

Deutscher Ethikrat – Web-Event

KI im Klassenzimmer – Ethische Fragen zu ChatGPT und Co.

Transkript der Veranstaltung

13. Dezember 2023, 17:30–19:00 Uhr

Online

Programm

Begrüßung und thematische Einführung	2
Elisabeth Gräß-Schmidt · Mitglied des Deutschen Ethikrates	
Was sagt der Deutsche Ethikrat zu KI und Bildung?	3
Judith Simon · Mitglied des Deutschen Ethikrates	
Was können ChatGPT und Co. – und was nicht?.....	6
Doris Weßels · Fachhochschule Kiel	
Wie wird in Schulen auf die neuen Angebote reagiert und wie werden sie genutzt?	12
Felicitas Macgilchrist · Carl von Ossietzky Universität Oldenburg	
Diskussion.....	16
Moderation: Elisabeth Gräß-Schmidt · Mitglied des Deutschen Ethikrates	

Hinweis: Bei dem folgenden Text handelt es sich nicht um eine wörtliche Transkription. Der Text wurde lektoriert, um eine gute Lesbarkeit herzustellen.

Begrüßung und thematische Einführung

Elisabeth Gräß-Schmidt · Mitglied des Deutschen Ethikrates (00:01)

Sehr geehrte Referentinnen, sehr geehrte Damen und Herren im Livestream, herzlich begrüße ich Sie zu unserem heutigen Web-Event zum Thema „KI im Klassenzimmer. Ethische Fragen zu ChatGPT und Co.“ Mein Name ist Elisabeth Gräß-Schmidt, ich bin Mitglied im Ethikrat und führe Sie heute Abend durch die Sendung.

Bildung ist gegenwärtig in aller Munde und hochaktuell. Letzte Woche eröffnete die neue PISA-Studie ein erschreckendes Ergebnis: Das Bildungsniveau von Schülerinnen und Schülern ist auf dem niedrigsten Niveau seit der Messung der ersten PISA-Studie im Jahr 2000. Es hapert beim Lesen und Schreiben, und die Frage ist: Was tun? Kann Digitalisierung hier eine Lösung bieten, wie es der Digitalpakt von 2019 vorsieht? Kann KI [künstliche Intelligenz] dieser Bildungsmisere abhelfen, eine Wende erreichen, indem sie Bildungsprozesse unterstützt? Oder behindern die Werkzeuge der KI wie ChatGPT Lernerfolge oder Kreativität? Deren Leistungen sind jedenfalls kritisch und differenziert zu betrachten. Gleichwohl können wir uns ihren Herausforderungen nicht entziehen und dürfen gegebenenfalls nicht davor zurückschrecken, einem Kulturwandel ins Auge zu blicken. So hat der Bildungsexperte Gottfried Böhme jüngst gewarnt: „Das strukturkonservative deutsche Bildungssystem steht vor einer Zäsur, die weit über die Erfindung des Taschenrechners hinausgeht. Künstliche Intelligenz bricht der Schule, so wie wir sie heute kennen, das Rückgrat.“ Starke Worte, die uns fragen lassen, ob unser Bildungssystem generalüberholt werden muss.

Der Ethikrat hat sich in seiner jüngsten Stellungnahme zum Thema „Mensch und Maschine“ bereits zu den Herausforderungen durch KI auch im Bereich der Bildung geäußert. Dabei waren wir von einer Leitlinie für die schulische Bildung ausgegangen. Zu ihr gehört zwar auch die Aneignung und Nutzungsfähigkeit von Wissen. Diese sind aber immer vor dem Hintergrund eines grundlegenden Bildungsziels zu betrachten. Aus ethischer Perspektive sieht der Ethikrat dieses Ziel in der Förderung und der Entwicklung junger Menschen zu mündigen, urteils- und verantwortungsfähigen Personen. Damit geht es immer darum, die Handlungsfähigkeit des Menschen zu erweitern und nicht zu vermindern. Dieses emanzipative Ziel ist kritischer Maßstab, an dem sich das Lehren und das Lernen in der Schule orientieren sollte.

Noch nicht im Blick unserer Stellungnahme war das gut vor einem Jahr lancierte Large-Language-Model ChatGPT 3, das gleich von vielen ausprobiert wurde, gerade auch im Bildungsbereich, an Universitäten und Schulen, und dort für Beunruhigung sorgte. Ein Chatbot, der in Sekunden schnelle Referate schreibt, Aufgaben löst, Fragen beantwortet, Bilder herstellt, Wissen liefert – muss dieses Werkzeug nicht verboten werden?, so manch überstürzte Reaktion. Denn Prüfungsformate wie Hausarbeiten, Referate, Lerntests – das ist so nicht mehr möglich.

Schnell wurde aber klar: Wir können ChatGPT nicht einfach verbieten. Es ist ein neues technisches Werkzeug, das, hat es einmal Eingang in unsere Gesellschaft gefunden, nicht einfach wieder verschwindet. Und das soll es vielleicht auch nicht. Denn Technologien sind immer ambivalent: Sie können nutzen oder schaden oder beides. Können Werkzeuge generativer KI uns vielleicht sogar den ethischen, sozialen und politischen

Zielvorstellungen von Bildung wie Bildungsgechtigkeit, Inklusion und Demokratiefähigkeit näherbringen?

Es gilt also, im Blick auf ChatGPT wie allgemein auf KI kritisch diese Ambivalenz wahrzunehmen und möglichst nur die Chancen mitzunehmen und die Risiken zu minimieren. Dieser konstruktiven und kritischen Sicht sind die folgenden Impulse verpflichtet.

Wir hören zunächst den Vortrag von Judith Simon. Judith Simon ist Mitglied des Deutschen Ethikrates und Professorin für Ethik in Informationstechnologien an der Universität Hamburg. Sie war AG-Sprecherin der Stellungnahme zu Mensch und Maschine und spricht zum Thema: Was sagt der Deutsche Ethikrat zu KI und Bildung?

Was sagt der Deutsche Ethikrat zu KI und Bildung?

