer drin. Und die Dinger funktionieren. Und da ist keiner, der die Kontrolle übernehmen würde.

Wir haben schon eine Überwachung, ja, aber es ist nicht für jedes Auto dediziert, ein Fahrer. Und die müssen ja auch schnell reagieren, wenn jemand über die Straße rennt, dann muss das Ding bremsen. Und das war auch so ein Spaß für einige Leute, als die angefahren sind, angefangen haben mit dem Waymo.

Und die wussten, die müssten sicher sein, sind die Leute hingegangen, haben sich vor die Autos gestellt und dann die angehalten. Oder so ein Pylone auch vorne auf das Auto drauf gesetzt und so. Und das ist nicht so, dass dann jemand noch hinten reagieren kann. Nein, die Dinger, die müssen bremsen.

Und die haben volle Kontrolle über das Ganze. Also ich fand es fantastisch, weil, und wo du gesagt hast, würde ich ja mit dem Flugzeug fliegen, das keinen Piloten hat, es ist doch relativ überschaubar, wie der Verkehr im Luftraum ist. Der Verkehr in einer Stadt wie San Francisco ist nicht mehr ganz so überschaubar.

Und Cable Cars hast du, ich bin gefahren und habe plötzlich jemandem im Auto die Tür aufgemacht und das System hat reagiert, ist schön ausgewichen und so, das war super. Das heißt jetzt mit dem Cloud-Bahn-VW und VW rein, da geht schon auch genug Abstand.

Normalerweise würde ich vielleicht Rolle lassen oder langsam bremsen, aber wenn wir das System, das ebenfalls machen, oder würdest du dann sagen, okay, jetzt muss ich bei der Mindestabstandsstelle gehen? Nein, das fährt ganz smooth. Also, und also eine fantastische Situation, die ich erlebt habe, Fußgängerüberweg.

Ganz viele Leute laufen über den Fußgängerüberweg. So, was würdet ihr als Mensch machen, wenn immer mehr Leute kommen, immer mehr Leute kommen? Warten, okay, oder mal so ein klein bisschen vorfahren, irgendwann lässt uns einer durch, oder? So, hat das Auto, das selbstfahrende Auto gemacht.

Klein Stückchen vor, klein Stückchen vor, und irgendwann haben die Leute angehalten und haben das Auto fahren lassen. Also, diese Situation fand ich fantastisch, dass es wirklich ähnlich reagiert wie wir Menschen auch, aber das fährt ganz gut, nicht plötzlich bremsen oder so, eigentlich viel sanfter als von Autofahrern.

Auf der Dampfer machen, haben wir ein anderes Problem, und zwar, wer kontrolliert schlussendlich, wenn wir automatische Maschinen haben? Wir sind in einem Forschungsprojekt und genau das, so Human in the Loop Sachen, die untersuchen auch mit Psychologen zusammen. Natürlich wollen wir 20 selbstfahrende Autos überwachen.

Es ist mega langweilig, bis es dort passiert. Bis man checkt, oh, was ist da? Also, das ist nicht, aus meiner Sicht, ein ungelöstes Problem. Wie wird überwacht? Menschen. Wenn die Systeme viel, viel besser sind als Menschen, kann man es blass sein.

Wer will AI-Recht studieren, hat eine schöne Zukunft, oder? Wer ist der Schuld? Der Ingenieur? Der Käufer?

Ich schlage vor, wir machen eine Pause bis Viertel nach elf, und dann machen wir weiter. Ich habe dann noch was.

Wir richten das am Cluster ein, das ist DeepSeek 5.3.8, ein neues Chip.

Okay, wir haben ein bisschen gesprochen, was Agenten sind. Ja, ich habe hier noch eine Folie, wo ein paar Sachen dazu aufgeschrieben sind.

Der Agentenbegriff, warum man den verwendet, aber wichtig: Agenten haben Sensoren und Aktuatoren, heißt glaube ich auch Aktuatoren, also nehmen die Umwelt auf, können auf die Umwelt eben auch reagieren und sie beeinflussen aufgrund von ihrem internen Verhalten.

Und sie müssen wahrnehmen können und müssen Aktionen ausführen können. Das gehört zu einem Agenten dazu. Und das steht noch nicht, dass es KI sein muss, aber man muss die Umwelt interpretieren und auf die Umwelt einwirken. Das ist das, was ein Agent machen kann. Und dann gibt es ein Modell.

Ein Modell beschreibt das Verhalten eines Agenten. Also wenn man den Input kriegt, dann wird das mit dem, was wir als Modell bezeichnen, verbunden und das Modell sagt dann, was ist der Output. Und Modelle, wir kennen Sprachmodelle, ja, das ist ein Modell. Ein Modell kann auch eine Wissensbasis sein.

Das ist einfach eine Beschreibung des Wissens über die Realität, die den Input kriegt und dann sagt, wie man wie das der Agent reagieren soll.

Ja, und ja, Agentenprogramme, Agenten in der KI, da geht es darum, dass wir dort die Ressourcen, die wir als Rechenressourcen haben, benutzen, um nicht triviale, komplexe Aufgaben zu machen.

Und wir haben mal hier so ein Bild, ja, wir haben den Agenten und wir haben die Umgebung in der Akte. So, und ein Agent soll autonom auf die Umgebung reagieren können, also nicht, dass wir ihn steuern, sondern er reagiert auf die Umgebung. Deshalb ist das Beispiel, das wir hatten, mit dem Auto, war ein schönes Beispiel.

Es ist ein Agent, das Auto reagiert selbstständig, wir haben nicht mehr den Fahrer. So, und was braucht man dafür? Man muss sich erstmal die Umgebung wahrnehmen. Also wir haben hier in diesem Fall, mit dem Auto hätten wir Sensoren, aber Wahrnehmung ist ein Teil von dem Agenten, sonst kann man nicht auf die Umwelt reagieren, wenn man die Umwelt nicht kennt.

Ja, also muss die Umwelt aufnehmen können, also wir haben einen Input, die haben wir nachher im System, das können Daten sein, das kann eine Anfrage sein, irgendetwas, was wir aus der Umwelt kriegen. Und dann kommt unser Modell. Das Modell ist das, was nachher die Entscheidung trifft, die Aktion bestimmt.

Wir können das als Riesen bezeichnen, wir können das als Berechnung bezeichnen, Matrizenrechnung, was immer das ist, aber das Arbeiten. Und dann produziert es den Output. Also die Aktuatoren werden aktiviert und die reagieren dann in der Umwelt.

Ja, das, was hier bunt ist, das ist der Agent und wir hatten eine Interaktion mit der Umwelt, aber da ist etwas, was mit der Eingabe passiert und da ist etwas, was bestimmt, was mit der Ausgabe

passieren soll. Also beim Auto wäre das zu bremsen, zu steuern, anzufahren und so weiter. Das wären die Aktuatoren, die das System sagt und dann hat es eine Auswirkung auf die Realität.

