Digitalgespräch Folge 36

Wie es Computern gelingt, eigenständig mit Sprache umzugehen

Mit Chris Biemann von der Universität Hamburg, 16. Mai 2023
https://zevedi.de/digitalgespraech-036-chris-biemann/

[Der Vorspann mit Musik und Ausschnitten aus dem Gespräch beginnt.]

Marlene Görger [mg]: Herr Biemann, Sie sind Professor am Fachbereich Informatik der Universität Hamburg, und Sie forschen und lehren zu einem breiten Spektrum der Aspekte maschineller Text- und Sprachverarbeitung.

Chris Biemann [Biemann]: Die Forschung der letzten, keine Ahnung, 30, 40 Jahre ist eigentlich hinfällig, bis auf, dass wir sie gebraucht hatten, um dorthin zu kommen, wo wir jetzt sind. Die Herangehensweise in diesen neuen Sprachmodellen ist eine strukturalistische Herangehensweise im Gegensatz zu der platonischen Höhle der Ideen.

Petra Gehring [pgg]: Kann man sagen, dass Sprache dann vielleicht nicht logisch ist, aber doch irgendwie so aus sich selbst heraus ziemlich effizient, so dass die Maschine diese Effizienz nochmal nacherfindet?

[Biemann]: Sie müssen für das Trainieren von einem großen Sprachmodell ungefähr die Kosten eines schicken Autos rechnen. Also wir sind ganz schnell in Bereichen, dass sich das eigentlich bis auf die ganz großen Firmen keiner mehr leisten kann. Wenn ich das auf Quatsch trainiere, dann kommt auch Quatsch raus. Es ist ja auch nicht so, dass wir Modelle machen wollen, die nur das British English der BBC verstehen.

[Der Vorspann endet, das Gespräch beginnt.]

[mg] Die Welt von uns Menschen ist durchdrungen von Sprache. Indem wir lernen, Gesagtes zu verstehen und selbst zu reden, Texte zu lesen und selbst zu schreiben, wird uns die Sprache, die uns umgibt, zur zweiten Natur. Die vielschichtigen Regeln, mit denen wir in unserer Muttersprache umgehen und wie man sie geschickt brechen kann, spielerisch, kreativ oder innovativ, verinnerlichen wir mit dem Heranwachsen wie von selbst. Für Fremdsprachen allerdings müssen wir Grammatik, Vokabeln und Redewendungen mehr oder weniger mühsam lernen. Dabei hilft uns, dass Sprachen schon seit Jahrtausenden auf in ihnen verborgene Regelmäßigkeiten hin analysiert wurden, so dass wir wie selbstverständlich Grammatiken und Wörterbücher haben, in denen die wichtigsten Merkmale und Begriffe systematisch erfasst sind. Die Botschaft lautet: Sprachen lassen sich all ihrer Lebendigkeit zum Trotz, zumindest in Momentaufnahmen, in Systematiken bannen und von außen betrachten. Dennoch sind wir mit einer fremden Sprache konfrontiert, gibt es nur ein allmähliches Annähern an den kompetenten Umgang mit ihr und man kann ihre abstrakten Regeln bis ins Detail kennen, ohne sie je ganz zu beherrschen. Der Frage, wie man Computern

beibringt, Sprachen der Menschenwelt mithilfe von Algorithmen maschinell zu verarbeiten, widmet sich seit den 1960er Jahren das Fach Computerlinguistik. In unseren Tagen scheint diese Frage weitgehend beantwortet zu sein. Denn ganz offensichtlich können Computer digitale Texte nicht nur verarbeiten und dabei sowohl mit riesigen Textsammlungen als auch einzelnen sprachlichen Äußerungen sinnvoll umgehen. In den spektakulärsten Anwendungen generieren Maschinen selbst Texte und können Dialoge simulieren. Viele Vorstellungen von hochentwickelter künstlicher Intelligenz sind eng mit dieser Form der Textarbeit verknüpft. Wie entstanden also die modernen Sprachmodelle, die uns so ins Staunen versetzen? Worin unterscheiden sich unterschiedliche Ansätze und welche Bedeutung haben sie heute für die Forschung? Wie gelangt man von der maschinellen Textanalyse zum maschinellen Generieren von Text, und welches Verständnis davon, was Sprache ist, wird in solcher Technik abgebildet? Darüber wollen wir heute im Digitalgespräch reden. Mein Name ist Marlene Görger. Ich bin Physikerin und Technikphilosophin am Zentrum verantwortungsbewusste Digitalisierung.

[pgg]: Und ich bin Petra Gehring, Professorin für Philosophie an der Technischen Universität Darmstadt. Unser Gast und Experte für das Thema ist uns, wie immer, über Videokonferenz zugeschaltet. Heute dürfen wir mit Professor Dr. Chris Biemann sprechen, der sich aus Hamburg zu uns eingewählt hat. Herzlich willkommen im Digitalgespräch, Herr Biemann! Vielen Dank für Ihre Zeit.

[Biemann]: Ja, vielen Dank, dass ich dabei sein darf.

[mg] Herr Biemann, Sie sind Professor am Fachbereich Informatik der Universität Hamburg. Dort leiten Sie die Language Technology Group. Und Sie forschen und lehren zu einem breiten Spektrum der Aspekte maschineller Text- und Sprachverarbeitung. Sie sind einer der Autoren des einzigen deutschsprachigen Fachbuchs zum Textmining, dem Standardwerk „Wissensrohstoff Text“. Das Buch ist 2022 in der zweiten Auflage erschienen, und es eignet sich nicht nur als Einführung für Fachleute und Studierende, sondern es kann auch interessierten Laien Eindrücke verschaffen, wie die Computerlinguistik mit Text und Sprache umgeht. Und genau das wollen wir heute von Ihnen erfahren. Zum Einstieg würden wir gern zurück zu den Anfängen. Wie hat man sich denn ursprünglich die Aufgabe zurechtgelegt, Sprache algorithmisch abzubilden?

[Biemann]: Ja, also die Anfänge der Computerlinguistik reichen eigentlich in die 1950er Jahre. Dort wurde vor allem die Anwendung der automatischen Übersetzung vorangetrieben. Man wollte schlicht und einfach die Fremdsprache Russisch aus amerikanischer Sicht im Kalten Krieg verstehen und hat dann sehr vollmundig Versprechungen gemacht, dass innerhalb von fünf Jahren Roboter einfach Russisch verstehen werden können und das Ganze auf Englisch oder auf Amerikanisch in Echtzeit übersetzen werden. Diese Anwendungen gibt es heute, aber damals natürlich noch nicht. Und das ist auch etwas, was sehr häufig passiert ist in Künstlicher Intelligenz, dass die Ankündigungen sehr vollmundig waren und dann aber festgestellt

wurde: Oh, das ist ja doch deutlich komplexer, als wir uns das vorgestellt haben. In den Anfängen der Computerlinguistik war die Verarbeitung von Sprache regelbasiert. Das heißt, die Idee war: Man nimmt sich ein Grammatikbuch, wie man das aus dem Schulunterricht kennt, codiert diese Regeln irgendwie so in den Computer hinein, dass diese Regeln befolgt werden und dass dann entsprechende Repräsentationen entstehen. Zum Beispiel so ein Aktionsplan, dass man einen Befehl ausführt oder eine Übersetzung gebaut werden kann oder eben was sonst die Anwendung ist von Sprachverständnis und Sprachgenerierung. Und wie vieles wurde das in den 50er, 60er Jahren stark vorangetrieben. Dann kam es zu dem sogenannten AI-Winter, dass die Geldgeber Ende der 60er Jahre dann festgestellt hatten, also es war hauptsächlich das Verteidigungsministerium der USA damals: Die Leute kommen irgendwie nicht aus dem Knick, das scheint irgendwie nicht zu funktionieren, und deswegen wurde da sehr viel Geld erstmal wieder abgezogen. Und solche regelbasierten Systeme sind erst mal attraktiv, weil man glaubt, die Kontrolle zu haben. Und man kommt auch relativ schnell zu relativ guten Ergebnissen auf den wenigen Beispielen, die man sich überlegt hat. Allerdings werden solche Regelsysteme wahnsinnig komplex, irgendwann auch nicht mehr beherrschbar, und wurden nie so weit getrieben, dass sie eine Performanz erreicht haben, dass man das auf normale Leute hätte loslassen wollen. Das zweite Paradigma, das dann entstand.