Judith Simon · Mitglied des Deutschen Ethikrates (05:45)

(Folie 1, 2)

Vielen Dank. Ich habe genau das vor, was gerade angekündigt ist: Zuerst werde ich kurz sagen, was wir in der Stellungnahme zu Mensch und Maschine geschrieben haben, vor allem mit Blick auf schulische Bildung, und dann fragen: Gibt es neue Herausforderungen durch ChatGPT, die wir in der Stellungnahme noch nicht so im Blick hatten? Und wenn ja, welche sind das?

(Folie 3)

Zunächst zum Hintergrund der Stellungnahme. Im März 2020 wurde durch die damaligen Bundestagsfraktionen und den Bundestagspräsidenten Schäuble eine Anfrage an uns gestellt, uns mit

dem Thema Mensch und Maschine, künstliche Intelligenz zu beschäftigen.

Im August 2020, ein bisschen verzögert aufgrund der Coronapandemie, haben wir die Arbeitsgruppe eingesetzt. Inhaltlich haben wir die Befassung mit dem Thema im November 2022 abgeschlossen. Das war genau der Monat, in dem ChatGPT auf den Markt kam.

Die Stellungnahme wurde im März dieses Jahres veröffentlicht. Sie ist mit fast 300 Seiten sehr umfangreich. Wir haben sie in drei Teile gegliedert:

Der erste Teil beschäftigt sich mit den Grundlagen und wir erklären unseren Schwerpunkt: die ethische und anthropologische Grundorientierung. Im zweiten Teil haben wir uns dann vier Anwendungsbereiche angeschaut, einer davon ist Bildung. Im dritten Teil haben wir zehn Querschnittsthemen identifiziert, also Themen, die bei allen Sektoren immer wieder aufgeschieden sind, wenn auch in unterschiedlicher Art und Weise.

(Folie 4)

Der Fokus war: Was passiert, wenn wir Tätigkeiten, die zuvor Menschen getan haben, an Maschinen delegieren? In welcher Art und Weise erweitert oder vermindert dies die Handlungsfähigkeit unterschiedlichster Beteiligter?

(Folie 5)

Wir haben uns den Einsatz oder die Verwendung von KI in vier Anwendungsbereichen angeschaut: in der Medizin, der schulischen Bildung, der öffentlichen Kommunikation und Meinungsbildung sowie der öffentlichen Verwaltung. Im Folgenden möchte ich kurz auf den Bildungsbereich eingehen.

(Folie 6)

Wir haben uns vor allen Dingen drei Systemarten angeschaut: erstens intelligente Tutorsysteme, die

eine lange Tradition im Einsatz von KI im Bildungsbereich haben, zweitens Telepräsenzroboter und drittens Classroom Analytics, das heißt, die Verwendung von KI im Sinne von Mustererkennung auch im Monitoring von Klassenräumen, unter anderem durch Audio- und Videoanalysen.

(Folie 7)

Auf Basis dieser Analysen haben wir dann von diesen drei Feldern Schlussfolgerungen gezogen. Das ist zum einen der Hinweis darauf, dass es auch um die Bildung der Persönlichkeit geht, dass Digitalisierung also kein Selbstzweck sein soll, sondern sich immer an den Zielen von Bildung ausrichten muss.

Wir haben auch analysiert, in welcher Art und Weise sich die Grade der Ersetzung des Menschen oder menschlicher Tätigkeiten bei unterschiedlichen Technologien unterscheiden.

Dann haben wir beurteilt, welche Auswirkungen dieses Delegieren, dieses Ersetzen auf die Handlungsfähigkeit hat (wir haben das auch Autor-schaft genannt), und zwar von unterschiedlichen Akteuren wie den Schüler:innen und dem Lehrpersonal.

Daraus haben wir elf Empfehlungen abgeleitet, auf die ich aufgrund der Kürze der Zeit nicht näher eingehen werde, aber auf die wir vielleicht im Laufe der Diskussion zu sprechen kommen.

(Folie 8)

Worauf ich aber kurz zu sprechen kommen möchte, das sind die Querschnittsthemen, die wir bei allen Sektoren identifiziert haben, da diese auch für den Bereich der schulischen Bildung relevant sind.

Der Grund, warum das hier so bunt ist, ist, dass ich versucht habe, unterschiedliche Themen zu clustern. Bestimmte ethische Fragen stellen sich

insbesondere in Bezug auf den Umgang mit Daten und in Bezug auf das Verfahren von maschinellen Lernverfahren. Das sind alle Fragen rund um den Schutz der Privatsphäre, Fragen von Überwachung und Datensouveränität, aber auch Fragen rund um systematische Verzerrungen und Diskriminierungen sowie Transparenz und Nachvollziehbarkeit. Das sind die hier rot markierten Punkte. Das sind Aspekte, die in vielen ethischen Stellungnahmen zu KI immer wieder diskutiert werden.

Wir haben uns auch damit beschäftigt (in Blau), welche Auswirkungen es insgesamt auf die Handlungsmöglichkeiten von unterschiedlichen Betroffenen hat, wenn sie KI einsetzen und verwenden, welche Auswirkungen das auf menschliche Kompetenzen hat und wie es um die Autonomie im Umgang mit solchen Technologien bestellt ist.

Einen dritten Blick (hier in Gelb) haben wir geworfen auf die Rolle von KI und digitalen Technologien als Infrastrukturen. Wie können wir sicherstellen, dass wir Abhängigkeiten und Vulnerabilitäten vermeiden und uns bestimmter Pfadabhängigkeiten sicher werden?

Auf einen Punkt möchte ich noch eingehen, weil der auch für den Bereich der Bildung oder für alle Bereiche von KI relevant ist, nämlich die Frage: Was macht es prinzipiell, dass wir uns stärker auf statistische wahrscheinlichkeitstheoretische Vorhersagen verlassen und dementsprechend das Individuum in den Hintergrund tritt und wir es stärker als Teil einer Gruppe betrachten?