Und wir können die KI als Agenten bezeichnen. Jetzt haben wir heute diesen Begriff Agentive AI, ja, in meinen Augen auch ein bisschen Missbrauch von dem Wort Agenten. Ich glaube, du hast am Anfang das als Beispiel auch gebracht.

Ja, jetzt haben wir den Agentenbegriff, wir fokussieren ihn eigentlich nur auf Sprachmodelle, auf generative KI, aber Agentensysteme sind weit mehr als nur mit Sprachmodellen zu arbeiten oder generative KI zu haben.

Genau zusammen, das ist die Struktur, das so würde ich mir merken. Also wenn jemand fragt, was ist ein Agent, würde das da eigentlich als Antwort erwarten. Umwelt, Agent, Input und Aktion.

Das wird überall gebraucht, was wir jetzt gerade gesehen haben.

Und Mensch ist auch ein Agent. Wir agieren in unserer Umwelt. Die Wahrnehmung machen wir mit unseren Sinnen. Nicht Sensoren, sondern wir hören, wir sehen, wir riechen, ja, also alles, was wir fühlen, alles das sind unsere Sensoren.

Das nehmen wir auf, aufgrund von dem arbeitet quasi unser Gehirn. Und dann reagieren wir darauf. Das ist unser Modell, das ist das Gehirn. Und wir haben ein Modell der Umwelt. Wir wissen, was das bedeutet, dass da ein Tisch steht, dass ich mich da festhalten kann, dass ich den rumschieben kann und alles. Das ist das interne Modell.

Und wir nennen es dann ein mentales Modell. Das ist in unserem Gehirn. Also das Gehirn ist nicht das Modell. In unserem Gehirn haben wir ein Modell der Realität, das uns sagt, wie wir agieren können in einer bestimmten Situation. Bewusst oder unbewusst, das spielt hier nicht die Rolle, aber es ist ein mentales Modell, das wir haben.

Und ja, was habe ich gesagt? Es ist ein gedankliches Abbild. Also Autos wären ähnlich, dass wir wissen, wie ein Auto funktioniert, was ein Auto ist, wie wir es bewegen können. Das funktioniert nur, weil wir wissen, weil wir das als mentales Modell haben. Deshalb können wir überhaupt ein Auto steuern.

Wenn wir das nicht wüssten, können wir gar nicht voraussehen, was die Auswirkung einer Reaktion wäre, wenn wir kein internes Modell hätten. Das ist also nicht nur, dass wir einfach steuern, sondern wir haben, wir wissen, wenn wir bremsen, halten wir an. Wenn ein Hindernis kommt, wir müssen bremsen, das bedeutet, wir überfahren das Kind, nicht um bei dem Beispiel von vorhin zu sein.

Vielleicht auch noch etwas. Also die Ansicht, wenn ich kein inneres Modell habe, kann ich nicht planen. Ich kann nicht voraussehen, was soll sein. Also ich brauche ein inneres Modell, das die Umgebung einigermaßen darstellt. Und wenn ich so ein inneres Modell habe, kann ich mich selber auch in das innere Modell hineintun.

Das ist ziemlich nahe am Selbstbewusstsein, selbstbewusst als Akteur selber drin sein.

Gut, wir passen uns auch an, das hängt nachher mit dem Lernen zusammen. Aufgrund von Erfahrungen wissen wir, wie wir reagieren können. Da habe ich das Beispiel vom Fahrlehrer gebracht. Wenn ich das wieder bei dem Auto schreibe, wir lernen, wie wir fahren sollen, der korrigiert uns, hilft uns dann auch, das mentale Modell zu entwickeln. Und auch danach geht es weiter.

Also Fahren lernt man ja eigentlich nicht in der Fahrschule, sondern nachher, weil man fährt, weil man weiß, wie man mit einer bestimmten Situation umgehen muss. Man passt sein mentales Modell an. Und dann haben wir nachher die Aktionen. Ja, das ist, was wir mit den Händen machen. Auch Reden ist eine Aktion.

Ja, also wir haben Aktuatoren, die wir steuern mit unseren mentalen Modellen. Und dadurch verändern wir wieder die Umwelt, haben einen Einfluss auf die Umwelt. Und das auch in verschiedener Art. Beim Auto wäre das, dass man fahren kann, dass man auch bestimmte Sachen erkennen muss.

Ich habe das mit dem Tanken, ja, also wir je nachdem tanken wir oder oder laden unser Auto. Das gehört mit zum Fahren dazu. Das wissen wir, wenn wir erkennen, okay, der Tank oder die Anzeige geht runter. Wir müssen auf das reagieren. Aber es sind auch andere Einflüsse, die wir haben.

Also wenn wir jetzt hier unterrichten, haben wir auch einen Einfluss auf euch hoffentlich, dass ihr euer mentales Modell anpasst. Das ist auch ein Einfluss auf die Umwelt. Es ist nicht nur direkt, kann auch indirekt sein. Michael, wenn man Mathematik macht im Kopf, kann man etwas Neues lernen, rein im Modell selber.

Ist das intern auch ein kleiner Agent? Gedächtnis.

Also wenn ich ein mentales Modell habe von meiner Umgebung, von meinem Universum, ich denke mir selber etwas aus, in einem Traum, in einer Geschichte oder in einer Rechnung, kann man das als Agent bezeichnen? Wirklich schärfen, was ein Agent ist, oder? Begriff.

Hat man einen Agenten in einem Agent?

Ja, also ich glaube, der Agent braucht ja das Equipment, um eine Aktion oder eine Aufgabe zu bringen. Es braucht eine Umgebung, oder? Das ist das Hirn, natürlich kommunizieren. Wenn man sagt, das Gehirn ist ein Teil vom Hirn, dann würde es ja.

Aber ich sage, ja, wenn nicht, wenn das separiert wird, dann nicht. Ja, also ich komme mit mir selber reden. Es gibt auch Wandelhandel in verschiedenen Universitäten. Da siehst du manchmal so Leute, bla bla bla bla, und reden mit sich selber. Ich habe das am Gedanken machen, oder wenn ich Schach spiele oder irgendjemand etwas überlege.

Ich will auf etwas hinaus, oder ich will die Definition, was ein Agent ist, ursprünglich hinaus. Da müsste jetzt kein Politik kommen, wenn es ja nur intern bleibt, dann ist es ja kein Output. Ja, wirklich. Theoretisch, wenn es jetzt Traum wäre, dann wäre es ja quasi auch ein Output, oder nicht?

