Talk Science to Me #63: Wie machen wir KI fairer?

2.6.2026 - Birgit Baustädter | Technische Universität Graz

Elisabeth Lex möchte künstliche Intelligenz und Algorithmen fairer und zugänglicher machen. Wie das geht, und wie sich ihre Forschung seit 2023 geändert hat, erzählt sie im Interview.

Dieser Text ist ein Transkript der Podcast-Folge und wurde im Sinne der Verständlichkeit leicht angepasst.

Künstliche Intelligenz arbeitet mit den Daten, die ihr im Training offeriert werden. Diese Daten können aber Ungleichgewichte und Biases von Natur aus enthalten. Wie diese Daten verbessert werden können und insgesamt mehr Fairness in Algorithmen hergestellt werden kann, erforscht Elisabeth Lex. Auch mit ihr habe ich 2023 zuletzt im Podcast gesprochen und sie blickt heute mit mir auf die vergangenenJahre zurück. Mein Name ist Birgit Baustädter und ihr hört Talk Science to Me, den Wissenschaftspodcast der TU Graz.

Talk Science to Me: Wir haben das letzte Mal 2023 miteinander über künstliche Intelligenz und über deine Arbeit in diesem Bereich gesprochen. Du hast dich damals vor allem damit beschäftigt, wie man mittels Daten und KI Aussagen über das menschliche Verhalten machen kann. Hat sich bei dir etwas geändert in der Forschung?

Elisabeth Lex: Also im Kern ist das noch immer mein Ausgangspunkt. Ich schaue auch heute noch darauf, wie man aus Daten etwas über das menschliche Verhalten ableiten kann. Was sich aber noch erweitert hat, ist mein Fokus, wo ich jetzt genau hinschaue. Also ich schaue heute noch viel stärker darauf, für wen diese Systeme geeignet sind, für wen sie eigentlich funktionieren oder auch nicht funktionieren. Und wie wir sie bauen können, sodass sie für alle einen Mehrwert bieten, nicht nur für eine spezifische Benutzer*innengruppe.

Das heißt, ein zentraler Teil von meiner aktuellen Forschung ist eben das Thema Accessibility und Inklusion mit einem speziellen Fokus weiterhin auf Personalisierung und auf Recommender-Systeme. Also auch das ist noch stabil geblieben. Aber es geht jetzt eigentlich auch darum, dass man sagt, welche Daten fehlen den Algorithmen grundsätzlich jetzt in Richtung Accessibility und Inklusion, damit sie überhaupt sinnvoll funktionieren können, zum Beispiel für Leute, die eine Bewegungseinschränkung haben oder für Leute, die neurodivers sind. Es geht weiterhin darum, welche Gruppen von Menschen sind unterrepräsentiert in Daten und Algorithmen oder welche anderen überrepräsentiert.

Als konkretes Beispiel habe ich etwas aus meiner Forschung mitgebracht, da geht es um Point of Interest Recommender Systeme. Das ist etwas, das kennt man wahrscheinlich aus Google Maps oder aus jeder Art von Online-Kartendienst. Das heißt, man möchte gerne bestimmte Orte in einer Stadt, wo ich jetzt als Tourist*in beispielsweise unterwegs bin, anschauen und möchte da die besten Vorschläge bekommen. Funktioniert grundsätzlich auch ganz gut. Aber das Problem ist, sobald ich eine Einschränkung habe, was jetzt die Zugänglichkeit betrifft, funktioniert es schon nicht mehr so gut. Also diese Orte haben oft eine Kennzeichnung, ob sie geeignet sind für Rollstuhlfahrer*innen, aber nicht unbedingt jetzt genau wie, was das jetzt genau bedeutet. Ist es jetzt der Eingang Rollstuhlfahrer*innen geeignet, ist es das WC, ist es der Parkplatz, wie auch immer. Und meistens