(Folie 9)

Kommen wir jetzt kurz zu Herausforderungen durch ChatGPT. Da ist natürlich die Frage: Was ist das Besondere an diesen großen Sprachmodellen oder anderen Formen von generativer KI, was wir vielleicht nicht in der Tiefe in der Stellungnahme betrachtet haben?

(Folie 10)

Um eine Brücke zum Bildungsbereich zu schlagen: Wir haben in der Stellungnahme geschrieben:

„Alle technologischen Möglichkeiten der Gestaltung der Bildungsprozesse sind [...] daraufhin zu überprüfen, ob sie dem [...] Verständnis des Menschen als einer zur Selbstbestimmung und Verantwortung fähigen Person entsprechen oder [...] diesem [...] entgegenstehen.“

(Folie 11)

Schon damals haben wir darauf verwiesen, dass ChatGPT hier durchaus neue Fragen stellt. Wir haben geschrieben:

„Die Notwendigkeit einer solch grundlegende Betrachtung wird auch in den aktuellen Debatten rund um ChatGPT sichtbar. [...] Ging es vordergründig zunächst um die Frage, wie Prüfungen fair gestaltet werden können, wenn man nicht weiß, ob Schülerinnen und Schüler ihre Hausarbeiten selbst geschrieben oder von ChatGPT haben produzieren lassen, so erfordert ChatGPT darüber hinaus eine erneute Auseinandersetzung damit, was Bildung ist und sein soll. Es geht um eine Vergewisserung und Neubestimmung dessen, was Ziel und Wert von Bildung ist, was relevantes Wissen ist und welche Fertigkeiten und Fähigkeiten Lernende weiterhin benötigen und welche vielleicht an Relevanz verloren haben.“

(Folie 12)

Was ist also das Besondere an ChatGPT? Meines Erachtens ist es bezeichnend, dass nicht nur bei uns, sondern auch im rechtlichen Bereich, im KI-Recht der Fokus nicht so stark auf diesen großen Sprachmodellen lag und dann eine große Überraschung letztes Jahr und vor allen Dingen im Laufe dieses Jahres zutage getreten ist. Diese disruptiven Effekte haben meines Erachtens zwei Ursachen:

Auf der einen Seite kann generative KI sehr einfach plausible Texte und Bilder von hoher Qualität, aber ohne Wahrheitsbezug herstellen. Es gibt eine sehr breite Anwendbarkeit, und weil Sprache eine so große Bedeutung für Information, Kommunikation und Emotion hat, ist das hochgradig relevant.

Der zweite Aspekt ist, dass wir es mit einem einfachen Interface zu tun haben und dass es einen freien Zugang über das Internet gab. Das heißt, wir haben es mit einer Technologie zu tun, die so einfach zu verwenden war wie eine Suchmaschine. Man brauchte wenig technische Vorbereitungen außer Zugang zum Internet und wenig Kompetenzen, um dann Texte und Bilder von sehr hoher Qualität produzieren zu können.

Diese beiden Sachen haben dazu geführt, dass ChatGPT schnell bekannt wurde und stark genutzt wurde.

(Folie 13)

Das hat dazu geführt, was ich andernorts schon als das Problem der dreifachen Täuschung benannt habe. Zunächst einmal: Weil ChatGPT ein Chatbot ist, geht es hier zunächst um die Frage: Können wir getäuscht werden darüber, dass wir mit KI interagieren? Diese Frage hat man sich schon mit ganz frühen Chatbots oder Vorläufern davon gestellt, wie dem Programm ELIZA von Josef Weizenbaum, das ihn zu einem sehr frühen Kritiker von digitalen Technologien gemacht hat. Selbst wenn man weiß, wie diese Technologien funktionieren, ist man stark verführt, ihnen Fähigkeiten und Fertigkeiten zuzuschreiben, die diese gar nicht haben.

(Folie 14)

Das bringt mich zum zweiten Aspekt der Täuschung: der Täuschung über die Fähigkeiten von KI. Wir alle werden uns vielleicht verwundert die Augen gerieben haben über diese Zuschreibungen, die an generative KI im Laufe des letzten Jahres gemacht wurden in Bezug auf Verständnis und Bewusstsein, was diese Systeme natürlich alles nicht haben. Aber die Interaktion scheint es nahezulegen, dass es mehr ist als Statistik, auch wenn es nur reine Statistik ist.

(Folie 15)

Die dritte Form der Täuschung ist die Täuschung durch KI-generierte Inhalte. Die kann in manchen Bereichen noch relativ amüsant sein (wie das Bild vom Papst im weißen Daunenmantel), aber in politischen Kontexten natürlich hochgradig problematisch werden.

(Folie 16)

Wichtig ist mir zu sagen: Wir haben dieses dreifache Problem der Täuschung, aber alle herkömmlichen Herausforderungen durch KI, die ich vorher mit Bezug auf die Stellungnahme genannt habe, sind weiterhin relevant, wie die Fragen rund um Bias und Diskriminierung, Mangel an Transparenz und Nachvollziehbarkeit, Schutz der Privatsphäre oder auch die prekären Arbeitsbedingungen in der Herstellung der Systeme.

(Folie 17)

Ich komme zum Schluss. Mit Bezug auf unsere Stellungnahme möchte ich noch mal betonen, dass das Ziel des Delegierens an KI-Systeme immer darin bestehen muss, menschliche Handlungsmöglichkeiten und Autorschaft zu erweitern und die Verminderung zu verhindern.

Die Auswirkungen sind sehr unterschiedlich für verschiedene Betroffene. Der besondere Fokus muss daher auf denen liegen, die bereits vulnerabel sind.

Der Teufel steckt hier oft im Detail. Das heißt, wir müssen einen genauen Blick auf die Technologien selbst verwenden, auf die Algorithmen, auf die Schnittstellen, aber auch auf die institutionellen und organisationalen Rahmenbedingungen.

Generative KI hat alle üblichen Probleme von KI, die hier auch relevant sind, und das Problem der dreifachen Täuschung kommt noch hinzu.

Damit bin ich am Ende meines kurzen Impulses. Vielen Dank.