[pgg]: Wenn ich kurz mal fragen darf: Kann man sich das so vorstellen, dass dieser regelbasierte Ansatz ein bisschen daran gescheitert ist, dass man sich Sprache als viel zu logisch vorgestellt hat? Es gibt irgendwie da so ein Regelsystem, irgendeine große Logik, der eine Sprache gehorcht und die Vorstellung, so logisch sei sie, hat sich dann zerschlagen. Kann man das so sich vorstellen?

[Biemann]: Ja, so könnte man das tatsächlich ausdrücken. Das sieht man auch, dass die Beispiele, bei denen das geklappt hat, also wo dann das Regelsystem hingehauen hat, doch immer sehr eingeschränkte Domänen waren. Also da gibt es von Winograd die sogenannte Blockswelt. Terry Winograd ist ein Pionier der Künstlichen Intelligenz. Der hat Anfang der 70er Jahre in Stanford, glaube ich, darüber promoviert und das war beeindruckend, wie man ein System sprachsteuern konnte, welches irgendwie bunte Blöcke, die verschiedene Formen und Farben hatten, aufeinanderstapelt oder nicht. Also in solchen Bereichen ging es immer sehr gut. Aber Sprache ist eben mehr als nur Logik. Sprache ist inhärent mehrdeutig und die Abbildung von Sprache zu Logik funktioniert nicht. Logik ist eben nur ein Teilbereich der Sprache. Und ja, das ist tatsächlich eines der Kernprobleme.

[mg] Sie sagten jetzt grade das zweite Paradigma, da würde ich gerne weiterhören.

[Biemann]: Genau das zweite Paradigma, das ist dann das statistische Paradigma. Man hat dann festgestellt: Ja, man könnte natürlich auch einfach Statistik über Sprachmaterial betreiben und jetzt nicht so regelbasiert top down vorgehen, dass ich dem Computer sage, was zu tun ist, sondern den Computer quasi merken lasse, wie das normalerweise in den Daten passiert, und es dann für mich irgendwie

gewinnbringend nutzbar mache. Anfang der 1980er Jahre wurde das dann sehr stark vorangetrieben, hauptsächlich durch IBM Research. Die hatten verschiedene Systeme, zum einen zur Übersetzung, zum anderen aber auch für die automatische Spracherkennung. Also wenn ich gesprochene Sprache in geschriebene Sprache übersetzen möchte. Und Fred Jelinek, der dieses Team damals geleitet hat, von dem stammte der Spruch: „Every time I fire a linguist, my recognition rate goes up." Also immer, wenn ich einen Linguisten gehen lasse und ihn durch Statistik ersetze, wird mein System besser. Er hat sich danach bei so einem Lifetime Award irgendwie entschuldigt mit dem Titel: „Some of my best friends are linguists." Also, das ist, no hard feelings. Aber das zeigt es ganz gut. Denn es ist vielleicht auch ein Prinzip, was wir jetzt immer noch sehen, dass quasi Datengetriebenheit auf lange Sicht dem Regelbasierten, dem Regelgetriebenen deutlich überlegen ist. Aber dazu muss man sagen: auf lange Sicht. Denn beim statistischen Paradigma kann ich nicht einfach ein paar Regeln schreiben und dann funktioniert das System. Ich muss zunächst mal Material sammeln. Das ist auch erst mal nicht trivial. Also wir reden von den 1980er Jahren. Damals gab es noch kein Internet, was man sich in Teilen oder ganz hätte runterladen können. Damals gab es Digitalisierungsprojekte, so in der Größenordnung einer Tageszeitung. Viel mehr konnte man da auch nicht auf den Maschinen sinnvollerweise speichern. Da ging es um sehr viel Geld damals noch, solchen Speicherplatz überhaupt vorzuhalten, geschweige denn die Verarbeitung, um damit Statistik zu betreiben. Bei dem statistischen Paradigma gibt es zwei Hauptrichtungen. Das eine ist das statistisch Unüberwachte. Hier mache ich Statistik über Wörter und deren Vorkommen und kann dann so Dinge tun wie typische Begriffe, typische Assoziationen zu einem Begriff. Das ist etwas, was immer noch im Textmining stattfindet, dass ich zum Beispiel statistisch herausfinde, dass Butter und Brot irgendwas miteinander zu tun haben. Das Andere ist überwacht. Und was ich hier mache, ist: Ich sammle Sprachbeispiele, ich annotiere sie, das heißt, ich mache da irgendwelche Markierungen dran, welche ich will, dass der Computer irgendwann selbst machen kann, und bringe dem Computer das mit einer sogenannten Klassifikationsaufgabe bei. Also ein Beispiel wäre zum Beispiel, wenn wir Namen im Text suchen möchten, weil Namen ja vielleicht interessanter Anker sind für alles Mögliche, was mit Fakten zu tun hat. Denken Sie an Nachrichtentexte: Wer macht was mit wem wo? Da sind Namen sehr wichtig. In diesem Paradigma würde ich mich jetzt hinsetzen und in irgendwelchen Texten die Namen finden, sie als solche markieren und dann dem Computer zum Lernen geben. Das ist dann ein Paradigma des maschinellen Lernens. Und in der Forschung ist dieser Paradigmenwechsel ungefähr Ende der 90er Jahre, Anfang der 2000er in einem Maße angekommen, dass es nur noch sehr wenige Leute gab, die dort regelbasiert noch danach gearbeitet haben. In der Industrie oder in der Anwendung ist das sehr viel später angekommen, denn da gibt es einen Unterschied, und der Unterschied ist der Entwicklungszeitraum. Ich muss jetzt also nicht nur Korpora also Textsammlungen irgendwie geeignet vorhalten und sammeln, sondern ich muss jetzt auch noch diese Markierungen dranmachen. Und damit tun sich Menschen auch schwer. Also das klingt sehr einfach, aber der Teufel steckt im Detail. Ich muss dann also immer mindestens zwei Leute an die gleiche Aufgabe setzen, eine dritte Person, die sagt: Das ist richtig, das ist falsch. Also da bin ich

dann schon mal ein halbes Jahr beschäftigt mit so einem Team, bis da so viele Daten rauskommen, dass ich da auch sinnvoll was lernen kann. Und gerade in der Anwendung muss es schnell gehen. Ein halbes Jahr hat niemand, der Ausgang ist ungewiss. Deswegen funktionierten bis in die 20er Jahre des 21. Jahrhunderts hinein viele Systeme in der Industrie tatsächlich noch regelbasiert. Genau. Und das war dann die statistische Welt. Und in der statistischen Welt ist dann sehr viel passiert. Das hat dann die Richtung Maschinelles Lernen eröffnet und hat eben auch die Richtung eröffnet, dass wir Sprachmodelle, klassische statistische Sprachmodelle, trainieren und sie für verschiedene Anwendungen nutzbar machen.