Ja, die Umgebung wäre das, was ich mir vorstelle im Traum. Ich wäre der Agent darin, oder der Mensch. Ja, vom Traum habe ich Informationen und könnte selber auch agieren. Dann wäre ich jetzt eigentlich auch noch ein Agent. Ich habe ein bisschen im Kopf, was ein Agent ist.

Nach der Definition Nummer 1. Wir haben eine Umgebung, nämlich meine Gedankenwelt. Wir haben jemanden, der agiert in meiner Gedankenwelt und ich habe Input von meiner Gedankenwelt. Es gibt Leute, die das Schachspiel im Kopf, die haben so ein Bild, die schaffen Tic-Tac-Toe, zum Beispiel Schachfile.

Stell dir vor, neun Felder, oben rechts ein Rundomalie, wo tut der Zick her? Mittig, ja, genau, Mittig, das kommt schnell, oder? Wo ist das Nächste? Oben rechts, oder wo? Oben links, oder schon Thumbnails, schnelles Denken, oder?

Okay, die machen was? Die sind selber, drehen das Modell, das ist die Umgebung, und die reagieren, die tun es ja drin, ändern den Zustand. Ist ein Agent.

Ja, das geht jetzt darum, auch was die Umgebung sein kann. Die kann real sein, die kann aber auch virtuell sein. Wenn du jetzt Spiele spielst, Gaming, das heißt eigentlich in der virtuellen Umgebung, auf die ja auch einen Einfluss ausübt. Und ja, in dem Denkmodell, das Wandel gebracht hat, könnte auch unsere interne Umgebung könnte die.

Unter interne Vorstellungen könnte die Umgebung sein, auf die wir dann Einfluss haben. Wenn du ein Beispiel weisst, wo es darum geht, Selbstbewusstsein als Agent zu sehen. Das muss kein visuelles Output sein, sondern es kann auch schon ein Input sein, wenn eine Berechnung stattfindet. Exakt, exakt.

Ich würde darauf hinweisen, dass es ein Schema ist, dass man das schematisch anschauen kann, wo ich die vier Sachen Input, Wahrnehmen, Agent, Umgebung und Aktion gleich sind.

Also höher gesagt, ein elektrischer Schalter
ist ein Agent.

Das ist jetzt aber schon ein bisschen weit hergekommen, oder? Darum habe ich es nicht gemacht. Genau. Aber was wichtig ist, ist, bei Agenten geht es auch nicht um die Größe. Wenn wir heute von Agent AI reden, dann haben wir es einen kleinen Agenten, den wir anschlagen. Ein Mensch ist ein Agent, das ist etwas Komplexes.

Wir können spezifische Agenten haben, die für ein Problem da sind, das Input bearbeiten, genau dieses Problem lösen und wieder den Output liefern. Also die Größe, was das Ding macht, das definiert nicht ein Agent. Ein Agent ist, wie Manuel gesagt, definiert durch die vier Sachen.

Wir haben eine Umgebung, wir können die Umgebung wahrnehmen, wir können eine Berechnung machen, ein Reasoning, was passieren soll, und wir können die Umgebung beeinflussen. Das sind die vier Sachen, die ein Agent ausmacht.

Ob das groß oder klein ist, die Umgebung, der Umfang der Aktion, das spielt keine Rolle, ob wir das als Agent bezeichnen wollen. Jetzt können wir uns mal, das war jetzt ein natürlicher Agent, der Mensch. Ja, wir können uns auch künstliche Agenten vorstellen.

Ja, dann, das ist die Art, wie man sich eigentlich ein KI-System vorstellen kann. Deshalb bringen wir das auch einfach. Im Prinzip, was wollen wir mit der KI machen? Wir wollen Agenten machen, die in der Umgebung etwas aufnehmen können, die intern ein Reasoning machen können und sagen, was ist ja das Ergebnis anschauen, dass man reagieren kann.

Und das Modell in diesem Fall ist dann etwas, was wir mit KI generieren. Das ist ein KI-Modell, das einen Input bearbeitet und einen Output generieren kann. Und das entspricht diesem mentalen Modell.

Diese Analogie mit den neuronalen Netzen haben wir schon gehabt, dass man das macht als Analogie, wie das menschliche Gehirn funktioniert. Das ist eine Möglichkeit. Wir können auch eine Wissensbasis haben da drin. Dann ist es ein KI-Modell, das das Wissen des Menschen, das bewusste Wissen für rationales Denken drin hat und dann reagiert.

Das ist einfach ein Modell, ist die Abbildung von einem Teil der Realität. Das ist das, was wir als KI-Modell bezeichnen. Und es ist eben automatisch verarbeitbar. Und was wir als Input dort haben, sind

dann in der Regel Sensordaten. Aber es können auch die Eingabe sein, die wir dann mit ChatGPT einfach eintilgen.

Das wäre dann auch der Input. Ja, das Modell leitet dann daraus, was ein Resultat hat und das Resultat ermöglicht dann die Aktion. Das kann auch wieder eine Sprachausgabe sein. Das kann aber auch, wenn wir einen Roboter haben, ein Arm sein, der sich bewegt, der etwas greift. Das kann das Auto sein, das sich bewegt.

Die Aktoren sind auch unterschiedlich, je nachdem, welchen Leistungsumfang sozusagen der Agent haben kann. Vielleicht auch noch, wenn ich den Kreislauf anschau, den kann ich brauchen, um den Agenten zu verbessern.

Oder er macht eine Aktion, dann kommt er wieder Informationen zurück und kann korrigieren. Er kann das innere Modell, das mentale Modell, kann er verbessern. Dem sagen wir Lernen. Und das wird gemacht, das heißt Reinforcement Learning oder selbstverstärkendes Lernen auf Deutsch. Das wäre genau das.

Also im Prinzip ist ein RL-Agent doppelt gemoppelt. Ein Agent kann selber sich verbessern. Das macht ihr ja selber auch, oder wenn ihr irgendwie etwas lernt. Ah, jetzt falsch machen, macht es irgendwie auch gut.

Das kann man beliebig weitertreiben, wenn man AGI will oder im Prinzip Modelle will, die sehr, sehr gut sind, kann ich die Umgebung so einstellen, dass sie immer mehr und mehr lernen. Das kann das Internet sein, das kann eine Umgebung sein, das kann ein Roboter sein.

Es braucht ja aber einen Feedback, bis es etwas Prozess von einem Künstler haben. Genau. Und der Feedback entsteht genau dann, wenn ich eine Aktion mache, die Umgebung anschau oder das Innere, das KI-Modell entsprechend ändern.

Wenn ich noch ein Zusatzfeedback gebe, ich kann jemandem sagen, ob er es gut gelernt hat oder nicht. Ich kann zusätzlich eine Note geben, ich kann sagen, du hast es gut gemacht. Ein kleines Kind zum Beispiel, wo du immer Feedback, oder habe ich es gut gemacht oder so.