Elisabeth Gräß-Schmidt (15:47)

Vielen Dank. Nun hören wir den Impuls von Doris Weßels. Sie ist Professorin für Wirtschaftsinformatik an der Fachhochschule Kiel und Mitbegründerin des Kompetenzzentrums „Schreiben lernen und lernen mit künstlicher Intelligenz – Tools und Techniken für Bildung und Wissenschaft“. Sie spricht zum Thema: Was können ChatGPT und Co. – und was nicht?

Was können ChatGPT und Co. – und was nicht?

Doris Weßels · Fachhochschule Kiel (16:18)

Vielen Dank. Der Titel meines Vortrags ist: Was können ChatGPT und Co. und was können sie nicht? Ich habe als Hintergrundbild zur Visualisierung DALL-E 3 bemüht, also generative KI, in dem Fall ein Bildgenerator, und habe mich inspirieren lassen durch die Fortschritte, die wir mit ChatGPT erlebt haben, und habe gedacht: Diese Treppe geht steil nach oben. Da sind viele Stufen erklommen worden, aber die Frage ist: Wie viele Stufen kann ein solches System oder die Gruppe von Systemen noch nehmen? Und wo sagen wir: „Bis dahin, aber nicht weiter“? Das ist täglich in der Diskussion: Was ist technisch machbar? Was wollen wir? Was ist ethisch, moralisch, gesetzlich vertretbar und was nicht?

(Folie 2)

ChatGPT hatte gerade, am 30.11., den ersten Geburtstag. Das erste Jahr mit ChatGPT war beeindruckend. Das Logo wurde am 30.11. veröffentlicht, und am Anfang des Jahres stand häufig die Frage im Vordergrund: Was kann ChatGPT überhaupt? Es wurden viele Kritikpunkte geäußert. Man hat sich mokiert, man hat sich lustig ge-

macht, und im Laufe des Jahres wurden die neueren Versionen veröffentlicht (das Sprachmodell in der Version 4 am 14.3., jetzt die Variante Turbo am 6.11.) und jetzt ist die Frage eher: Was kann ChatGPT eigentlich nicht? Und das nehme ich in meinem Umfeld ganz stark wahr: Der Fokus und die Blickrichtung, die Perspektive haben sich drastisch gewandelt.

(Folie 3)

Zur heutigen Agenda: Die Reise, auf die ich Sie jetzt mitnehmen möchte, ist, erst mal ein bisschen in die DNA von KI-Sprachmodellen einzutauchen, dann die Frage: Welche Potenziale sehen wir heute, aber auch welche Limitationen? Eenden möchte ich mit meinem Herzensthema, der Task-Force „KI-Bildung“. Denn ich glaube, dass wir hier eine große Chance nutzen müssen, Lehrende in ganz Deutschland schnell fit zu machen und ein Grundverständnis über diese Technologie in die Gesellschaft zu bringen. Nach dem Vortrag können wir dann gemeinsam eine Diskussion führen.

(Folie 4)

Ich nehme Sie jetzt mit in die KI-Sprachmodelle und wir tauchen ein bisschen tiefer ein, um die Charakteristika generativer KI zu verstehen.

(Folie 5)

Diese Darstellung soll das beleuchten. Judith Simon hat das schön beschrieben: Das ist pure Statistik. Es ist maschinelles Lernen. Es ist ein künstliches neuronales Netz, wo man versucht, das, was bei uns im Gehirn in der Verbindung von Neuronen und Synapsen passiert, in Software zu gießen und abzubilden, und das Ganze ist ein spezieller Algorithmus. Der ist nicht deterministisch (das heißt nicht Eingabe gleich Ausgabe), sondern probabilistisch. Hier regiert quasi der Zufall und hier regiert statistische Plausibilität.

Wenn man jetzt eingibt als Mensch (das ist ja ein natürliches Erkennen von Sprache): „Angela Merkel flog nach“ und wir diesen Algorithmus in Gang setzen, dann ist ein solches Modell darauf trainiert, Text fortzuführen. Das ist das Trainingsziel. Es geht nicht darum, Fakten zu suchen oder zu ermitteln. Es ist keine Suchmaschine, sondern es geht darum, Text zu produzieren, und zwar in einem kontinuierlichen Fluss, das heißt, hier oben sehen wir das Symbol des Würfels.

Die Textfortführung basiert auf dem, was in den Trainingsdaten steckt. Hier sind Hunderte von Millionen von Texten herangezogen worden. Auch das ist ein Diskussionsthema: Durfte man das überhaupt? Oder war das der größte Datendiebstahl der Menschheit? Hier wurden also Texte genommen. Diese Software hat Strukturen und Muster in Hunderten von Millionen von menschlichen Texten erkannt. Das bildet dieses Gesamtsystem, und aufgrund dieser Strukturen und Muster ist das Modell in der Lage, Text fortzuführen. Das heißt, Stück für Stück, Wortsilbe für Wortsilbe, hier Token genannt, geschieht dieser Algorithmus, den ich jetzt im Detail noch einmal zeigen werde.

(neue Folie, nicht im PDF)

Ich nehme Sie dazu mit auf den sogenannten Playground von OpenAI. Wenn man hier als Mensch eingibt: „Angela Merkel flog nach“ und dann die Taste „Submit“ betätigt, dann ist das Modell aufgefordert, meinen Text fortzuführen. Wir bewegen uns hier in dem Modell 3.5 und lassen uns in Farben die Wahrscheinlichkeiten anzeigen, wie diese Wortsilben gewählt wurden. Je grüner die Farbe, desto wahrscheinlicher.

Das Spiel beginnt, erste Runde. Ich betätige die Submit-Taste und wir sehen hier bunte Wortsilben, die reingeflogen sind. Wenn ich jetzt mit der

Maustaste darübergehe, sehe ich Wahrscheinlichkeiten. „China“ hatte eine eher rötliche Farbe, das heißt, in diesem Spektrum der potenziellen Kandidaten ist das ein nicht ganz wahrscheinlicher Kandidat. Wenn ich hier darübergehe, sieht man: Diese Wortsilben sind Stück für Stück zusammengesetzt auf Basis der Trainingsdaten, die hier drinstecken.