[pgg]: Es gibt unheimlich viele Informationen, die wir jetzt gerade im Moment auch irgendwie versuchen uns zurechtzulegen, wo überall mit Sprache automatisiert schon gearbeitet werden kann und wie man es macht. Und wahrscheinlich ist auch vieles jetzt schon so auf maschinellem Lernen irgendwie aufgesetzt, manches vielleicht auch nicht. Das klingt manchmal so wie so ein Werkzeugkoffer, dass es verschiedene Zugriffe gibt, gewissermaßen auf den Gegenstand. Was sind denn so die Standardwerkzeuge, wo Sie sagen würden: Naja, das ist schon mal klar so, in der Computerlinguistik, das hat man so im Koffer, wenn es um Sprache geht?

[Biemann]: Da muss ich tatsächlich noch mal ausholen und das dritte Paradigma, nämlich das neuronale Paradigma, beschreiben. Denn dieser Werkzeugkoffer hat eine sehr starke Veränderung erfahren. Ich kann gerne den Werkzeugkoffer des statistischen Paradigmas beschreiben. Also im Werkzeugkoffer des statistischen Paradigmas ist die sogenannte linguistische Pipeline, die man aufgebaut hat in der Computerlinguistik, und das ist eine Verarbeitungskette, welche der Verarbeitung des Menschen, oder so wie Linguist:innen sich das vorstellen, wie der Mensch sich das verarbeitet, nachempfunden ist. Und das ist auf einer groben Ebene eigentlich relativ unstrittig. Also irgendwie kommt jetzt das Sprachmaterial hinein, sei es jetzt auditiv oder sei es visuell durch Lesen. Und ich muss zunächst mal dafür sorgen, dass ich die Segmentierung richtig kriege: Also zu wissen, wo fängt ein Wort an, wo hört es auf, was sind Satzeinheiten? Das ist sowas, was zum Beispiel Kinder schon im Mutterleib anfangen zu lernen: Wo sind die Silben, wo sind die Grenzen der Wörter, um diese Segmentierung richtig zu kriegen. Dann würde man weitermachen mit der morphologischen Ebene, also man schaut sich die Einzelwörter an und sagt: Ah, okay, da ist jetzt vielleicht eine Endung dran. Was könnte das sein? Ist es eine dritte Person Singular, ist das ein Plural? Wie passt das Ganze zusammen? Dann würde man die Wortarten festlegen. Und mit Wortarten kann man dann schon mal sowas extrahieren, wie zum Beispiel die Nominalphrasen. Also das sind die Dinge, um die es geht im Text. Der grüne Frosch zum Beispiel. Da kommt man auch schon sehr viel weiter, auch ohne die Verben zu betrachten. Die nächste Stufe wäre dann die syntaktische Verarbeitung, dass man sagt: Okay, was bezieht sich auf was im Satz? Da werden dann entweder Syntaxbäume gemalt, wie man sie aus der Schule kennt, oder andere Arten von grammatischer Repräsentation. Dann geht es in die semantische Verarbeitung. Und jetzt wird es auch langsam vom Formalismus her schwierig, denn in dieser Ebene würde man dann die Logik abbilden. Nun wissen wir ja schon, dass die Logik eben

nicht reicht. Das heißt, jetzt habe ich die Voraussetzung, überhaupt erstmal eine Beschreibung der Semantik zu finden, welche allgemeingültig genug ist, um Sprache abzudecken. Das gibt es im Allgemeinen irgendwie auch noch nicht, oder zumindest gibt es da keine, wo alle sagen: Genau, das ist es jetzt. Und die nächste Ebene ist dann die Pragmatik: Was will ich eigentlich damit? Also die Funktionalität. Und je nachdem, was ich für ein System bauen möchte und welche Sprachfunktionalität ich einbauen möchte, brauche ich Teile dieser Pipeline oder alles in dieser Pipeline plus irgendwas obendrauf. Zum Beispiel, wenn ich wissen möchte, wer mit wem heute in der Zeitung was zu tun hatte, dann reicht es völlig, diese Wortartenebene zu machen, zu gucken, was davon sind Namen und dann bin ich fertig. Da brauche ich keine syntaktische Verarbeitung. Wenn ich sowas wie Sentimentanalyse mache, da möchte ich schon syntaktische Verarbeitung, weil ich möchte schon wissen, auf was sich jetzt das positive oder negative Statement bezieht. Und wenn ich Übersetzungen mache, ist das eine Frage, inwieweit ich eine explizite oder implizite semantische Repräsentation brauche, um die Semantik über diese Sprachgrenzen hinweg transportieren zu können. Und das war der Standardbausatz. Also man würde quasi bei einer computerlinguistischen Aufgabe dann gucken: Welche Pipeline passt? Wie viel brauche ich von dieser Pipeline? Habe ich diese Komponenten in den entsprechenden Sprachen? Benötige ich die noch? Und das ist dann so eine Art Grundsprachverständnis in der Maschine, und dann wird das angepasst auf die spezielle Aufgabe. Jetzt würde es sich ganz gut anschließen, über das neuronale Paradigma zu sprechen.

[pgg]: Ganz genau. Jetzt sind wir gespannt.

[Biemann]: Genau, weil an und für sich ist diese Unterscheidung, das Grundverständnis von Sprache und das auf die Anwendung bringen, weiterhin gültig. Es ist nur so, dass diese linguistische Pipeline, die sehr kleinteilig und mit sehr viel Arbeit aufgebaut wurde, weil ich ja, egal ob ich jetzt regelbasiert oder statistisch arbeite, für all diese Ebenen genug Daten brauche oder genug Regeln brauche und genug Testdaten brauche, um zu gucken: Wie gut bin ich denn eigentlich? Weil nur was ich testen kann, kann ich auch verbessern. Es ist eine empirische Wissenschaft an der Stelle. Und das nimmt uns jetzt das neuronale Paradigma ab, indem es sagt: Dieser gesamte Block, dieser gesamte Vorverarbeitungsschritt dieser Pipeline, den lernen wir komplett automatisch, rein durch sehr, sehr viele Daten, sehr, sehr viele Sprachdaten, die wir einfach in so ein Large Language Model hineinstecken, und dieses Large Language Model tut, was es muss, um dieses Sprachverständnis zu erlangen, welches wir dann ubiquitär so einsetzen können, wie wir das eben wollen für unsere Anwendungen. Und dann muss ich wieder gucken, wie bekomme ich jetzt dieses Kernsprachverständnis mit meiner Anbindung verbunden an der Stelle.

[pgg]: Das heißt, es wird quasi alles auf einmal abgetastet und alles auf einmal einsortiert und klassifiziert und ins Verhältnis zueinander gebracht?