ist auch diese Abdeckung nicht so ganz gut. Es gibt aber auch noch andere Anforderungen wie zum Beispiel sensorische Belastungen, also wenn jemand sehr lärmempfindlich ist, sehr lichtempfindlich, dann ist das eigentlich eine Information, die steckt in den Daten ganz schlicht einfach nicht drinnen. Und wenn ich jetzt aber diese Einschränkungen habe, dann kann ich mit den Empfehlungen wenig anfangen. Das Gleiche ist, wenn ich jetzt Rollstuhlfahrer*in bin und nicht genau weiß, ist dieses Restaurant, das ich mir rausgesucht habe, wirklich komplett barrierefrei, dann kann ich da eigentlich nicht hinfahren, weil wenn ich dann vor der Tür stehe und ich komme da nicht rein, habe ich den Weg umsonst auf mich genommen. Das heißt, diese Algorithmen, ja, da gibt es schon eine systematische Exklusion, sage ich einmal, von Menschen, die nicht in diesen Standardbedürfnisse, die Algorithmen immer wieder so annehmen und von einem, wie man zu sein hat, hineinpassen. Da schaue ich jetzt noch viel, viel verstärkter hin. Wie kann man diese Probleme transparent machen und wie kann man die Algorithmen auch verändern? Wie kann man die Datenlage positiv beeinflussen.

Personalisierungsalgorithmen sind ja Tools, die sehr hilfreich sind, aber gleichzeitig binden sie unsere Aufmerksamkeit enorm. Weil sie grundsätzlich darauf optimiert sind, dass Menschen möglichst lange in einem System gehalten und dabei möglichst gut unterstützt werden . Aber das funktioniert auch nicht für jeden gleich gut. Dazu sagt man Engagement, wenn man so lange in einem System drinnen bleibt. Wenn man sich jetzt beispielsweise neurodiverse Personen anschaut, und das kann jetzt ein ganz breites Spektrum sein von ADHS oder Autismus oder ich habe einfach irgendwas auf diesem Spektrum, was mich jetzt halt auch auszeichnet, dann sieht man schon, dass dieses Engagement durchaus negative Konsequenzen haben kann. Wir haben da auch eine Studie vor kurzem durchgeführt, wo wir einfach verglichen haben, so Engagement-basierte Recommender-Systeme und so im Bereich Short-Content-Videos, man kennt das von TikTok, YouTube-Shorts und so weiter. Und dann sollten die Leute eben berichten, wie es ihnen geht mit ihrem Zeitmanagement, wenn sie so eine Plattform angefangen haben zu schauen, wie es ihnen geht mit ihren Emotionen, ob sie sich dadurch positiver oder negativer fühlen, wie es mit ihrer generellen Aufmerksamkeitspanne ist. Und da sieht man, diese Systeme haben negative Auswirkungen auf alle.Es ist so, dass durchwegs die Leute berichten, dass ihnen emotional oft schlechter geht danach, dass sie sich schwer tun die Plattform auszuschalten, aber die neurodiversen Personen, die tun sich einfach verhältnismäßig so viel schwerer, da das eigene Zeitmanagement im Blick zu behalten und eben auch die Emotionen zu kontrollieren. Das heißt, sie sind stärker betroffen von einer Designentscheidung, die so Plattformen für ihre Algorithmen getroffen haben. Und das ist auch etwas, wo ich mich jetzt verstärkt damit beschäftige, weil es eben nicht diesen einen Standardmenschen gibt. Menschen haben ganz unterschiedliche Bedürfnisse. Wie können wir unsere Algorithmen so bauen, dass diese unterschiedlichen Bedürfnisse unterstützt werden und nicht da etwas verschlechtert wird oder sich die Leute dann wirklich auch schlechter fühlen, wenn sie das System benutzen? Das heißt, es geht jetzt nicht mehr nur um Voraussagen von Präferenzen, sondern es geht darum zu schauen, für wen ist das wie gut und wie muss ich das System verändern, damit das für alle gut bleibt.

Die Professur Human-Computer Interfaces and Inclusive Technologies von Elisabeth Lex wird vom Verein „Erzherzog-Johann-Gesellschaft Initiativ für Kinder und Jugendliche mit Behinderungen“ finanziell unterstützt.