Wenn ich das Ganze noch mal mache (ich kann das alles löschen; neues Spiel, neues Glück), werden wir einen neuen Text erleben. Hier flog sie nach Jerusalem. Ich kann das beliebig wiederholen. Jerusalem ist rot, nicht so wahrscheinlich. Washington wäre eigentlich viel wahrscheinlicher gewesen aufgrund der Trainingsdaten.

(Folie 6)

Das müssen wir uns immer bewusst machen: Bei diesem Mechanismus geht es nicht um Fakten. Das Trainingsziel dieser Modelle ist gewesen, Text zu produzieren. Solche Modelle erzeugen Unikate und auch keine Plagiate, das ist für uns im Bildungsbereich eine ganz wichtige Unterscheidung. Ich könnte das Ergebnis nicht reproduzieren, sondern es ist wie ein munteres Wortsilbenwürfeln, aber geprägt durch das, was in den Trainingsdaten drinsteckt. Es spiegelt uns das auch wieder und ist getrieben durch statistische Plausibilität. Deshalb ist es auch keine Suchmaschine, zumindest in dieser Form nicht. Es ist eine Inspirationsmaschine und es imitiert uns Menschen in einer beeindruckenden Form.

Ich habe extra in Klammern gesetzt: in der geschlossenen Variante keine Suchmaschine. Denn wir haben im Laufe des Jahres gesehen, dass diese Technologie ganz häufig mit anderen Technologien, in dem Fall mit Suchmaschinen verbunden wird. Das heißt, wir haben jetzt das Beste aus zwei Welten, wenn man so will. Wir haben Modelle und Systeme, die uns in natürlicher Sprache

verstehen können, aber sie benutzen zusätzlich eine Suchmaschine und liefern dann etwas in natürlicher Sprache in einer sehr ansprechenden textuellen Darstellung zurück. Das ist zum Beispiel ein Perplexity AI. Das Zusammenspiel unterschiedlicher Technologien erleben wir auch bei der Suchmaschine Bing von Microsoft.

(Folie 7)

Ganz wichtig sind qualitätssteigernde Maßnahmen, wie wir in diesem Jahr erlebt haben. Unter ethischen Aspekten sind die unzulässigen, unerwünschten Inhalte ganz wichtig, die auch in den Trainingsdaten drinstecken. Die sollen natürlich nicht ungefiltert herausgegeben werden. Ich möchte nicht jedem den Bombenbau oder den perfekten Mord erklären. Hier hat man von der Anbieterseite stark nachgearbeitet, um diese Gefahr zu reduzieren. Das ist auch gelungen.

Wenn das Modell konfabuliert und wir als Menschen sagen, das ist Blödsinn, dann nennt man das Halluzination. Diese Gefahr ist durch entsprechende Mechanismen kontinuierlich reduziert worden. Auch dort haben wir deutliche Fortschritte erlebt.

(Folie 8)

Man kann sich so ein Sprachmodell vorstellen im Kern immer noch ganz pur, wie es am Anfang war, und auf den äußeren Schalen ist das ganze Jahr über von allen Anbietern nachgearbeitet worden. Es ist ein Feintuning, es sind Filtermechanismen, es sind Sekundärsysteme. Das heißt, anbieterseitig bemüht man sich darum, diese Modelle immer sicherer, besser und leistungsstärker zu machen. Das haben wir das ganze Jahr in beeindruckender Form erlebt.

(Folie 9)

Diese Sprachmodelle werden manchmal auch als General Purpose AI bezeichnet, weil sie wie eine

Allzwecklösung wirken und sehr vielseitig in der Einsetzbarkeit sind. Man kann das über den Playground machen; das haben wir vorhin gemacht, wo wir solche Modelle erleben. ChatGPT ist die niedrigschwellige, einfachste Variante, wo ich einfach nur etwas eingeben muss, eine Enter-taste drücke und heraus kommt ein Text.

Aber wir haben auch ChatGPT schon als einen kleinen Kosmos erleben dürfen, denn in der kostenpflichtigen Variante kann ich viele integrierte Software-Funktionen nutzen. Im März konnte man in ChatGPT ungefähr 70 Funktionen aufrufen: Man konnte Reisen buchen, man konnte sich im Restaurant einen Tisch reservieren, man konnte mathematische Aufgaben lösen. Jetzt, im Dezember, kann man ungefähr 1400 Software-Anwendungen benutzen, wenn man mit ChatGPT arbeitet. Da ist ein Riesenkosmos entstanden, der auch für den Bildungsbereich faszinierend ist.

Auch bei der Aktualität der Daten hat sich viel getan. Wir hatten noch vor Kurzem einen Stand der Daten aus dem Jahr 2021, und wir haben jetzt in der aktuellsten Modellversion von OpenAI den Stand von April 2023. Das heißt, die Aktualität der Datenbasis hat sich auch deutlich verbessert.

(Folie 10)

Jetzt kommen wir zu den Potenzialen und den Limitationen.

(Folie 11)

Ich habe als Beispiel den Cornelsen Bereichsindex gewählt. Dort gibt es eine sehr positive Sicht der befragten Experten auf solche Systeme. In dem Fall waren es 250 Personen, die im deutschen Bildungssystem in verschiedenen Rollen tätig sind. 90 Prozent betrachten Systeme wie ChatGPT als positiv, als Hilfsmittel und als Innovationsstreiber, 10 Prozent demzufolge nicht. Bei den

Lernenden sind es etwas mehr, die eine etwas kritischere Sicht haben: Das sind 13 Prozent. Das heißt, wir haben eine erstaunlich positive Sicht auf die Potenziale dieser Technologie gemäß dem Cornelsen Bildungsindex.