Wenn wir jetzt ChatGPT zum Beispiel benutzen würden und wir alle drinnen würden, ChatGPT den gleichen Prompt geben und ergibt uns einen Output, der vielleicht der gleiche ist oder sollte der gleiche sein, dann könnten wir alle sagen, nein, das stimmt nicht, eigentlich so und so, und dann könnten wir das Modell überfordern. Exakt. Ist gemacht worden.

Man hat zuerst, ich glaube, das war Microsoft, haben ein paar Leute gefunden, da machen wir eine NATI aus dem LLM. Und der Feedback, also die einfach wirklich in die Richtung haben, die die Antwort gegeben haben, und das Modell hat sich adaptiert. Also man kann sehr duktil das Modell ändern.

Zum Teil, ChatGPT 5 hat, glaube ich, den Sample nicht, oder wo man so auf und ab machen kann. Ich kann nicht mehr Feedback geben, da bin ich mir ganz sicher. Aber andere Modelle, da kann ich sagen, hast du eine gute Antwort gegeben und so weiter. Das ist eine kleine Geschichte, nur nebenbei. Also, ja, es ist eben sehr mündlich, oder?

Zwischen, die immer den Daumen hoch gehabt. Das ist völlig sinnlos, weil wenn es immer gut ist, kann man nichts lernen, oder? Also hat man das ein bisschen gefiltert. Aber haben sie die Regeln gefiltert oder so, weil dann... Doch, doch.

Ja, also wenn ihr immer den Daumen hoch hebt, dann passiert das, ihr habt vorne weggenommen, dann werdet ihr einfach tiefer gewichtet.

Und die Antworten selber, die ihr gebt, werden ausgewertet, ob ihr welche Nachfolgefragen stellt und so weiter wird angeschaut.

Nochmal, der einzige Grund, warum das ein Reinforcement Agent ist, ist, wenn ich aufgrund der Umgebung, wenn ich eine Aktion mache, aufgrund von dem, mein innerer mentaler Modell gewechselt wird oder verbessert wird, geändert wird. Wenn es statisch ist, ist es einfach ein Agent.

Gut, ich komme nochmal zurück zu den verschiedenen Arten von Agenten, was vorhin gesagt hat, symbolisch und subsymbolisch.

Also wenn ich jetzt symbolische KI habe, ja, dann habe ich meinen Agenten, der in der Umgebung läuft, ich nehme etwas wahr, dann kommt hier meine Wissensbasis, die ich explizit symbolisch repräsentiert habe, aus der kann ich eine Handlung ableiten und die beeinflusst die Welt.

Ja, so, wir können mit der symbolischen KI einen Agenten bauen und wir können mit der subsymbolischen KI einen Agenten bauen. Dann ist zum Beispiel ein neuronales Netz oder ein statistisches Verfahren, aufgrund von dem wir die Eingaben verarbeiten und eine Ausgabe generieren, die den Output bekommt.

Also der Agentenbegriff ist nicht abhängig von der Art, welche Form unser Modell anwendet. Eine Frage. Beim subsymbolischen Modell lernt ja die KI, indem es ja die Richtlinien anpasst. Wie sieht es dann beim symbolischen Modell aus?

Wie zum Beispiel bei ELIZA ist es eigentlich einfach ein großes Switch-Case, wenn das Wort kommt, machst du das. Kommt dann der Feedback-Loop einfach nur vom Menschen aus, wenn er etwas anpasst? Ja, das Ganze setzt nicht ein Feedback voraus. Wir sagen, wir haben einen Agenten, der agiert.

Lernen ist dann etwa eine Meta-Ebene drüber, die sagt, wie passen wir unser Modell an. Im Prinzip gibt es das, aber wir können auch beim symbolischen System Lernen einbauen, aber es ist nicht eine Voraussetzung, damit etwas ein Agent ist, dass es sich auch selbst anpasst. Auch hier, wir können, wir trainieren das neuronale Netz.

Ja, so, nach dem Training kann man das rausgeben, sagen, okay, jetzt haben wir die Anwendung und trainieren es nicht mehr. Ja, und dann wäre das der Agent, den wir vorher außerhalb in der anderen Phase trainiert haben, und dann bringen wir ihn in die Anwendung. Das neuronale Netz muss ja nicht ständig angepasst werden.

Es kann auch sein, dass man es einmal trainiert und dann wird es verwendet. Das ist aber beim Menschen nicht so. Wir lernen durch. Wir können nicht nicht lernen. Es hat eine ganz, ganz blöde Eigenschaft.

Wenn die schönen Erinnerungen immer wieder ins Gedächtnis sollen, dann werden die mit der jetzigen Welt überschrieben. Das heißt, es ändert dauernd.

Darum ist es bei einem Gerichtsfall, wenn man x-mal befragt wird, wie soll da plötzlich grün, ob es schon das rote ist gewesen. Ich habe das so suggestiv erfragt, oder? Also die Erinnerung, die ihr habt, wenn ihr die vielmals abfragt, die ändert.

Da kann man die Roma da im Prinzip ausbügeln ein bisschen. Ja, und so kann man auch die Aussagen von Leuten beeinflussen, wenn man dem vorher redet, ihnen was sagt und dann plötzlich ist es beeinflusst. Neuer Vergangeneheit ist nicht so geschehen, wie ihr es im Kopf habt.

Was ist die Farbe von der Tomate?

Klingt aber ein schönes Beispiel, oder? Wer kennt das? Fragt mal. Wie ist die Farbe von der Tomate? Was ist die Farbe von Rot? Bei welcher Farbe darfst du mit der über die von der Ampel darfst du über die Straße fahren? Die Leute sagen Rot. Weil du vorher sehr viele Sachen über Rot gefragt hast, so kann man die Leute beeinflussen.

Gut, okay. Weitere Art von Agenten. Ja, wir haben gesagt, Agenten können auch interagieren. Also, was ist ein Roboter? Ein Roboter hat seine Objekte in der Umwelt, die er wahrnehmen muss. Ja, und das ist über die Kameras zum Beispiel, wenn wir an Shake denken.

Wir können die Elemente identifizieren. Können wir auch über eine Position machen, wo die Sachen sind. Aber das wäre eine Variante. Wenn er selbstständig ist, hat er eine Kamera. Dann hat man die Recheneinheit, die das berechnet, was nachher passieren kann.

Und dann Motoren, die die Steuerung machen von den Armen, die den Roboter irgendwo hinfahren, was aufheben und so weiter. Und wir haben hier Recheneinheiten oder Code geschrieben, weil es gibt viele Roboter, die haben nichts mit KI zu tun.