Wie kann ich in so einen Algorithmus so eingreifen? Also welche Möglichkeiten gibt es da dazu zu verbessern?

Lex: Ich kann das Ziel des Algorithmus verändern, das heißt die sogenannte Optimierungsfunktion, weil eigentlich sind es Algorithmen-Methoden, wo man etwas optimiert. Das heißt man optimiert zum Beispiel auf Genauigkeit, man könnte aber auch auf Diversität optimieren. Oder auf ich möchte was haben, was überraschend für mich ist. Das heißt Serendipitous Recommendation dann in dem

Fall. Das heißt ich muss mir überlegen, und das ist im Wesentlichen eine mathematische Funktion, eine Optimierungsfunktion, wie ich dieses Ziel, diese Optimierung beschreibe, so dass es eben keine negativen Konsequenzen hat. Was in dem Bereich Short-Content-Recommender-Systeme ganz üblich ist, ist, dass man extern Trigger setzt. Es gibt ganz viele Tools und Apps, die sagen, okay, nach einer Stunde kriegst du einen Reminder. Jetzt bist du eine Stunde in dem System. Oder wenn du eine gewisse Anzahl von Videos angeschaut hast, kriegst du einen Reminder. Und da haben wir auch so User*innen dazu befragt, wie sie das empfinden und ob sie das als hilfreich empfinden. Und interessanterweise sagen da viele, sie empfinden es nicht als hilfreich. Erstens schauen sie vielleicht dann trotzdem länger. Und zweitens fühlen sie sich bevormundet durch so einen externen Trigger. Das heißt, man muss das eigentlich subtiler machen. Und das ist aktive Forschung, sage ich einmal. Da gibt es meiner Meinung nach keine richtige Lösung. Es gibt einmal so Tools, aber die sind alle nicht personalisiert. Das ist so ein Timer für alle. Und das wird dann oft als paternalistisch empfunden und nicht unbedingt als hilfreich. Das ist auch etwas, wo ich mich gerade mit meiner Forschungsgruppe dahin stark beschäftige, wie man solche Interventionen setzen kann, sodass die dann auch als hilfreich wahrgenommen werden.

Das heißt, in deinem Bereich hat sich sehr viel getan. Wie schaut es mit der Forschungscommunity im Großen aus? Gerade im Bereich KI hat sich ja sehr, sehr viel, sehr, sehr rasch verändert. Welche neuen Möglichkeiten gibt es da, aber auch welche neuen Hürden?

Lex: Ja, es hat sich wirklich viel verändert, das muss man echt sagen, in den letzten paar Jahren. Und es wird sich auch noch viel verändern, da bin ich mir auch ganz sicher. Die größte Veränderung ist sicher die Dominanz von sogenannten Foundation Models. Das sind diese großen Modelle der Welt, sage ich einmal. Man kennt das unter Large Language Models auch. Das sind so diese Algorithmen und Modelle, die hinter Tools wie ChatGPT beispielsweise stehen. Das heißt sogenannte generative KI-Systeme. Das heißt, die haben sehr viele Daten gesehen in ihrem Training und generieren darauf Neues, was ähnlich ist zu dem, was sie gesehen haben, aber es ist halt neu. Und dadurch, dass die auf so vielen Daten gelernt wurden, haben die echt wirklich breite Repräsentationen von verschiedenen Problemstellungen drinnen. Und damit haben sie auch viele Fragestellungen verschoben. Das heißt, früher hat man eigentlich so einzelne Modelle für so abgegrenzte Fragestellungen entwickelt und die dann auch arbeiten lassen. Jetzt haben wir eher so generalistische Systeme, die man aber auf Kontexte anpasst. Das ist der aktuelle Stand, sage ich einmal.