(Folie 12)

Ich habe selber mal eine solche Liste gemacht: Auf der linken Seite sind die positiven Aspekte und auf der rechten Seite die kritischen Aspekte, Limitationen und Herausforderungen. Auf der linken Seite [„Potenziale“]:

- Man ist sehr schnell und sehr produktiv (ich glaube, das kann jeder bestätigen, der damit gearbeitet hat).
- Man kann auch beim Automatisieren dadurch, dass man einen Schreib- oder Diskussionspartner hat (je nachdem, wie man die Rolle einer solchen generativen KI oder auch von ChatGPT wahrnimmt), die Qualität steigern, wenn man damit kompetent umgeht.
- Man hat eine Inspirationshilfe. Kreativität habe ich in Anführungsstriche gesetzt, denn es ist eine andere Form von Kreativität, aber sie ist da.
- Man kann individualisiert und selbstbestimmt lernen und trainieren (das zeige ich gleich an zwei Beispielen). Das ist aus meiner Sicht das größte Potenzial, was wir hier sehen.
- Wenn wir damit entsprechend umgehen, fördern wir die Digitalkompetenz und die Multilingualität. Das wird im Bildungskontext häufig nicht so genannt, erstaunlicherweise: Denn dass wir kaum noch eine Sprachbarriere haben, dass wir sehr einfach übersetzen können, immer einfacher und immer besser, ist ein großer Schatz.
- Die Multimodalität ist auch im Bildungsbereich faszinierend: dass ich von der textuellen Darstellung in eine grafische Darstellung, eine

Visualisierung gehen kann. Wir haben inzwischen so viele Formen, die wir generativ nutzen können: Wir können Texte generieren, Bilder, Videos, Musik, Softwarecode, 3D-Welten, es gibt immer mehr Möglichkeiten. Wir können außerdem von der einen Darstellungsform in die andere springen.

Auf der rechten Seite [„Limitationen und Herausforderungen“]:

- Hier stehen die Halluzinationen, also wenn einfach Blödsinn produziert wird und es keine Fakten sind. Das ist insbesondere dann eine Gefahr, wenn die noch über Social Media oder andere reichweitestarken Plattformen verbreitet werden.
- Die Bias-Problematik, die Voreingenommenheit, dass sich Stereotype dort zeigen, reproduzieren und verstärken, sehen wir ganz stark in der Bildgenerierung. Da gibt es sehr viele Beispiele. Das sieht man unmittelbar, wenn man es ausprobiert.
- Wir haben viele rechtliche Fragestellungen. Wir haben das Urheberrecht. Wir haben gerade die Diskussion, EU-AI Act [Artificial Intelligence Act der EU], also Fragen über Fragen, und je tiefer man eintaucht, desto mehr Fragen entstehen auch.
- Datenschutz-Problematik: Wir haben die Problematik der DSGVO (Datenschutz-Grundverordnung).
- Wir haben eine große Abhängigkeit von den Anbietern.
- Wir haben Kompetenzverschiebungen. Das sind an der einen Stelle Kompetenzverluste, aber an anderer Stelle entstehen auch neue Kompetenzen.
- Wir haben ökologisch gesehen den Ressourcenverbrauch, der auch thematisiert wird. Auch da gibt es glücklicherweise positive Entwicklungen und neue Versuche, das zu lindern.

- Es gibt (das beobachte ich und nicht nur ich) auch die Abhängigkeit im Online-Bereich, dass also die persönliche Interaktion, der persönliche Dialog und Austausch zurückgefahren wird und stattdessen der Austausch mit KI-Tools und Techniken zunimmt. Das ist sicherlich auch eine gesellschaftliche Gefahr.

In der Mitte ist der gelbe Bereich [„Abhängig von der Nutzungsmöglichkeit“]. Das ist ein starker Hebel, der positiv auf die linke Seite, auf die Potenzialseite, wie auf die rechte Seite wirken kann. Denn die Frage ist: Wer hat eigentlich Zugang zu diesen Werkzeugen?

Wir laufen Gefahr, dass wir eine Spaltung der Gesellschaft verstärken, auch eine Spaltung der Lernenden im Bildungskontext, denn zum Beispiel bei ChatGPT haben wir die kostenlose und die kostenpflichtige Variante. Die kostenpflichtige Variante basiert auf einer höheren Modellversion. Sie ist deutlich leistungsstärker als die kostenlose Variante, aber kostet halt. Wir benötigen hier für eine Bildungsgerechtigkeit dringend für alle Lernenden oder eigentlich für alle im Bildungsbereich den Zugang auf die gleiche Technologie, auf das gleiche Sprachmodell, natürlich eine hohe Version. Das haben wir bis heute nicht geschafft und das finde ich ganz wichtig. Denn das ist wirklich eine gesellschaftliche Gefahr.

(Folie 13)

Zu einem positiven Aspekt, was vielleicht nicht so bekannt ist, aber jeder, jede kann sich einfach einen kleinen Trainer bauen. Ich habe hier mal ein kleines Beispiel gebaut: „Du bist mein Vokaltrainer für Englisch. Erstelle mir ein Quiz mit 10 Fragen für englische Fachbegriffe. Warte immer meine Antwort ab, bevor du mir die nächste Frage stellst. Belohne mich mit einer kleinen Überraschung oder einem Smiley, gib mir am Ende eine

Tabelle und Tipps, wie ich meine Leistung verbessern kann.“ Das kann man in natürlicher Sprache für jedes Fach eingeben, und dann hat man seinen persönlichen Lerntutor.

Jetzt ist natürlich die Frage: Wann entstehen die Halluzinationen? Es sind ja Lernende, die das nicht unbedingt fachlich beurteilen können. Das heißt, wir müssen das didaktisch flankieren. Wir müssen auch Sorge dafür tragen, dass sich nicht falsches Wissen verstärkt. Das ist nicht ganz risikofrei, aber wir erleben kontinuierliche Leistungssteigerungen der Systeme, und dieses Potenzial zu nutzen ist, glaube ich, ein großer Schatz.

(Folie 14)

Wenn man das noch professioneller machen will, kann man Folgendes machen. Das kann man seit Anfang November, seit dem 6.11., dieser Developer-Konferenz. Kurz danach konnten wir kleine eigene Bots bauen, das sind diese GPTs. Ich habe hier verschiedene Lernbots gebaut, hier zum Beispiel einen Lernbot fürs Projektmanagement. Den kann man sehr einfach bauen. Man wird ein bisschen an die Hand genommen und dann hat man sehr schnell – man kann auch eigene Dokumente, eigene Skripte hochladen, und dann kann man mit diesen Dokumenten lernen und das sehr schön gestalten.