[Biemann]: Ja, und das ist wirklich faszinierend, dass das so klappt. Aber man hat natürlich sehr lange daran rumprobiert, bis man das irgendwie hinbekommen hat. Man muss sich diese Modelle, wie zum Beispiel das ChatGPT, was jetzt in aller Munde ist – aber da gibt es ja auch einige an Vorläufern –, so vorstellen, dass das neuronale Netze sind, die verschiedene Ebenen haben. Es gibt quasi die unterste Ebene, wo das Ganze hineinfließt, und es gibt die oberste Ebene, wo dann irgendwie was passiert, und sehr viele Ebenen dazwischen. Und als diese Modelle rauskamen, gab es eine Reihe an Arbeiten, die untersucht hat, was diese Modelle eigentlich tun. Und tatsächlich ist es so, dass in diesen verschiedenen Ebenen Dinge passieren, die man mit der linguistischen Pipeline in Korrelation bringen kann. Also sehr starke Korrelationen, also wie als ob quasi implizit aus der Aufgabe, die diese Modelle lernen, das Modell merkt, dass es notwendig ist, so etwas zu haben wie eine syntaktische Repräsentation, sowas zu haben wie eine Gruppe von Wörtern, welche ähnliche Eigenschaften haben, wie zum Beispiel Namen, und das dann in den internen Repräsentationen, das sind einfach irgendwelche Zahlenvektoren, die man natürlich versuchen kann zu interpretieren, aber die so groß sind, dass man das für gewöhnlich nicht komplett durchdringen kann als Mensch, dass diese dort angelegt werden. Und das finde ich sehr interessant, weil diese Modelle quasi so groß sind, dass sie emergentes Verhalten zeigen. Also, dass ich nicht explizit sagen muss, was sie tun sollen auf diesen Ebenen, und sie tun es trotzdem.

[mg] Das heißt, die Modelle reproduzieren eigentlich das, was man vorher schon dachte, wie man das Problem zerlegt?

[Biemann]: Richtig. Und ich bin mir relativ sicher, dass wenn wir das nicht vorher probiert hätten, mit diesem Modellzerlegen vorher, dann wären wir nicht dort hingekommen, diese Architektur so aufzubauen. Da hat das Feld natürlich sehr viel gelernt. Also jetzt die Forschung der letzten, keine Ahnung, 30, 40 Jahre ist eigentlich hinfällig, bis auf, dass wir sie gebraucht hatten, um dorthin zu kommen, wo wir jetzt sind.

[pgg]: Kann man sagen, dass Sprache, also so unsere Sprache, so wie sie so ist, in ihrer ganzen Vielfalt, wie wir sie alle benutzen, dann vielleicht nicht logisch ist, aber doch irgendwie so aus sich selbst heraus ziemlich effizient, so dass die Maschine diese Effizienz nochmal nach erfindet?

[Biemann]: Ja, also das Interessante an Sprache diesbezüglich ist, dass ich, wenn ich eine Sprache nutze, irgendwie immer die Balance finden muss zwischen: Sage ich es explizit genug, dass die Nachricht, die ich übermitteln möchte, ankommen kann, aber sage ich es kurz genug, dass das noch effizient ist? Wenn ich jetzt es übertreibe mit explizit, dann rede ich in logischen Formeln, die natürlich irgendwie weiter, weiter, weiter runterspezifiziert werden könnten. Und wenn ich zu kurz bin, werde ich unverständlich. Und Sprache geht genau diesen Mittelweg. Das heißt, Sprache ist auch deswegen inhärent mehrdeutig, weil wir diese Mehrdeutigkeiten dazu verwenden können, dass wir Sachen unterspezifizieren, die wir gar nicht spezifizieren müssen, weil

sie aus dem Kontext klar sind. Das ist vielleicht so der starke Unterschied zu logischen Formeln oder Programmiersprachen, wo ich alles bis ins Einzelne spezifizieren muss, sonst kann man quasi nicht handeln. Die Herangehensweise, die jetzt hier sehr konsequent angewendet wird in diesen neuen Sprachmodellen, ist eine strukturalistische Herangehensweise, im Gegensatz zu der platonischen Höhle der Ideen. Also es ist nicht so, dass die Konzepte irgendwie formal irgendwo schon definiert sind und ich kann sie mir einfach holen und in eine Welt einbauen, sondern das Verständnis von Sprachsemantik, welches hier implementiert wird, ist eigentlich eins, das auf de Saussure zurückgeht oder Wittgenstein, wo ich sage, die Bedeutung eines Symbols, also in dem Fall eines Wortes, ergibt sich nur und ausschließlich aus den Kontexten, in denen ich dieses Symbol beobachte. Und das ist natürlich eine radikale Konsequenz, weil ich jetzt dann eigentlich nur noch fragen muss: Wie kann ich das formalisieren? Und ich kann mir das so formalisieren, indem ich sagen kann, ich nehme für die Repräsentation der Bedeutung eines Wortes die Summe aller Kontexte, in denen dieses Wort auftritt. Und dann habe ich plötzlich die Möglichkeit, Worte anhand ihrer Kontexte zu vergleichen, weil vielleicht haben die ja genau die gleichen Kontexte, teilweise, teilweise eben nicht. Und schon kann ich so was aufspannen wie eine Ähnlichkeit zwischen Wörtern oder zwischen Sätzen oder zwischen Dokumenten. So was kann man dann auf verschiedene Ebenen bringen. Und was jetzt passiert, ist, dass ich eine verteilte Repräsentation habe. Was meine ich damit? Das heißt, dass nicht die Bedeutung eines Wortes als Definition, als logische Formel vorliegt, die dann aus irgendwelchen atomaren Eigenschaften irgendwie zusammengesetzt werden muss, sondern eigentlich als mathematischer Vektor vorliegt. Und dieser Vektor bezeichnet eben alle möglichen Kontexte, die es gibt. Und dann steht in diesem Vektor drin: So oft war es in diesem Kontext, so oft war es in diesem Kontext. Und das erlaubt mir, Wörter in Computern als Vektoren zu repräsentieren. Und das ist eine Art von Repräsentation, die wir auch in menschlichen neuronalen Netzen, sage ich mal, sehen. Wir haben kein einzelnes Neuron pro Wort in unserem Kopf, sondern wir haben eine Konfiguration von Neuronen, welche irgendwie diese Bedeutung abbildet. Und genau so ist die Idee jetzt quasi auch bei diesen neuen Sprachmodellen.

[pgg]: Wenn ich selber was sagen will, kann ich mir irgendwie ungefähr vorstellen, wie ich diese dann ja sozusagen sehr fluide bereitliegende Zugriffskompetenz, die ich da habe, indem ich eigentlich alles nutzen kann, um das zu sagen, was ich sagen will, aber ich nutze einen bestimmten Pfad, weil sich das durch die Kontexte als, ich sage mal, vielleicht nicht effizientester, aber optimaler Pfad so schon irgendwie erweist. Vielleicht habe ich da auch Erfahrung damit und außerdem will ich ja was sagen. Wenn jetzt so eine Maschine mit so einem großen Modell arbeitet, dann will die ja nicht unbedingt was sagen. Wie kriegt die diese Pfadwahl hin?