Es gibt natürlich sehr viele Möglichkeiten, weil man hat dann so ein "Weltwissen" in so einem Modell und adaptiert das dann auf den eigenen Task. Aber es hat natürlich auch Hürden wieder, sage ich einmal. Weil wir haben grundsätzlich noch komplexere Systeme, noch weniger Transparenz in den Systemen, es ist noch schwerer reproduzierbar, warum da ein gewisser Output gekommen ist. Und wir sind extrem von Daten und von Ressourcen abhängig.Also einerseits wirklich riesige Datenmengen und andererseits braucht man schon noch diese großen Datenzentren, um diese Modelle zu trainieren und auch dann laufen zu lassen. Das heißt, ja, viele Möglichkeiten, aber auch viele neue Herausforderungen.

Das, was ich vorhin angesprochen habe, dass es diese Datenlücken gibt, das ist weiterhin das gleiche Thema. Das ist noch immer das gleiche Problem. Bestehende Datenlücken werden durch solche Modelle auch weiterhin verstärkt. Man sieht es im Kontext von zum Beispiel Sprache, weil so Übersetzungstools sind ganz stark auch von diesen Foundation Models geprägt, so was wie Google Translate oder DeepL, die funktionieren ja auch sehr gut, sage ich einmal, für die meisten Sprachen. Aber sobald ihr jetzt in eine Sprache geht, die von weniger Menschen gesprochen wird und wo es weniger Texte gibt, auf denen die KI trainiert werden kann, dann funktioniert das einfach schon wieder nicht mehr so gut. Das bring natürlich auch sehr viele Probleme, weil diskutiert wird, solche Übersetzungssoftware zum Beispiel im Bereich Asyl einzusetzen, als automatische

Übersetzungssoftware, damit man sich die Übersetzer*innen quasi sparen kann oder vielleicht auch halt einfach unterstützen kann. Aber dann hat man das Problem, dass es gesprochene Sprache ist, mit Dialekten, mit Ausdrucksweisen, die vielleicht jetzt nirgends niedergeschrieben sind. Dann hat man das Problem, dass es eben dadurch, weil es so wenig Daten gibt, die KI das nicht so gut verstehen kann und so weiter und so fort. Das heißt, neue Biases sind auch dort wieder da. Also von dem her gibt es natürlich viele Möglichkeiten, aber auch viele Probleme.

Und dann das Dritte, was eine Herausforderung ist, dass man eben auch sich überlegt, was heißt das jetzt, wenn ich KI-Systeme irgendwo einsetze, welcher Art von regulatorischen Aufwand muss ich da jetzt noch on top betreiben, zusätzlich zu der technischen Challenge.

Was sind denn aus deiner Sicht aktuell die großen Forschungsfragen?

Lex: Es sind sehr interdisziplinäre Fragen, denke ich, die mich am meisten umtreiben. Also wie bauen wir Systeme, die nicht nur leistungsfähig sind, sondern die auch zugänglich sind für eine breite Bevölkerung, ohne dass man dieser Bevölkerung schadet mit dem, was man da anbietet. Das betrifft die Daten, die man zur Verfügung hat, wie ich schon gesagt habe, die Modelle, aber auch die Interfaces, die User Interfaces und die Designentscheidungen, die man schon trifft, wenn man Technologien eben Menschen zur Verfügung stellt. Also diese Chatbots zum Beispiel, die wirken ja immer sehr überzeugend. Und die sind ja rein von der Designentscheidung auch so, man schreibt so seine Anfrage rein und dann tröpfeln so die Antworten zurück. Da wird nicht auf einmal ein Text zurückgegeben, was man tun könnte. Sondern es kommt so Wort für Wort. Als würde man mit Freunden schreiben. Das ist eine Designentscheidung, damit das Ganze humanisiert wird, meiner Meinung nach. Heißt aber nicht unbedingt, dass das jetzt dafür alles gut ist. Also ich denke, das ist eine große Herausforderung. Und wir müssen eben verstehen, wo diese Datenlücken ein Problem sind und wie man diese Datenlücken in gewisser Weise schließen kann. Und das Optimieren auf Engagement, das habe ich vorher auch angesprochen, das ist vielleicht nicht das Beste. Ich verstehe, warum Firmen das tun, weil sie natürlich auch ihre Werbung verkaufen wollen, die Leute möglichst lange in ihrem Ökosystem halten wollen. Aber aus gesellschaftlicher Sicht sollte uns das Wohlbefinden der Menschen schon auch ein großes Anliegen sein. Und das jetzt abzuwägen, dass man einerseits diese geschäftlichen Ziele erreicht und andererseits eben dieses Wohlbefinden und die Sicherheit der Personen, die die Systeme nutzen, im Blick behaltet, ist aus meiner Sicht auch eine große Forschungsfrage.