Ich habe das bewusst so gestaltet, dass es auch Spaß machen muss, also dass es viele Anreize gibt für Lernende, sich zu beschäftigen. Das wird aus meiner Sicht das beherrschende Thema im nächsten Jahr sein: diese Form von Lernbots, die man in unterschiedlichen Kontexten einsetzen kann.

(Folie 15)

Jetzt kommt eine schwierige Geschichte, denn die Lernenden können wie gesagt per Knopfdruck Hausarbeiten, schriftliche Ausarbeitungen schreiben. Das war das große Thema schon zu Beginn

des Jahres: Bedeutet es das Ende der Hausarbeit? Aus meiner Sicht nicht. Es hängt immer maßgeblich an der Aufgabenstellung. Wir als Lehrende können diese Technologie auch benutzen und dann das, was die Studierenden oder die Lernenden geschrieben haben, KI-gestützt bewerten. Wenn wir das machen und über die gesamte Prozesskette nachdenken (also die Seite der Lernenden generiert Texte, wir bekommen diese Texte und generieren dann das Gutachten oder die Bewertung, auch wieder KI-gestützt), dann haben wir eine Prozesskette, die man in ihrer Gesamtheit natürlich kritisch hinterfragen muss. Ich glaube, wir schlittern in so eine Entwicklung hinein, und das müssen wir uns bewusst machen. Das heißt, wir müssen hier ganzheitlicher denken, wir müssen in Prozessketten denken, wir müssen auch das menschliche Verhalten berücksichtigen, und die Entwicklung zeigt – und das kann ich auch bestätigen, ich habe auch so einen Prüf-Bot gebaut - das ist ethisch und rechtlich schwierig, grenzwertig, aber technisch geht das schon sehr gut. Und wir müssen uns nichts vormachen: Das wird auch benutzt werden, und darüber muss man nachdenken, das muss man diskutieren.

(Folie 16)

Abschließend meine Empfehlungen zu diesem Thema.

(Folie 17)

Das habe ich schon im April beim Bundestagsausschuss dargelegt. Wir haben schon im April geglaubt, dass wir dringend eine Task-Force benötigen und hier schnell eine Bildungsstrategie fahren müssen, um Lehrende an Schulen wie an Hochschulen schnell fit zu machen und ihnen ein Grundlagenverständnis zu geben, damit sie es an die Lernenden weitergeben können und wir dann große Teile der Gesellschaft erreichen, um auch als Gesellschaft diskursfähig zu sein über die Art

und Weise, wie wir mit dieser Technologie umgehen wollen. Das habe ich hier formuliert und kann man auch nachlesen. Es ist entsprechend veröffentlicht worden.

Damit beende ich meinen Vortrag und freue mich auf die weitere Diskussion.

Elisabeth Gräß-Schmidt (34:57)

Vielen Dank. Ich gebe jetzt das Wort an Frau Felicitas Macgilchrist. Sie ist Professorin für Digitale Bildung in der Schule an der Universität Oldenburg, Mitbegründerin von The Basement, dem Digital Lab des Leibniz-Instituts für Bildungsmedien und des Leibniz Wissenschafts-Campus – Postdigitale Partizipation in Braunschweig. Sie spricht zum Thema: Wie wird in Schulen auf die neuen Angebote reagiert und wie werden sie genutzt?

Wie wird in Schulen auf die neuen Angebote reagiert und wie werden sie genutzt?

Felicitas Macgilchrist · Carl von Ossietzky Universität Oldenburg (35:37)

(Folie 1)

Vielen Dank für die Einladung. Schön, dass ich hier sprechen und Impulse aus der Kultur- und Erziehungswissenschaft geben darf.

Ich werde auch über Schule, Lehrkräfte und Schüler:innen sprechen, aber ich möchte die Aufgabe der Kulturwissenschaften auch benutzen, um grundsätzlich zu irritieren, was sie ja versuchen. Den Begriff „KI“ habe ich in Anführungsstriche gesetzt, denn durch meinen Input zieht sich die Idee, wie der Begriff „KI“ überhaupt zu benutzen ist. Dem steht auch ein ethisches Handeln entgegen, denn wenn wir über KI sprechen (das ist ja

der Begriff, der uns heute zusammengebracht hat), machen wir KI zu einem Ding und machen das Narrativ, das gerade in der Gesellschaft so prägend ist, noch stärker.

(Folie 2)

Auf dem Weg dahin möchte ich drei Fragen beantworten:

[1] Wie werden ChatGPT und Co. bereits im Bildungskontext eingesetzt? Da sage ich etwas über Schule, Schüler:innen und Lehrkräfte.

[2] Welche gesellschaftlich problematischen Auswirkungen sind beobachtet worden, die sich anhand dieser Nutzungsweisen in den Schulen entfalten?

[3] Was ist jetzt zu tun? Wie kann generative KI in der Schule nicht nur anders genutzt werden (ohne nur die Nutzungsweisen und die Verantwortung bei den Nutzer:innen zu sehen), sondern auch anders entwickelt, anders reguliert, anders finanziert usw. werden?

(Folie 3)

Bevor ich zu diesen Fragen komme, habe ich ein paar Vorbemerkungen. Hier beziehe ich mich auf Lucy Suchman und andere Arbeiten in den Science and Technology Studies, die sagen: KI ist kein unkontroverses Ding an sich, sondern KI ist ein Narrativ. Wenn wir über KI sprechen, bezeichnen wir viele verschiedene Techniken, Methoden und Technologien.

Wenn wir auf KI schauen und das Wort „KI“ benutzen, dann stabilisieren wir das, was wir gerade versuchen zu dekonstruieren und zu destabilisieren. Es ist ein sehr verführerisches Narrativ, gerade in der Gesellschaft, wie es besprochen wird, und wir brauchen Arten und Weisen, wie wir dieses Narrativ dekonstruieren und anpassen können. Deswegen fand ich es sehr schön, dass die Veranstaltung heute auch über ChatGPT und Co.

spricht, weil ich heute vor allem über diese Teile von KI sprechen werde, die die großen Sprachmodelle genannt werden.