Roboter, die in der Fabrik arbeiten und die die Werkzeuge oder die suchen und die Objekte bewegen, etwas einbauen und so weiter. Die haben zum großen Teil gar keine Intelligenz drin, in dem Sinne, was wir sagen, ein mentales Modell, das ist einfach Code.

Aber die können auch die Umgebung wahrnehmen, können die Umgebung beeinflussen. Und damit sind sie Agenten. Was haben wir noch für ein Beispiel? Ein Chatbot. Ja, das haben wir ja vorhin schon gesagt. Wir haben die Benutzer, der Benutzerin, die mit dem Interagiert.

Die Wahrnehmung ist einfach der Text, den man empfängt, den das System empfängt. So, das ist der Input, den muss man interpretieren. Dann haben wir das Sprachmodell, das die Interpretation macht und eine Antwort ableitet, die dann nachher ausgegeben wird und die dann Einfluss hat auf das, was der Benutzer oder die Benutzerin empfängt.

Ist Elsa ein Modell? Eine Agent? Elsa? Wie heißt sie? Elisa.

Sehr gut. Die Agentin.

Okay, also warum wir das gefragt haben, ist, dass dieser Begriff von einem Agent auf viele, viele Sachen passt. Als wir mal den Studiengang definiert haben, war unsere Vision ein intelligentes Unternehmen. Und das können wir auch als einen Agenten bezeichnen. Was wäre der Input für das Unternehmen?

Der Input wäre Informationen über Kunden, über den Markt. Ja, dann reagiert das Unternehmen und bringt Produkte auf den Markt, verkauft Produkte. Und das kann man auch mit KI unterstützen.

Also auch in der Organisation könnte wie ein Agent aufgefasst werden, das auf die Umgebung reagiert, indem es dann Entscheidungen trifft und dann wieder auf die Umgebung zurückspielt. Das heißt, die Umwelt ist der Input und Output beeinflusst später die Umwelt.

Genau, das ist ja, stimmt. Und die Umwelt kann, kann man jetzt mehr oder weniger groß fassen. Betrachten wir nur die die Wirtschaftswelt, dann gucken wir den Markt an. Wir können die Umwelt

auch breiter fassen. Dann hätten wir das, was wir so im Umgangssprach auf das Umwelt bezeichnen, ja, Sustainability und so.

Dann ist eine breitere Sicht auf die Umwelt. Ja, aber ja, die Umwelt ist im Prinzip für ein Unternehmen das, was wir als, was es beeinflusst. Man kann Unternehmen eben auch als einen Agenten sehen. Also noch weiter, als es vorher gesagt hat, ein Mensch, ein Agent, und Unternehmen wäre auch ein Agent.

Gut, und wir werden auch das immer mal wieder bringen, den Begriff von einem Agenten, wenn wir verschiedene Arten von Anwendungen haben. In diesem Fall ist der Agent schon so, was ist das Modell in diesem Fall? Was ist die Eingabe? Was ist die Ausgabe oder Input, Output in den verschiedenen Fällen? Mit der Erinnerung von einem Agenten.

Normalerweise wird in aktueller Forschung von Multi-Agenten-Systemen gesprochen. Wenn ihr schaut, wie DeepSeek oder ChatGPT 5 funktioniert, sieht man mehrere Agenten, die konkurrieren, zum Beispiel während die beste Antwort.

Oder wir machen Air Traffic Control-Simulationen, wo jedes Flugzeug ein Agent ist, der selber agiert. Sie sind in Konkurrenz zueinander.

Also, wenn ich drei oder vier Agenten habe, dann in Konkurrenz, dann werden die immer besser, weil die challengen sich selber, die gibt es jetzt die gleiche Umgebung, und die können daran nochmal spielen. Oder es gibt kooperative Agenten, die kennt man aus der Logistik. Systeme, die ein Auto ausladen.

Mehrere Agenten, die sich absprechen, und ja, wenn man holt etwas aus, ein Agent holt etwas aus dem Auto raus, bringt das ins Lager, der Nächste kommt, holt das Nächste.

Man kann auch dann Agenten haben, die spezialisiert sind auf bestimmte Waren, die sie transportieren können, leichte oder schwere, und die kooperieren, um ein Problem zu lösen. Also, man kann es über Konkurrenzieren machen, damit man sich challenged, besser wird.

Man kann es über Kooperation machen, weil kein Agent alleine in der Lage ist, alle Aufgaben zu machen. Es müssen sie sich absprechen, wer welche, in unserem Fall, das ich gerade gesagt habe, wer welche Produkte aus dem Auto.

Gut, wie gesagt, das war so eine Art, wie wir auf KI-Systeme schauen können, wie sie als Agenten zu sehen. Haben wir noch ein paar Folien zu, ja, was ist eigentlich KI, KI heute, KI morgen, um so etwas dazu zu sagen? Ja, vielleicht kann ich euch etwas zeigen, wo mich auf der Socke geholt hat.

Und zwar habe ich mit DeepSeek eine Unterhaltung gehabt, und ich würde euch die schnell zeigen. So als Einstieg. Die Frage ist, ist es wirklich intelligent oder nicht? Und ich probiere das schnell dazu zu schalten, oder so.

Ja, würde gerade ein bisschen ins System drücken, wir können sharen, ich zeige es dann hier, oder? Gut. Und zwar, das ist zufällig passiert, auf dem Weg im Zug nach Basel.

Okay, genau. Schema.

Ich weiß nicht, was es da Schemes gibt. Ja, doch. Genau.

Und zwar habe ich auf Deutsch interagiert.

Und ich habe, seht ihr das irgendwo? Schon schließe ich es vor. Dann mache ich weiter runter. Oben nachher sage ich, irgendwie vielen Dank so und so weiter. Und sage ich super, vielen Dank.

Ja, per Zufall habe ich mit dem Banglüdschen, habe ich jetzt einfach super vielen Dank geschrieben. Und so. Dann kommt zurück Gengsche. Das ist ein chinesisches Modell. Falls du noch mehr Hilfe brauchst, sage ich und so weiter. Und dann haben wir uns über Banglüdschen unterhalten.

Es ist eine knuddelig Sprache und so weiter, auf Banglüdsch. Gerade auch in Zürich-Deutsch gibt es vielleicht andere Antworten. Und dann, was extrem seltsam ist, wo ich nochmal anschau, ist folgendes. Das Modell sagt mir, hey, hilf mir Banglüdschen lernen, gib mir mehr Informationen.

Genau, und das wäre eine Frage an euch fürs nächste Mal. Aber was ich hier, was mir so Socken holt, was ein bisschen Gänsehaut gibt, ist, das Modell fragt mich und will lernen. Das verstehe ich noch nicht. Weil das ist dann auch, wenn man LLMs anschaut, wie das überhaupt funktioniert.