Was mir auch noch als große Forschungsfrage einfällt, ist der Zugang zu Informationen. Wir leben in einer Welt, wo Zugang zu Information immer stärker durch KI mediiert wird. Also wir sehen das jetzt in Suchinterfaces, es gibt eine AI-Summary in einem PDF-Reader. Viele Leute benutzen eben irgendwelche Chatbots, um ihre Fragen zu stellen. Und man hat irgendwie so das Gefühl, man hat so super tollen Zugang zu Information, aber gleichzeitig ist der eben stark mediiert durch die Plattformen, die diese Informationen anbieten. Ich habe mal irgendwann einen Spruch gehört, dass in ein paar Jahren die Menschen reich sind, wenn die noch ein Brockhaus im Regal stehen haben, weil quasi die Information dann ihnen gehört. Also ja, das ist aus meiner Sicht auch etwas, wo man sich noch viel Gedanken drüber machen muss, weil gerade mit Informationen wird ja auch so viel Missinformation betrieben. Und wenn da die Informationen in den Händen weniger sind, dann ist das natürlich eine große, große, könnte es eine große Manipulationsmaschine auch werden.

Du hast 2023 auch davon gesprochen, dass es immer schwerer wird für euch in der Forschung, an Daten zu kommen, vor allem jetzt von den Social Media Plattformen. Wie ist denn das heute? Wie hat sich das entwickelt?

Lex: Ja, das ist tatsächlich so. Der Zugang zu Daten ist wirklich sehr restriktiv mittlerweile, weil die Plattformen, welcher Art auch immer, typischerweise sehr genau kontrollieren, welche Daten sie zur

Verfügung stellen und welche nicht und unter welchen Bedingungen. Da geht es oft um viel Geld, sage ich einmal. Ich verstehe auch die Plattformen, die wollen ja auch nicht unbedingt gratis Trainingsdaten für Large Language Models oder Foundation Models anbieten. Das ist schon auch so ein Grund, warum das passiert ist. Aber aus der Perspektive der Forschung ist es natürlich schon schwer. Also wir müssen halt jetzt schauen, welche alternativen Datenquellen wir benutzen können für unsere Forschungsfragen.

Der Trend geht immer mehr dazu, dass man auch Simulation verwendet, also dass man versucht, sich ein eigenes Modell von einer Umgebung zu machen und das möglichst genau abbildet und mit simulierten Daten arbeitet oder ebenso partizipative Ansätze, dass man sagt, man ruft jetzt diese Bevölkerungsgruppe auf, die man jetzt untersuchen möchte oder für die man jetzt einen Algorithmus beispielsweise bauen möchte und bietet sie um eine Datenspende. Oder man bindet sie in den Prozess ein, sodass die Menschen die Möglichkeit haben, auch freiwilligere Daten für die Forschung zur Verfügung zu stellen. Aber es ist auf jeden Fall ein großes Problem. Es gibt in Europa schon so regulatorische Ansätze, die hilfreich sein können, weil es gibt ja den Digital Services Act und dort gibt es die Möglichkeit, dass wenn man eine Forschungsfrage hat, wo man sagt, man vermutet ein Problem in einer Plattform, dass man dann auch quasi einen Antrag einreichen kann und dann Daten von der Plattform zur Verfügung gestellt bekommt, um diese Forschungsfrage zu untersuchen. Aber der Prozess ist sehr komplex. Wir haben ihn in meiner Forschungsgruppe noch nicht durchexerziert, wir haben es vor, wir beobachten noch ein bisschen, wir haben uns diesen Prozess schon sehr genau angeschaut und ja, ist auch die Frage, wie das dann nachher wirklich umgesetzt werden kann. Aber das ist ein großes Problem.