[Biemann]: Ja, das ist eine sehr gute Frage. Ich würde es gerne am Beispiel von ChatGPT machen und von alten Sprachmodellen. Also man hat früher angefangen, diese Sprachmodelle so zu bauen, dass man sagt: Wir machen Statistik, wir schauen einfach, welche Wörter hintereinander vorkommen. Und dann kann ich sowas machen wie: Wenn ich die ersten beiden Wörter weiß, was ist jetzt die

Wahrscheinlichkeitsverteilung für das dritte Wort? Schaue ich einfach: Wie häufig habe ich die beiden gesehen? Was habe ich danach gesehen? Wie häufig ist das passiert? Und dann kann ich beim Generieren jetzt mir eins von diesen raussuchen, und zwar eben proportional nach Häufigkeit. Wenn ich sowas mache, dann kommen da Texte raus, die man sehr gut vorlesen kann. Allerdings sind die teilweise sehr unsinnig, es sind fast alle ziemlich unsinnig. Und das hat damit zu tun, dass da irgendwie, wie Sie sagen, einfach kein Ziel dahinter ist und einfach auch der Horizont von diesen Modellen relativ beschränkt ist. Also wenn ich nur zwei, drei Wörter nach hinten gucken kann, oder lassen Sie es fünf sein, aber mehr kriegt man nicht hin. Hat was mit Statistik zu tun und mit seltenen Ereignissen. Dann wird das Ganze lokal konsistent, aber insgesamt ziemlich inkonsistent. Wenn ich jetzt diese großen Modelle nehme, die funktionieren ein wenig anders. Die sagen, sie lernen quasi Lücken vorherzusagen im Text, und dazu können die die Kontexte verwenden, und zwar sehr breite Kontexte. Und die selektieren sie selber und lernen, wie man sie selektiert. Ich bin nicht mehr gebunden auf das Wenige, was ich mir statistisch leisten kann, sondern ich kann die Historie nach hinten, so nenne ich es mal, jetzt auf, ich sage mal, viele 100 oder einige 1000 Wörter erweitern. Wenn ich das mache und wenn ich auch noch das Ding lernen lasse, dass es ja Sätze auch beenden muss. Und das lernt es dadurch, dass es ab und zu mal rein von der Häufigkeit mal den Punkt vorhersagen muss. Dann bekomme ich jetzt nicht mehr nur irgendwie Aneinanderreihungen von Dingen, die lokal irgendwie vorlesbar sind, sondern bekomme ich kohärente Texte. Diese Texte sind aber, wie Sie sagen, immer noch nicht an eine Aufgabe gebunden, sondern es generiert so vor sich hin. Und was jetzt obendrauf kommt, ist etwas, was wir außen rum bauen. Und das ist genau das, was irgendwie Sprachkompetenz in diesem Modell von Anwendung unterscheidet. Ich muss, um das nutzbar zu machen, das in irgendetwas einbauen. Und bei ChatGPT zum Beispiel wurde es eingebaut in ein Dialogsystem. Da haben dann Leute sich hingesetzt, haben gesagt, so soll die Antwort sein, oder haben gesagt, die Antwort gefällt mir nicht, schreib es bitte so um, oder mal Daumen hoch, die Antwort war gut. Und dann wird quasi auf diese Komponente, die Sprachverständnis an sich hat und jetzt schon ganz gut und kohärent generieren kann, noch etwas draufgesetzt, was jetzt das Bedürfnis hat, den Zweck zu erfüllen. Und das lernt es quasi in einem extra Schritt, der obendrauf ist. Dafür brauche ich wieder Daten, und das ist anstrengend und das ist nicht emergent. Da muss ich wieder Leute fragen, sogenannte Annotator:innen, welche mir eben sagen, was richtig und was falsch ist.

[pgg]: Da kommen dann auch im Grunde wieder Spielregeln ins Spiel: Löse die Aufgabe und löse sie richtig, und achte auf alle Details der Aufgabenstellung oder sowas.

[Biemann]: Ja, wobei das, also wenn Sie sagen jetzt Spielregeln, dann denkt man ja da wieder an regelbasierte Systeme. Das würde es quasi implizit lernen. Also wenn ich der Maschine eine Aufgabe gebe und die wird nur teilweise gelöst und ich gebe dann das Feedback, da fehlt noch etwas. Dann lernt die Maschine über viele Beispiele und über

viele Iterationen, das dann eben nicht mehr zu tun. Und plötzlich achtet sie auf die Details der Aufgabe.

[mg] Ich hätte noch mal eine andere Frage, die sich auf Daten bezieht, mit denen gelernt wird. Sie hatten ja vorhin potenziell das ganze Internet im Grunde als Datensatz genannt, auf dem die Maschine lernen kann. Ist das überhaupt eine sinnvolle Idee? Ich kann mir vorstellen, dass sozusagen, also ich kann mir vorstellen, dass die Texte und auch die Kontexte und die Art, sich auszudrücken, und auch die Abweichungen, die ja Menschen produzieren, die irgendwie online kommunizieren, eigentlich viel zu inhomogen sind, als dass das irgendwie sinnvoll verarbeitet werden könnte, oder unterschätze ich da die Maschinen schon?

[pgg]: Ich hätte jetzt umgekehrt vermutet, dass da viel zu ähnliche Sachen im Netz immer nur so kurze Informationen, kaum Langtext oder so, also in beide Richtungen könnte man sagen: Das ist doch ein sehr spezieller Diskurs. Und wenn es ein Diskurs der Diskurse ist.

[Biemann]: Ja, natürlich muss man dazu sagen, zum einen garbage in garbage out, wie wir sagen. Also wenn ich das auf Quatsch trainiere, dann kommt auch Quatsch raus. Und Quatsch kann in dem Fall natürlich auch bedeuten, dass das Ding sexistisch, rassistisch ist. Also irgendwelche Biases, die ich quasi dort beobachte. Oder das Internet ist ja voller Englisch, was nicht von Muttersprachlern ist, das heißt, die Sprachqualität im Netz, gerade bei Englisch, ist zweifelhaft. Das ist so die eine Position. Diese Position hatten wir auch in der Computerlinguistik. Das war immer die Frage nach: Was ist das richtige Korpus? Da gibt es Bestrebungen zu Nationalkorpora, also das British National Korpus, es gibt das American, noch ein paar andere, Deutsch hat keins. Und da war immer so die Frage: Wie kann ich denn das jetzt so balancieren, dass ich da die richtigen Sprachbeispiele und die richtige Komposition davon habe, dass ich eben die Sprache an sich abbilde? Das hat sich aber auch überholt. Gerade schon im statistischen Paradigma geht das schon los. Die Beobachtung: Mehr ist besser. Und zwar unbesehen der Qualität. Und das mag jetzt erstmal verwundern. Aber es ist so, dass diese Modelle bezüglich thematischer Domänen von verschiedenen Sprachen oder von verschiedenen Subsprachen fragmentieren. Was meine ich damit? Es gibt quasi innerhalb des Modells Untermodelle. Wenn ich jetzt, um beim einfachen Beispiel zu bleiben, wenn ich jetzt ein Modell gleichzeitig auf deutschen und englischen Texten trainiere, meinetwegen auch so ein einfaches Sprachmodell, was nur zwei Wörter nach hinten guckt und ihm jetzt zwei deutsche Wörter gebe: Es war. So, dann fängt das an, Deutsch zu generieren und es bleibt im Deutschen. Es gibt sehr wenige Stellen, die für das Modell für Englisch und Deutsch gleich aussehen, und ohne, dass ich dem Modell sage, welche Sprache es generieren soll – es bleibt im Deutschen, weil es sehr wenige Zweiwortkombinationen gibt, die in beiden Sprachen vorkommen. Und je größer die Modelle und je schlauer, desto mehr habe ich diesen Effekt. Und es hat sich eigentlich durchgesetzt zu sagen: Wir nehmen am besten alles, weil alles ist Teil der Sprachkompetenz. Und es ist ja auch nicht so, dass wir Modelle machen wollen, die jetzt nur irgendwie das British English der BBC verstehen, sondern wir wollen ja