Und das ist etwas, was ich nicht verstehe. Das ist einfach ein bisschen ein Einstieg. Ja, woher geht es? Bei Basel-Deutsch kann man in eine andere Sprache gehen, kann man das Gleiche probieren, was passiert. Aber hier der Punkt ist, das System fragt mich und will lernen.

Also das ist in, das will etwas machen, das hat ein Ziel. Oder das ist das, was mich so ein bisschen komisch tut. Und das passiert auch mit Agenten. Also es sind so Reasoning-Agenten, die gegeneinander konkurrenzieren. Und jetzt ist die Frage, was da passiert. Ja.

Aus dem Text, weil ich nicht glaube, das Wort lernen oder ob es dir einfach durch den Anschluss geben will, dass irgendeine Kommunikation stattfindet oder irgendeine Aktion oder so. Genau. Und genau das möchte ich gerne wissen.

Oder ob das irgendwie das chinesische Zimmer ist, wo einfach die kleinen Chinesen da irgendwie die richtige Antwort sowieso geben. Oder ob dort in der Influence of the Causal irgendetwas dahinter ist. Weil ich mache ein Reasoning, oder dahinter. Das ist eigentlich das Einfache. Ich habe keine Ahnung noch.

Aber ich gehe gerade schauen, dass wir jetzt lokal, also wir schauen das einfach so gewichtgenau an und wollen mal verstehen, warum das das macht. Vielleicht ist es aber genau das. Vielleicht ist es einfach eine richtige Antwort. Aber das ist das erste Mal in meinem Leben, wo ich ins LLM mich etwas frage. Ich frage ich ja immer. Einfach nur so als aktuelle Forschung.

Aber der ChatGPT, der ist jetzt auch verantwortlich irgendwann, wo du ein bisschen mehr über, ja, über die Verhaltung, über die Betroffenheit und über etwas... Genau. Dort weiß ich, wie es programmiert ist. Aber das Modell fragt mich, ich will lernen über das, was du weißt.

Das ist etwas anderes, als was du noch mehr wissen darfst. Wenn du es nicht programmierst, wenn du es nicht kurz machen willst oder einen bisschen räumen, dann willst du nicht nur das und das und das und das. Das könnte Grenzen haben.

Also einfach, ich glaube, die User einfach ein bisschen einfacher zu entfalten, dass es so lange wie möglich im Modell... Ja. Gut, von dem her schnell einen Schritt machen, weil ich habe ja gesagt, wir hatten mehrere Agenten gehabt. Und ich habe einen Agenten gehabt, der mich nachher fragt: "Hey, kannst du mir noch mehr Informationen geben?"

Kann der lernen?" Und das hintere mentale Bild von der Umgebung kann er verbessern, indem er es lernt.

Genau, kurz raus. Gut, wenn wir uns das Agentenbild mal anschauen, wir haben die Eingabe gehabt, wir haben dann das Modell und die Ausgabe gehabt. Heute ist es eigentlich so, dass man im System eigentlich mehrere Modelle hat.

Also wir haben ein Modell, das die Eingabe kriegt und dann die Eingabe interpretiert. Das ist ein eigenes Modell, nur um zu verstehen, was die Eingabe eigentlich bedeutet. Dann kann man Modelle oder mehrere Modelle haben, die die Verarbeitung machen.

Wir haben vorhin von Gender AI, Multi-Agenten-Systemen gesprochen, die die Eingabe, nachdem sie verstanden ist, verarbeiten und dann ein Modell, das nachher den Output generiert. Also ein Modell muss nicht ein Modell sein, das kann aus mehreren Teilmodellen bestehen. Haben wir auch im Kopf. Wir haben Modelleingabe zu verarbeiten.

Wir haben ein visuelles System, das uns hilft, den visuellen Output zu verstehen. Wir haben ein Modell, das uns hilft, den Sprach-, sorry, visuellen Input zu verstehen, den Sprachinput zu verstehen. Mit diesem Input können wir nachher arbeiten, können uns eine Lösung für ein Problem arbeiten und dann können wir nachher reagieren.

Und dann auch haben wir vielleicht auch wahrscheinlich verschiedene Modelle zum Sprechen und um die Hände zu bewegen und so weiter.

Gut, das haben wir beim autonomen Fahren, haben wir das auch, dass es verschiedene Ebenen gibt. Die Sensoren, dann das Fahren und so weiter. Die Sensoren da zu interpretieren, ist eine eigene komplexe Aufgabe. Dann zu entscheiden, wie das Auto fährt, ist die nächste Aufgabe.

Und dann das tatsächlich umzusetzen, ist dann eine dritte Aufgabe. Also auf allen drei Ebenen haben wir ein oder mehrere Modelle.

Gut, wenn man einen Roboter betrachtet, ist es auch so, ja, das Bild da oben, also man hat die Interpretation von den Sensordaten, Computer wischen, dann macht das System einen Plan, wie es, was das ist, der Roboter machen wird und dann wird der Plan ausgeführt.

Ein Auto ist ja, wie so ein pünktlicher Roboter, halt relativ komplex. Also das ist im Moment, was wie man dieses System aufbaut. Es ist nicht ein Modell, es können durchaus mehrere Modelle sein, die miteinander, hier haben wir ja vorhin auch schon angesprochen, dass es mehrere Agenten sein kann.

Und da kann man sich das Quartettsystem anschauen. Im Prinzip ist es ja so, wir haben, müssen das System trainieren und dann arbeitet das System mit Feedback, kann sich verbessern und liefert dann einen Output. Und da gibt es auch verschiedene Umsetzungen.

Wir haben, wenn du so eine Suche anschaust, dann ist es so, dass man eine Anfrage hat. Dann gibt es irgendwo einen Retrieval, die Information zu finden. Dann wird aus dem zusammen mit dem, was das System intern als Wissen hat, eine Antwort generiert. REG ist soähnlich.

Also macht ein Information Retrieval, Retrieval of Memory und dann Generation, die Antwort generieren, die dann nachher an die Umgebung geht.

Und dann gibt es noch Dialogsysteme, die gar nicht aufgrund von datenbasierter KI kommen. Also Systeme wie Dialogflow, du kannst selbst programmieren, was die Antwort ist auf eine Frage oder wie die Leistung gearbeitet hat.

Dann kommt das erstmal das Sprachverstehen, dann das Dialogmanagement, zu sagen, auf diese Art von Angabe, also auf diesen Incident, den man erkennt, in dem ersten Teil reagieren wir auf diese und

jene Art und dann wird der Output generiert, auch wieder in natürlicher Sprache, also Sprachgenerierung. Hier noch ein wichtiger Hinweis.