Hat sich an diesem Zugang durch die Beliebtheit von Large Language Models irgendwas verändert und wie ist es auch mit den LLMs zu arbeiten in dem Bereich?

Lex: Der Zugang hat sich auf jeden Fall verändert, hat sich verschlechtert durch LLMs. Genau deshalb, weil eben die LLMs ja auf großen Datenmengen trainiert werden und die in den Anfangsjahren einfach runtergesogen wurden von irgendwelchen Plattformen.Und dann haben die Plattformen sich schon gewehrt, dass sie da jetzt eigentlich nicht einfach nur ein Trainingsdatenlieferant sein wollen, gratis für OpenAI oder welche Firma auch immer. Man kann natürlich Large Language Models schon auch wieder dazu benutzen, um Daten zu generieren für Simulationen. Das funktioniert auch nicht schlecht, sage ich einmal. Je nachdem, was man für Fragestellungen hat. Also es ist von dem her irgendwie so ein bisschen ein zweischneidiges Schwert. Man muss halt immer sehr vorsichtig sein, weil es ist natürlich auch nur eine Abbildung der Realität, so ein Large Language Model und wie Daten diese generieren. Aber wenn man es vorsichtig mit dem Wissen, dass es eben Bias ist und alle möglichen Probleme da gibt, tut, dann kann man das schon machen. Also tun wir auch, tun wir auch. Es ist halt irgendwie so ein Behelfsmittel, sage ich einmal.

Wie siehst du den Bias in Daten heute? Wir haben ja damals darüber gesprochen, dass du daran arbeitest, dass eben genau diese Biases verschwinden oder dass man die irgendwie geringer macht in den Daten,dass es da Methoden gibt. Wie sich das entwickelt?

Lex: Ja, ich meine, das Problem der Biases besteht auf jeden Fall weiter. Gleichzeitig gibt es schon jetzt auch bessere Werkzeuge, damit man mit dem umgeht. Als wir das letzte Mal gesprochen haben, haben wir ja über den Bereich Musik und Popularity Bias und diese Themen gesprochen, dass es eben so Nischen-Musik gibt, die schwer repräsentiert wird und schwer empfohlen wird, weil es einfach wenig Daten gibt. Aber jetzt kann man mit diesen großen Modellen schon auch bessere semantische Verständnis von den User*innen-Präferenzen erhalten. Das ist natürlich gut, wenn man jetzt wenig über die User*innen weiß. Also so technische Fortschritte gibt es schon einige. Oder man hat eben multimodale Modelle. Das heißt nicht nur, dass man jetzt sagt, man hat Text oder Interaktionsdaten, sondern vielleicht auch Bilder, das Audiosignal von einem Musiktrack oder von

einem Video und auch zusätzliche Informationen. Das heißt, ich habe einfach mehr Input aus verschiedener Perspektive. Und damit ist es schon noch besser, sage ich einmal, auch so gebiasede Daten zu repräsentieren. Also gerade wenn man sehr spezifische Situationen hat. Ich würde jetzt nicht unterschreiben, dass das für jedes Szenario geht, überhaupt nicht. Aber per se ist auch der technische Fortschritt in den Modellen da durchaus positiv zu bewerten, aus meiner Sicht, was das Thema Bias betrifft. Nicht die Lösung, aber ein Schritt.

Du hast den Musikgeschmack eben auch abseits von Populärmusik jetzt schon, also von sehr populärer Musik erwähnt und du hast dich damals auch mit Wahlen beschäftigt und wie Social Media Wahlen beeinflussen kann. Wie hast du das Gefühl, dass sich das entwickelt hat? Werden die Leute heute mehr oder weniger beeinflusst? Hat sich da Awareness entwickelt oder ist das jetzt stärker geworden? Was hat sich da verändert?