Modelle machen, die Leute auch verwenden können. Und wenn die dicke Finger beim Tippen haben, dann haben die dicke Finger beim Tippen. Soll das Modell doch robust sein gegenüber solchen grammatikalisch sehr zweifelhaften Konstruktionen. Das ist überhaupt dann kein Problem, und deswegen macht man das so, deswegen schert man sich da gar nicht drum. Das ist zum einen einfacher, zum anderen gibt es aber auch eine ganze Serie an Arbeiten, die gezeigt hat, dass ich sehr viel Arbeit machen muss, um aus kleineren Kollektionen höherer Qualität die gleiche Leistung herauszuholen wie aus einer großen Kollektion, im Sinne: Ich schütte da einfach mal alles rein. Die Gefahr natürlich ist, dass diese Modelle dann Stereotypen reproduzieren. Dass es da Biases gibt, dass ich dann vielleicht auf irgendwelchen Foren trainiere, wo es um Frauenhass geht oder um – also es gibt ja nichts, was es nicht gibt im Internet. Das ist dann auch eine interessante Frage: Was machen wir dann damit und wie gehen wir damit um? Also die momentane Ansicht oder der momentane Ansatz ist eigentlich zu sagen: Da bauen wir jetzt was außen rum, wo wir aufpassen, dass es nicht zu viel von sich gibt, was wir eben nicht wollen. Aber da gibt es jetzt keinen inhärenten Mechanismus, dass man ihm jetzt Sexismus oder Rassismus abtrainieren könnte.

[mg] Aber so ein anderer Aspekt, dieses am besten mit allem trainieren, was kostet das dann an Ressourcen, diese Modelle zu trainieren und da dann auch so maximalistisch ranzugehen?

[Biemann]: Also da gibt es verschiedene Schätzungen und ich weiß es auch nicht im Einzelnen. Teilweise wird das auch hinter verschlossenen Türen gemacht. Aber Sie müssen für das Trainieren von einem großen Sprachmodell, wenn Sie das auf der Cloud machen, ja ungefähr die Kosten eines schicken Autos rechnen. Also da sind wir schon bei fünf-, sechsstelligen Eurobeträgen für einmal Trainieren dieses Modelles. Dabei bleibt es aber nicht, wenn Sie diese Modelle einfach nur einmal trainieren wollen und wissen, wie man trainiert, dann ist das ungefähr das, was Sie zu bezahlen haben. Wenn diese Modelle entwickelt werden. Dann gibt es natürlich, wie bei allem, sehr viele Stellschrauben, und diese Stellschrauben müssen optimiert werden. Und die werden so optimiert, nennt sich Hyperparameter, falls das interessant ist für die Runde, dass man die einfach ausprobiert in Kombination. Jetzt nicht alle mit allen, sondern da hat man dann gewisse Erfahrungen, was mit was zusammenhängt. Aber es gibt immer so 3, 4, 5 Einstellungen, die man dann pro Hyperparameter ausprobieren möchte. Und da sind wir ganz schnell dann im Millionenbereich. Also wir sind ganz schnell in Bereichen, dass sich das eigentlich bis auf die ganz großen Firmen keiner mehr leisten kann. Also insbesondere wir als mittelgroßer Lehrstuhl für Computerlinguistik/Sprachtechnologie haben weder Zugriff zu den entsprechenden Hardwareservern, das sind Grafikkarten, auf denen so was läuft, noch haben wir die Möglichkeit, das zu kaufen oder irgendwie zu mieten. Das ist einfach zu teuer, das ist zu viel Geld, das ist inzwischen in Hand der großen Player, also Facebook, Google, OpenAI, IBM, solche Firmen.

[pgg]: Das heißt, Forschung im engeren Sinne, öffentlich und offen und rein aus Wissensdurst betriebene Forschung, findet da ihre Grenzen, und es ist jetzt im Grunde schon industrielle Entwicklung, wenn da Sachen optimiert werden?

[Biemann]: Das ist richtig, aber da verschieben sich natürlich Sachen. Also ich mein, ich bin ja zum Beispiel auch froh, dass ich den Kugelschreiber, mit dem ich Notizen machen, nicht mehr selber bauen muss oder den Taschenrechner nicht selber bauen muss, indem ich meine Sachen ausrechne. Und ich kann natürlich auch mit solchen Modellen in anderen Kontexten Forschung betreiben. Was ich aber eben nicht mehr kann, ist die Grundlagenforschung an diesen Modellen an sich.

[pgg]: An was forschen Sie denn gerade?

[Biemann]: Ich forsche gerade an der digitalen Transformation der Wissenschaft. Ich leite im Neben- oder Hauptberuf inzwischen eine zentrale Einrichtung an der Uni Hamburg: House of Computing und Data Science. Und wir kümmern uns darum, wie man Wissenschaft digitaler macht, und zwar alle Bereiche der Wissenschaft. Und das ist natürlich auch Teil davon. Also solche Dinge setzen wir ein in der Anwendung, um zum Beispiel hermeneutische Prozesse zu unterstützen. Wir sagen, wir wollen Toolumgebungen schaffen, die uns ermöglichen, mehr Material mit computergestützten Methoden so zu verarbeiten, dass wir weiterhin dem Menschen die komplette Kontrolle überlassen, aber dadurch ermöglichen, durch die Sichtung größeren Materials nicht nur qualitativ, sondern eben auch quantitativ in so einem Mixed-Methods-Ansatz zu arbeiten.

[pgg]: Das heißt, ich übersetze mir das jetzt mal: Sie helfen wissenschaftlichen Leserinnen und Lesern beim Verstehen auch von Texten mit Blick auf anspruchsvolle Fragestellungen? Indem Sie da eine automatische Assistenz anbieten?

[Biemann]: Ja, so ungefähr.

[pgg]: Was für Fächer profitieren denn da?

[Biemann]: Also von diesem Ansatz sollten eigentlich alle profitieren können, aber die, die am meisten profitieren, momentan, sind die Fächer, die eigentlich traditionell eher weniger technikaffin sind. Also so was wie die Alte Geschichte, wie die Kulturanthropologie, wie weite Teile der Geisteswissenschaften. Wie Sie vielleicht wissen, haben wir ja noch ein Projekt zur digitalen Begriffsgeschichte, wo wir eben auch Google Books analysieren und die Bedeutungsveränderung von Schlüsselkonzepten über die Zeit. Und das ist etwas, was man mit sehr viel intellektueller Arbeit machen kann. Man kann es auch automatisch gestützt machen. Und ich denke, der Königsweg ist, beides zu verbinden und dann sich gegenseitig abzugleichen, und so zu einer quantitativ fundierteren Ansicht der Wahrheit zu gelangen, als das qualitativ allein möglich ist.