Und zwar, die Modelle sind spezialisiert. Die haben ganz bestimmte Aufgaben. Jedes Modell einzeln. Vorher hatten wir Modelle gehabt, die universell sind, die alles lernen können. Die können nur bestimmte Sachen lernen. Kleiner Gedanke für den Heimweg. Wir sind auch spezialisiert.

Also wir können nicht alles lernen.

Das ist etwas, was ich sagen will. Wir sind nicht universelle Lernmaschinen, sondern wir sind auch spezialisiert. Wir haben ein Hörsystem, wir haben ein optisches System. Wir haben irgendwie so ein bisschen einen präfrontalen Kortex, wo wir Reasoning machen können, also Sachen überlegen. Aber das ist alles limitiert.

Also die Wahrnehmung von der Umgebung, die Verbesserung vom Modell, vom inneren Modell, ist limitiert. Und die Tendenz geht in die Richtung, dass wir immer mehr spezialisierte Modelle haben. Dafür haben wir natürlich Agenten, die viel, viel besser auf die Umgebung reagieren können. Aber das ist vorgegeben.

Gut, jetzt wollen wir eigentlich noch darüber diskutieren. Haben wir noch drei Minuten Zeit. Was erwartet uns eigentlich? Was können wir uns erwarten für die KI? Also was wir heute haben, dann könnte man vielleicht sagen als Narrow Artificial Intelligence. Warum Narrow?

Kleinere Modelle, die eben spezifischen, kleinen, also in all den Sachen extrem gut sind und somit auch wenig Speicher und Lernaufwand. Ja, aber ChatGPT haben wir auch schon, oder? Das ist aber nicht gut in allen. Ja, genau.

Der Punkt ist schon richtig. Es muss nicht klein sein, aber es ist spezialisiert auf eine Sache. ChatGPT kann mit dir kommunizieren. Ja, dann, da ist es gut. Aber es hat keine visuellen Angaben, es kann agieren, also es ist spezialisiert auf etwas. Die meisten Agenten, die wir heute haben, sind auf irgendetwas spezialisiert.

Also von der allgemeinen KI reden wir definitiv noch nicht. Auch wenn manche meinen, mit ChatGPT haben wir allgemeine KI, das ist in dem Sinne Umwelt, in der Umwelt agieren, das ist ein kleiner Teil, weil es nur einen kleinen Teil von dem Input gibt.

Also Narrow Artificial Intelligence spezialisiert für konkrete Aufgaben und das kann unterschiedliche Sachen sein, Bilderkennung, Textanalyse.

Die Reichweite von dem, was wir kognitive Fähigkeiten am Anfang mal gesagt haben, ist relativ gering, wenn man es gegenüber den Menschen vergleicht. In allen Bereichen. Ja, was kriegen wir morgen oder übermorgen? GAI, AGI, was bedeutet das?

General Artificial Intelligence, oder? Ja.

Ja, Artificial General Intelligence. Ja, dann kommt es auf die Superintelligenz. Genau. Also General Artificial Intelligence oder Artificial General Intelligence. Also Anwendungen von der KI, die breiter sind, also mehr Fähigkeiten abdecken.

Im Prinzip war die Idee, dass man so etwas wie die allgemeine Intelligenz, die wir als Menschen haben, auf dem Computer abbilden kann. Kriegen wir das hin irgendwann? Oder wollen wir das überhaupt?

Ich glaube es nicht. Ich glaube, dass wir es nicht entwickeln können. Warum? Es wäre doch cool.

Ich glaube, dieses Feld wird 2027 heißen. Das sagt irgendwie, dass wir schon auf dem Weg sind und dass das total vorgenommen wird.

Von den OpenAI, glaube ich. Ja, das ist schon bei deren Automation. Also der wird geschasst und hat auf Millionen verzichtet, um gegen die freien Ebenen zu kämpfen. Ach so. Ich hänge euch das rauf. Das müsst ihr lesen. Das ist genau Webseite dazu.

So fünf Steps, oder sechs Steps, was wir jetzt hier machen. Kann man das verhindern, was wir nicht wollen?

Das ist ja kompliziert, das zu überwachen, solche Systeme. Weil man kann ja nicht so einfach irgendwie spezifische Teile aus dem ganzen Skelette AI rauslesen. Das ist sehr, sehr schwer, oder wenn gerade was benötigt.

Und auch so oder so, als Beispiel, kann ja jetzt eine KI komplett eine eigene Sprache generieren, die nicht entsteht. Und dann sind wir ja verloren darin. Wir machen ja in einem anderen Forschungsprojekt genau das, dass Typisch nicht entsteht mit verschiedenen Agenten.

Die müssen über Tokens, die wir verstehen, kommunizieren. Telegraph haben jetzt viel vor, da extrem viel investiert in KI. Und wenn man je nachdem, was für ein Bild man liest, sieht man schon, dass es irgendwo schon Ermittlungen gab.

Und das könnte auch das AGI dann aufhalten, dass die Leute sagen, ich will die Manpower und die finanzielle Investition zu machen. Aber ich glaube, Google, Microsoft, die haben recht viel. Ich habe dann ein brutales Beispiel. Wer braucht ChatGPT für Texte schreiben? Niemand. Alle.

Schön nach vorn. Die sind der einzige Klasse. Die müssen einen Deutschaufsatz schreiben, den ChatGPT nicht braucht, also die Hilfe nicht braucht.

Ich meine, die brauchen es. Also das heißt, wenn die als Nation oder irgendwie, zum Beispiel Waffenroboter oder Drohnen, nicht einsetzen, hat er einen Nachteil. Das heißt, es gibt einen gewissen Druck, das ganze System das anzuwenden. Es wird verhindert, weil das Kontrolle ist eins.

Wir zum Beispiel können wir, glaube ich, zugeben, wobei. Ich kann noch nachschauen, ob die mir ChatGPT mir den Text abgeben. Irgendwann ist das so fertig. Das machen wir dann. Muss man nicht mehr Deutsch können, muss man nicht mehr Texte schreiben.

Ich sage dir, das ist so eine Entwicklung, die man vielleicht nicht kaufen kann. Wenn ich einen AGI habe, einen super Kampfroboter, ich empfehle euch WarGames, das ist ein Satz für die. Du musst es dir mal anschauen. Und ich glaube, ich werde den Turn nur noch hacken. Aber die Methode ist dort. Kann ich so etwas aufhalten?

Wenn der Gegner das hat, ein super Rechencomputer, wo ich jeden Krieg gewinnen kann. Das mache ich, gehe ins Labor und baue noch etwas.

Das ist das Problem, oder? Und das kann man vorher verhindern, in der Nähe. Darum gibt es so Berichte, und darum. Ich weiß nicht, wie es ausgeht, oder? Das ist ein bisschen die Frage, die wir stellen.