Lex: Ja, ich würde schon sagen, dass der Einfluss größer geworden ist, aber er ist weniger sichtbar, also er ist weniger leicht zu fassen. Das liegt schon auch daran, dass eben diese Systeme, die da das Potenzial haben zum Beeinflussen, diese KI-Systeme,diese algorithmischen Systeme einfach in so vielen Interfaces integriert sind. Und das nicht unbedingt nur in so einem klassischen Feed, so wie es halt früher war, so eine Liste von Recommendations, eine Liste von Suchergebnissen. Da sind halt welche drinnen, die sind gebiased, sondern auch in diesen Assistenten, in den Chatbots, in den generierten Antworten sind da Biases drinnen und das ist oft nicht so ganz leicht zu fassen, sage ich einmal. Ja, also von dem her würde ich sagen, größer geworden und schwerer zu mitigieren, leider.

Siehst du da eine Chance, dass sich das wieder verändert?

Lex: Ja, sicher, natürlich. Das einzige was fix ist, ist die Veränderung, oder? Ich denke, auch da ist Regulierung ein Thema, das nicht unwichtig ist. Da ist ein generelles Bewusstsein ein Thema, schon ein bisschen ein Druck auf Plattformen und Informationsanbieter*innen, der da notwendig sein muss, dass man da eine breitere Perspektive abbildet. Also das ist jetzt nicht unbedingt früher. Aber aus meiner Sicht war, als man diese Listen gehabt hat und sich dann anschauen kann, was klicke ich an, was klicke ich nicht an. Da war die Verantwortung noch ein bisschen mehr beim Individuum, meiner Meinung nach, als es jetzt ist. Jetzt ist es aus meiner Sicht eine große Verantwortung von den Anbietern und die muss man dann auch in die Pflicht nehmen, dass sie da schon auch gewisse Standards einhalten. Man sieht ja jetzt diese Debatte, in den letzten Monaten hat man ja ganz viel gesehen, was das Erstellen von Missinformationsinhalten betrifft und gerade aus dem Amerikanischen ist es oft eine Debatte zwischen Meinungsfreiheit und Zensur, aber das ist eine sehr verknappte Diskussion, sage ich einmal. Ich denke, Europa sollte für sich da ein eigenes Wertebild finden und es auch umsetzen. Es ist nicht so verknappt, dass es entweder Zensur ist oder Meinungsfreiheit. Ich glaube, das ist nicht unbedingt ein technisches Problem, sage ich einmal. Das ist ein gesellschaftliches Problem.

2023 hast du KI vor allem noch als Werkzeug gesehen, das Menschen unterstützen kann. Es wird ja zunehmend, dass es Angst gibt, dass KI Jobs ersetzen kann. Wie siehst du das heute?

Lex: Ja, primär sehe ich KI noch immer als Werkzeug, muss ich sagen. Aber es hat schon tiefgreifende Auswirkungen auf die Arbeitswelt in ganz unterschiedlichen Berufen. Meiner Erfahrung nach ist es so, dass man KI-Tools, welcher Art auch immer, sehr gut nutzen kann, wenn man schon eine gewisse Expertise hat. Dann wird man effizienter, dann wird man schneller.

Man sieht das beim Programmieren. Es gibt schon wahnsinnig gute Unterstützungen, was das Programmieren betrifft, aber ich kann es eigentlich nur dann sinnvoll einsetzen, wenn ich schon ein gewisses Grundverständnis und Grundgerüst habe und Programmierfähigkeiten. Das unterschätzen unsere Studierenden manchmal. Die Code-LLMs machen das Gleiche wie die Sprach-LLMs. Die

generieren einfach etwas, basierend auf Statistiken, basierend auf Zusammenhängen. Die schreiben auch lange Codestücke hin und dann plötzlich steht was drin, was man eigentlich gar nicht kennt. Wenn man sich nicht auskennt, dann wird man das nie finden. Also ich denke schon, dass es eben viel Potenzial hat für Expert*innen, dass die noch schneller werden und noch stärker werden. Aber es gibt auf jeden Fall Tätigkeiten, die ersetzt werden.