[mg] Da drängt sich mir so ein bisschen die Frage auf, wie dann in solchen Zusammenhängen die Schnittstelle dann zu dem, was die Maschine ausgibt, gestaltet ist. Wir sind jetzt ja durch die öffentliche Diskussion und diesen Fokus auf ChatGPT so sehr in diesen Dialogsystemen drin. Man gibt da irgendwie was ein und dann kommt ein Text raus, und den nimmt man dann so, wie er ist, und fragt vielleicht nochmal nach. In den Forschungsprojekten, die Sie so vor Augen haben: Was sind da die Darstellungsweisen, mit denen man sich dann befasst? Ist das immer auch dann wieder Text, oder sind das Visualisierungen auch, sind das statistische Ergebnisse?

[Biemann]: Das sind meistens relativ komplexe Oberflächen, also Oberflächen, die jetzt nicht so intuitiv sind, dass man sich dann hinsetzt und das sofort kann. Da braucht man dann schon mal eine Einführung, eine Schulung, und sei es ein Viertelstundenvideo, um zu sehen, was man da machen kann. Denn – also das sehen manche anders – aber ich bin jetzt kein Fan von diesem Dialog ist alles Paradigma. Dialog hat auf jeden Fall seine Berechtigung, ist auf jeden Fall attraktiv, so als ja Sparringspartner, als Agent, mit dem man das quasi dann zusammen durchgeht. Aber bei manchen Sachen wollen sie kein Dialogsystem. Also denken Sie zum Beispiel an die klassische Googlesuche, wenn Sie die Ergebnisse sichten, um sich dann zu entscheiden: Auf was klicke ich denn jetzt auf der ersten Seite? Das wollen Sie nicht vorgelesen haben, oder Sie wollen es nicht in einem Text zusammengefasst haben, sondern Sie sind schon darauf geschult, irgendwie zu sehen: I know it, when I see it. Und dann klicke ich drauf. Dieses Prinzip haben wir ganz oft in den Oberflächen: ein Anbieten von Möglichkeiten. Ich gebe Ihnen ein Beispiel, was für uns so ein Feature ist, was wir sehr häufig verwenden: die satzbasierte Ähnlichkeitssuche. Wenn wir eine Kollektion haben, die wir geeignet irgendwie indiziert in unserer Oberfläche drinhaben, ist es ja häufig so, dass Leute jetzt irgendeinen Kontext, irgendeinen Satz, irgendein Versatzstück, aus irgendeinem Grund interessant finden, weil es für ihre, keine Ahnung, Diskursanalyse oder was auch immer die Leute machen, interessant ist. Jetzt können Sie sagen: Ich möchte mehr desselben haben. Die Maschine weiß gar nicht, was die Leute suchen, spuckt aber andere Sätze in anderen Dokumenten aus, wo eine ähnliche Semantik vorliegt. Dann habe ich wieder das Prinzip: Ich habe jetzt irgendwie so Suchergebnisse, die gerankt sind nach Ähnlichkeit in dem Fall, und dann suche ich mir die raus, die mir gefallen und die mir nicht gefallen. Sowas möchte ich nicht im Dialog haben. Oder wenn es darum geht, Visualisierungen über die Zeit zu machen, also so schnell zu sehen: Was ist ein Thema und wie sind die Themen verteilt und wie entwickelt sich das über die Jahre oder Jahrzehnte? Das sind so die typischen Anwendungen.

[pgg]: Das heißt, Sie setzen schon darauf, dass Expertinnen und Experten auf der anderen Seite sind und dass die auch Freiheitsgrade haben wollen, um dann selbst zu deuten, was sie jetzt aus diesen maschinellen Angeboten machen und dass sie da gewissermaßen noch so eine unfertige oder mehrdeutige Form von Ausgabe bevorzugen und jetzt nicht irgendwie den einen Vorschlag, den man dann entweder nimmt oder verwerfen muss?

[Biemann]: Genau richtig, denn das ist ja eine Anwendung in der Wissenschaft. Und Wissenschaft kann ja nur funktionieren, wenn es mehrere Zugänge und mehrere Möglichkeiten gibt. Und das hat natürlich auch was mit Transparenz zu tun. Also diese Systeme bringen einen ja auch recht schnell auf den einen Pfad, den man dann weiter austritt und vernachlässigen vielleicht die anderen Pfade und das ist jetzt auch auf Dauer wahrscheinlich nicht zuträglich.

[pgg]: Gehört ja auch zu den Kritikpunkten an Dialogsystemen vom Typ ChatGPT, dass man im Grunde nur dieses Ergebnis hat, ohne irgendwelche Hinweise aufs Zustandekommen. Man kann natürlich noch mal nachfragen und spezifizieren lassen, aber es ist doch ziemlich geblackboxt.

[Biemann]: Ja, und es ist dem System inhärent. Also ich hatte Ihnen ja beschrieben, wie das mit diesem einfachen Sprachmodell geht. So zwei Wörter nach hinten usw. und wenn ich nur Wörter zähle und deren Reihenfolge, dann habe ich überhaupt keine Möglichkeit zu sagen: Aus welcher Quelle kommt das eigentlich? Es ist in diesem Modell nicht vorgesehen. Natürlich könnte ich mir das für jede einzelne Stelle irgendwo mitspeichern, aber das ist ziemlich wahnsinnig. Aber trotzdem, ich kann es natürlich indizieren und dann für drei Wortbegriffe gucken: Wo kommen die vor? Bei diesen modernen Large Language Models wie ChatGPT ist es inhärent genauso. Also ich kann quasi das nicht. Ich kann das gar nicht mitspeichern. Es wird auf jeden Fall abstrahiert, aber da sind noch ein paar Abstraktionsebenen drüber, weil ich eben nicht auf Einzelwörtern operiere, sondern nur so auf Repräsentationen von diesen Vektoren, die ich vorher ein bisschen ansprach. Deswegen ist dieser Schritt zurück zur Quelle komplett verloren und lässt sich auch nicht leicht reparieren. Also das ist etwas, wo jetzt natürlich sehr viele Leute dran arbeiten, weil das natürlich das Interessante ist, wie mache ich das Ganze irgendwie zuverlässiger, weil diese Systeme ja weiterhin noch halluzinieren. Also die reden ja irgendwas teilweise oder hauen da irgendwelche nicht faktischen Statements rein, die aber so gut klingen, dass man es dann glauben mag, und das ist ja das Gefährliche an der Geschichte. Das ist ein aktives Forschungsfeld tatsächlich: Wie repariere ich das? Oder vielleicht sogar: Wie schaffe ich Modelle, die das irgendwie inhärent mitschleifen mit diesen Quellen?

[mg] Das leitet so ein bisschen auf eine Frage zu, die ich mir auch gestellt habe. Was hat die Computerlinguistik jetzt eigentlich noch zu tun? Also ist das sozusagen ein Noch-besser-machen eines jetzt schon sehr guten Ansatzes, jetzt schon sehr guter Systeme? Oder könnte man sich vorstellen, dass es auch noch mal einen Paradigmenwechsel gibt irgendwann?