Wer denkt, dass man das noch verhindern kann? Ja, es gibt noch ein paar andere Fragen, ob wir mit ChatGPT eigentlich auf dem Weg sind, also mit Sprachmodellen allgemeine KI zu erzeugen, oder? Ja. Das wären. Ja.

Werden Sprachmodelle irgendwann mal zu der allgemeinen Intelligenz? Was denkt ihr? Nochmal, bitte? Sind Sprachmodelle schon allgemeine Intelligenz, oder werden sie mal?

Ich glaube, dass Sprachmodelle nicht an sich AGIs sind, sondern jetzt bei denen. Ich habe ja nicht immer auch eine Lösung für etwas gefunden.

Aber ich weiß auch nicht, ob es möglich ist, dass Sprachmodelle allgemein bei LLMs AGIs werden, weil es einfach schwer ist, so große Datenmengen zu haben. Ich finde, da bräuchte man schon andere Modelle, um das hinzukriegen.

Weil ich meine, jetzt für LLMs braucht man ja schon so unglaublich viele Data Centers, damit das richtig läuft. Und. Aber sie verarbeiten mehr Daten, als wir jemals im Leben sehen werden. Aber wir werden gebraucht als für das Training, als wir sehen. Da gibt es eine Menge der Daten.

Wir sind doch auch allgemein intelligent nach weniger Daten.

Ja, aber es wird schon möglich sein, vor allem weil man eigentlich das neuronale Netz von ChatGPT einfach nehmen kann und auf eine andere Spalte kopieren kann. Diese Spalte dann in ein Roboter einsetzen kann und dann hat dieser Roboter schon das ganze Wissen von ChatGPT. Okay.

Es gibt ja im Moment jetzt von der EU, also es ist der Menschen, was verhindern würde, von behindert. Aus EU-Gründen, aus Angstgründen, aus verschiedenen Gründen.

Aber wenn ich zum Beispiel denke, eine Firma, wenn alle verarbeiten würden, ihre Daten würden teilen miteinander, dann würde das für das Wohl von der Menschheit sicherlich ein Thema Fortschritt sein, dass man mit KI bearbeitet. Dann könnte es neue Lösungen bringen zu Problemen, die wir haben.

Aber das Interesse der Firmen, der Personen dort, ist ein ganz anderes. Wäre ja auch die BITS. Wenn die BITS würde ihre Daten anordnen, das wäre sehr interessant. Aber ich glaube, das ist das, was das behindern würde bis zum Ergebnis.

Okay, das ist in die Richtung, wir können nicht alle Daten haben, dann bräuchten wir noch mehr Daten, die auch spezialisiert sind. Aber interessant fand ich auch diese Sache mit dem Roboter. Weil was kann denn ChatGPT machen?

Es kann aufgrund von Daten in einem Artikel erzeugen, aber kann es Bilder interpretieren, Videos interpretieren aus der realen Welt, darauf reagieren und wieder abagieren, zurzeit ist der Output ein Text, vielleicht noch ein Bild, aber nicht eine Aktion, die startet.

Also von daher ist im Sinne von Umfang, was die kognitiven Fähigkeiten angeht, die wir mal gesagt haben, die Einschränkung für Sprachmodelle relativ klein. Da wird es nicht so anders. Nein, nein, nein, das ist auch so. Aber. Genau.

Ich glaube, was die TIA-Empfehlungen am Testen sind, dass die Inputs von ChatGPT, von der KI, die Automatik macht. Genau. Übrigens Miro mit Abertos, wo wir sind, demnächst Cluster. Genau.

Also sobald wir die, wenn wir Zugriffe auf die richtige Welt haben, kann natürlich das Modell ziemlich viel machen und ist nicht kontrollierbar. Also ich kann hier auf dem Rechner, kann ich dann nicht installieren, aber ich kann sagen, hey, start Word, schreibe einen Text, verschrecke mir den.

Und das nächste Mal muss ich nur sagen, wenn ich es tokenisiere, kann ich sagen, schreibe mir Alex einen Brief, wo es darum geht, einen Termin abzumachen. Und das übernächste Mal sage ich, hey,

mach einfach, was ich das letzte Mal schon gesagt habe. Also dann kann man sich streiken. Ja. Durch den Durchlauf kann ich alles.

Exakt, exakt.

Aber zurzeit sind wir mit dem Ganzen eher auf das, was wir auch als bewusstes Wissen schon bezeichnet haben. Also das ganze Tested Knowledge, das wir haben, müssen wir zuerst in Sprache übersetzen können, damit wir damit interagieren können. Vieles von dem können wir gar nicht in Sprache übersetzen.

Von daher haben wir schon eine Beschränkung, was Sprachmodelle als Basis für allgemeine künstliche Intelligenz angeht. Weil wir viele Informationen gar nicht als Sprache darstellen können. Also vieles wissen, das wir haben. Deshalb habe ich das Beispiel gemacht mit dem Umweltinterpretieren.

Vieles erkennen wir gar nicht bewusst. Wir wissen gar nicht, was das für Objekte sind, die wir sehen. Aber wir wissen, da kommt etwas, das auf uns zugeflogen kommt oder zugerannt kommt. Und dem müssen wir ausweichen. Sollte es sein. Also es ist sehr vieles, was wir gar nicht als Sprache darstellen können, das aber unsere Handlung beeinflusst.

Und da ist sehr fraglich, ob das der Versuch Baustein, genau wie uns für uns Sprache ein Baustein ist. Aber ich bin ziemlich sicher, dass das nicht der einzige Baustein ist, dass wir noch viele andere Bausteine brauchen. SAI.

Das ist die Superintelligenz, von der wir gesprochen haben. Das können wir das nächste Mal diskutieren, noch als Einstieg, bevor wir dann in ein anderes Thema einsteigen. Okay? Vielen Dank fürs Mitmachen. Ich wünsche euch noch einen schönen Tag. Bis nächste Woche.

Ich kann

mich gerade nicht mehr bedanken. Kennst du Chloe?

Aber das ist etwas, was für mich ist das symbolisches Bistante. Wenn ich beschreibe, ist es ein symbolisches Bistante. Interessant sind die, die irgendwie diffus sind, dann kann ich nicht mehr darüber reden. Das, was wir mit der ganzen Röntgenbildung und so haben.

Ich kann versuchen, das zu beschreiben, aber ich bin nicht den Inhalt eines Bildes wirklich in den Text rein. Ich habe es nicht bei dir. Ich darf nicht Mittagessen. Ich muss jetzt in ein Meeting. Es läuft eigentlich schon seit zehn Minuten. Okay, du machst das von mir noch.