Man sieht das jetzt schon in den Firmen, dass die teilweise eine AI-first-Policy fahren und schauen, was kann die KI erledigen und wofür brauche ich überhaupt noch Personal. Beziehungsweise ich kenne auch Unternehmen, jetzt nicht unbedingt in Österreich, aber internationale Unternehmen, wenn da jetzt Software produziert wird, dann müssen sie das teilweise sogar genehmigen lassen, wenn sie es ohne KI machen wollen, weil das einfach mehr Zeit braucht. Also von dem her, wir werden auf jeden Fall sehen, dass es Automatisierung gibt, dass sich Aufgaben verschieben, dass sich Aufgaben neu verteilen. Und das macht sicher auch Probleme. Wie sich das auswirken wird, ist, glaube ich, schwer zu sagen. Ich bin grundsätzlich immer ein optimistischer Mensch. Ich denke, wir sollten diese Werkzeuge nutzen. Wir sollten das Beste daraus machen. Wir sollten damit auch komplexere Fragestellungen angehen, als wir das vielleicht bis jetzt machen haben können. Und wir sollten eben schauen, dass wir die Systeme, die wir dann auch verwenden, so gestalten, dass sie Menschen unterstützen und ganz unterschiedliche Menschen unterstützen in unterschiedlichen Aufgaben. Und ich denke, was ganz wichtig ist, ist die Bereitschaft, sich ständig weiterzubilden, die Bereitschaft, sich weiterzuentwickeln, die Bereitschaft, nicht stehen bleiben. Auf keinen Fall stehen bleiben. Und meiner Meinung nach ist eigentlich jetzt eine Zeit, wo, obwohl wir ja diesen ganzen Zugang zu Informationen haben, wo man selbst ganz viel wissen soll und muss. Das ist vielleicht sogar noch mehr. Also sich nicht aufs Nachschauen verlassen, sondern selbst arbeiten und sich dann unterstützen lassen.

Wenn du in die Zukunft blickst, wo kann das Ganze noch hingehen?

Lex: Ja, ich glaube, ich sehe schon viel Potenzial in so personalisierten Systemen, die einen wirklich unterstützen können .Natürlich, weil ich selber auch personalisierte Sicherungsforschung betreibe. Aber eben so Systeme, die nicht nur effizient sind aufgrund von irgendwelchen Businessmetriken, sondern auch eben unterschiedliche Bedürfnisse gut adressieren können. Also es ist nicht immer nur alles schwarz und weiß, wir werden alle ersetzt, sage ich einmal. Sondern es ist eigentlich so, dass wir Fähigkeiten haben, die man unterstützen kann. Ich habe jetzt gerade am Wochenende ein paar Videos angeschaut, in denen es um Exoskelette im asiatischen Bereich geht. Und wenn du dann siehst, dass es Menschen gibt, die plötzlich wieder aufstehen können, weil sie Exoskelette haben, dann ist das einfach eine wunderschöne KI-Anwendung. Ich glaube, man soll schon den Fokus egen, dass man jetzt wirklich sagt, da gibt es Probleme, die man adressieren kann.

Aber es ist eben wichtig, dass wir uns überlegen, was ist das Ziel von so einem System? Ist es eben zum Beispiel Engagement oder nicht? Diese Entscheidung muss man treffen. Ist es Wohlbefinden oder nicht Wohlbefinden? Ist es Autonomie oder nicht? Und ich denke, das ist dann so ein wertgetriebener Ansatz. Und wenn diese guten Werte da ja schon repräsentiert werden, dann kann die ganze KI-Richtung in eine sehr gute Richtung gehen, sodass es wirklich uns allen besser geht. Also es liegt einerseits an der Technologie selbst, ob wir das Potenzial nutzen können, aber auch an den Designentscheidungen, die wir treffen, an den Daten, an den Rahmenbedingungen, die wir als Menschen und als Gesellschaft setzen sollen.

Vielen Dank für das Interview.

Lex: Vielen Dank für die tollen Fragen.

Vielen Dank, dass ihr heute wieder mit dabei wart. In der nächsten Folge spreche ich mit

Alexander Felfernig vom Institute of Software Engineering and Artificial Intelligence.

https://www.tugraz.at/news/artikel/talk-science-to-me-63-wie-machen-wir-ki-fairer