[Biemann]: Also es ist schwer vorstellbar, dass es da jetzt in den nächsten Jahren oder wenigen Jahrzehnten einen ganz neuen Paradigmenwechsel gibt, weil das Level, was wir hier erreicht haben, an Sprachverständnis, und ich weiß nicht, ob Sie schon mal GPT4 gesehen haben – das ist ja nochmal eine Ecke besser, da ist das Feld eigentlich jetzt erst mal fertig. Also manchmal habe ich das Gefühl, tatsächlich: Es ist gelöst. Was machen wir denn jetzt? Natürlich gibt es da noch Einiges zu tun, was aber anders

gelagert ist. Also angesprochen: Wie schaffe ich das, mit Quellen zu verbinden? Angesprochen: Wie schaffe ich es, diesen Bias da rauszuholen? Man kann schon dem Ding Bias einzeln abtrainieren. Also wir hatten mal eine Arbeit laufen von Angelie Kraft. Sehr schöne Arbeit, die hat auch einige Preise gewonnen, wo sie es geschafft hat, so einem generativen Modell den Sexismus abzutrainieren. Das hat funktioniert. Das war dann nicht mehr sexistisch. Es war aber weiterhin rassistisch, weil das eine hat mit dem anderen nichts zu tun. Auf Modellebene. Und wie kann man das jetzt irgendwie mal so lösen, dass das prinzipiell die Biases, ohne jetzt irgendwie in triviale Modelle abzugleiten, die so schlecht sind, dass sie eigentlich gar nix können und nicht mal solche Eigenschaften haben? Das sind diese beiden Dinge. Und natürlich geht es jetzt darum, das in die Anwendung zu bringen, und ein Stichwort ist Few-Shot-Learning: Also wie kann ich es schaffen, dass ich solche Modelle oder solche Gesamtsysteme sehr schnell auf meine Fragestellungen anpasse – auf meine private Fragestellung – also personalisiert? Ich möchte es doch eigentlich an der Hand nehmen, ihm dreimal zeigen, was ich jetzt gerade interessant finde. Dazu bin ich bereit, vielleicht noch zwei, drei Mal zu sagen: Nö, das war jetzt nicht so und das war jetzt gut, und dann soll es doch das eigentlich können. Und dann hat es so ein Niveau erreicht, wo ich sage: Das ist ein Assistent, der bringt mir was und der frisst mir keine Zeit, weil ich die ganze Zeit irgendwie nur erklären soll, was es zu tun hat. Und das sind so die spannenden, herausfordernden Fragen, die weiterhin computerlinguistisch sind. Allerdings eben nicht mehr so in dieser Grundlagenforschung, für das wir damals alle angetreten sind. Von dem her: ja, also wir haben tatsächlich eine Selbstfindungskrise in der Computerlinguistik.

[pgg]: Ist Energieeffizienz auch so ein Thema? Man könnte sich ja vorstellen, dass man das jetzt, keine Ahnung, durch irgendwie günstigere Rechner vielleicht noch ein bisschen runterkriegt, diesen enormen Verbrauch. Aber eventuell gibt es ja auch auf der Ebene der Problemlösung, und damit ja auf der linguistischen Ebene, noch Vereinfachungsmöglichkeiten.

[Biemann]: Ja, vielen Dank, dass Sie das ansprechen. Energieeffizienz ist ein Thema, aber da wird es noch einiges an Jahren dauern, glaube ich, bis wir dort die entsprechenden Maßzahlen finden, welche es uns erlauben, solche Sachen zu vergleichen. Wie gesagt, das ist eine empirische Wissenschaft. Das heißt, wir müssen evaluieren, ob es besser geworden ist. Und wenn wir rein nach Sprachperformanz gucken, ist es relativ einfach. Da haben wir Aufgaben, und die kann ich dann lösen oder nicht. Und da gibt es halt Schwierige und Einfache und so und so viel Prozent habe ich gelöst. Wir haben es noch nicht geschafft, bezüglich Größe der Modelle, Energieeffizienz, Stromverbrauch, wie auch immer, eine Maßzahl zu schaffen, welche so belastbar ist, dass die Leute anfangen, daran ihre Modelle zu vergleichen. Das wird in den nächsten Jahren stattfinden, da bin ich mir sehr sicher. Ich weiß aber nicht, ob die Computerlinguistik da die Treiberin sein wird oder ob es vielmehr eigentlich was ist, was generell im Machine Learning stattfindet. Das kann man nie so sagen. Also zum Beispiel diese Transformerarchitekturen kommen aus der Computerlinguistik.

[mg] Wie ChatGPT eine ist, meinen Sie?

[Biemann]: Ja, wie ChatGPT eine ist. Und die haben aber Anwendungen in allen möglichen Bereichen. Also ich habe das letztens irgendwie bei der Vorhersage von Partikelschauern in physikalischen Detektoren gesehen, dass sie auch diese Architektur anwenden. Was eigentlich aus der maschinellen Übersetzung kommt, ist schon sehr spannend. Von daher ist es echt unklar, wo die Energieeffizienz dann herkommen wird. Vielleicht aus der Bildverarbeitung. Wer weiß.

[mg] Eine Einschätzung von Ihnen, vielleicht auch zum Abschluss. Die große mediale Aufmerksamkeit bekommen ja die Dialogsysteme gerade. In der Wissenschaft, haben Sie beschrieben, will man eigentlich einen ganz anderen Typen von Assistenten oder hat eine größere Bandbreite von Typen. Kann man das so sagen, dass die Vorstellung vom maschinellen Gesprächspartner eigentlich eher eine populäre Vorstellung ist und in der wissenschaftlichen Community die Maschine gar nicht so sehr jetzt als ein Gesprächspartner wahrgenommen wird, sondern stärker als ein Werkzeug? Oder ist das zu platt?

[Biemann]: Es ist relativ platt, aber nicht zu platt. Also ich denke tatsächlich, diese Popularisierung ist dadurch entstanden, dass man mit dem Ding ja wie sprechen oder chatten kann. Also dieses Modell an sich – GPT – gibt es ja auch schon seit drei Jahren. Aber es war eben nicht eingebettet in diese Funktionalität, dass man das leicht benutzen konnte. Es gibt auch Leute, die sagen: Ja, auch für diese wissenschaftlichen Anwendungen möchte ich Dialog haben. Und es gibt auch Leute, die an Dialogsystemen an sich forschen oder an soziotechnischen Systemen und deren Interaktion. Da ist natürlich Dialog ein sehr wichtiges Element. Also gerade in der Wirtschaftsinformatik ist das ein ganz großes Thema, weil die Frage ist: Wie bekomme ich das dann eigentlich in Unternehmenskontexten verortet? Und eine relativ leichte Möglichkeit ist, das irgendwie in den Chat einzubauen, den es zwischen den Leuten eh schon gibt. Also da muss ich irgendwie Anthromorphisierung machen, sonst weiß ich gar nicht, was das für ein komisches System sein soll. Und da bin ich mal gespannt, inwieweit das dann immer mehr in diese Richtung geht. Aber es wird so bleiben, dass ich für Spezialanwendungen nicht alles über Dialogsysteme machen möchte.
[Der Abspann mit Musik beginnt.]

[mg] Damit ist dieses Digitalgespräch zu Ende und wir bedanken uns bei Chris Biemann von der Universität Hamburg für diese spannende Diskussion und die faszinierenden Einblicke. Viele Grüße nach Hamburg! Und wie immer auch vielen Dank an Sie, liebe Zuhörerinnen und Zuhörer, für das Interesse und die Aufmerksamkeit. Und wenn Sie mögen, hören wir uns in drei Wochen wieder, zur nächsten Folge des Digitalgesprächs, dem Podcast von ZEVEDI, dem Zentrum verantwortungsbewusste Digitalisierung.