(Folie 4)

Hier beziehe ich mich auf eine Definition von Emily Bender, Timnit Gebru und anderen. Sie verstehen unter dem Begriff Sprachmodell Systeme, die für die Vorhersage von Zeichenketten trainiert werden, also die Vorhersage der Wahrscheinlichkeit eines Tokens (Zeichens, Wortes oder Zeichenkette), in Abhängigkeit von seinem vorherigen oder umgebenden Kontext.

Es geht also um Statistik. Das wollte ich noch mal betonen, auch wenn es schon gesagt worden ist. Es sind statistische Maschinen. In diesem Text wird das stochastische Papageien [„Stochastic Parrots“] genannt, und es wird explizit darauf eingegangen: Sie verstehen Sprache nicht, sie verstehen Menschen nicht. Darauf komme ich noch zurück.

(Folie 5)

Zur ersten Frage: Wie werden diese Systeme im Bildungskontext eingesetzt?

(Folie 6)

Ich habe hier ein paar Schlaglichter mitgebracht. Oben rechts: Der Bayerische Rundfunk hat ChatGPT das bayerische Abitur machen lassen. Da hat ChatGPT 3.5 nicht so gut abgeschnitten (in manchen Fächern ziemlich gut, in anderen Fächern nicht so gut). Aber ChatGPT 4 hat das bayerische Abitur laut Bayerischem Rundfunk mit einer Zwei abgeschlossen.

Zweites Beispiel: Als ChatGPT im November letzten Jahres rauskam, hat New York erst mal die Nutzung verboten. Viele haben kritisch darauf reagiert und gesagt: Ein Verbot funktioniert nicht. New York hat jetzt auch dieses Verbot zurückge-

nommen. Aber die Idee dahinter war, das zu verlangsamen, nicht gleich im Bildungssystem auf alles aufzuspringen, was rauskommt, dass das Bildungssystem immer auf die Technik reagiert, sondern erst mal zu verlangsamen und zu überlegen: Was heißt das jetzt und wie wollen wir langsam und mit Bedacht entscheiden, wie wir damit umgehen?

Als drittes Beispiel: Sachsen-Anhalt hat für alle Lehrkräfte in Sachsen-Anhalt zwei Systeme zur Verfügung gestellt: Das eine heißt EmuGPT und das andere heißt EmuDALL-E. Wenn die Lehrkräfte sich an dem System, was sie in Sachsen-Anhalt haben, anmelden (und das können alle Lehrkräfte, wenn sie ein Konto haben), können sie diese spezifische Version von ChatGPT und von DALL-E nutzen, um Bilder zu generieren, um Texte zu generieren. Darin ist auch die kostenpflichtige Version enthalten, sodass alle Lehrkräfte in diesem Bundesland das haben.

Das sind nur einige Schlaglichter. Auch dieses Bild ist in einem Bericht über EmuDALL-E und EmuGPT produziert worden von DALL-E. Man sieht hier bunte Farben, Instagram-Körperbilder, die Sonne scheint und man hat hinten eine Idee von KI, die nur in blauen Farben und ganz unrealistisch ist. Auf dieses Bild komme ich später noch zurück.

(Folie 7)

Die Regierung in Großbritannien hat im Sommer 2023 eine Umfrage gestartet, um zu eruieren, wofür Lehrkräfte ChatGPT und Co. eigentlich benutzen. Es kamen ca. 600 Antworten. Die meisten Antworten kamen von Early-Adopter-Lehrkräften, also denjenigen, die schon mit den Sprachmodellen gearbeitet haben, herumexperimentiert haben, also grundsätzlich relativ positiv eingestellt waren. Sie haben auch Sorgen geäußert, aber in der Liste, die Sie hier sehen, stehen die Punkte,

die sie gesagt haben, was sie vor allem benutzt haben, in der Reihenfolge der Häufigkeit. Diese generative KI wird also verwendet, um Lernmaterialien zu erstellen inklusive Ressourcen speziell für einzelne Schüler:innen mit einem besonderen Förderbedarf.

Es ging um die Gestaltung von Handouts, von Arbeitsblättern und Präsentationen, Bildern, Musterantworten und Quizfragen. In Grundschulen berichteten die Lehrkräfte, dass sie generative KI zur Erstellung von Lesematerialien oder für Verständnisfragen für die Schüler:innen erstellt haben.

Wichtig bei diesem ersten Punkt und eigentlich bei allen Punkten ist, dass die Lehrkräfte, die geantwortet haben, betont haben, wie wichtig es ist, die Lernmaterialien auf Richtigkeit und auf Übereinstimmung mit dem Lehrplan zu überprüfen, also eine manuelle Nutzung von automatisierten Systemen.

Andere häufig genannte Anwendungen waren die Unterrichtsplanung, der Live-Einsatz im Unterricht, wo sie mit Schüler:innen zusammen etwas ausprobiert haben, etwas gesucht haben. Sie haben sie für administrative Aufgaben verwendet, um Briefe an Eltern zu schreiben und andere standardisierte Sachen. Einige haben auch mit Sprachmodellen experimentiert für die Bewertung, die Benotung und um Feedback auf Arbeiten von Schüler:innen zu geben (da betonen sie aber auch, dass es bedingt hilfreich sein kann, das zu benutzen). Sie haben es benutzt für die Schulung der jungen Menschen, der Schüler:innen. Sie haben es als Recherche, als Wissensquelle benutzt und für das Korrekturlesen und Übersetzen sowie die Entwicklung und Überprüfung von Code.

Am problematischsten scheint der Punkt der Recherche zu sein, weil dann ChatGPT als Wissensquelle und als Suchmaschine benutzt wird. Aber

die Idee dahinter ist nicht, es als Wissensquelle zu benutzen. Und damit kommen viele Probleme.

Noch mal zu Bewertung, Benotung und Feedback: Wenn die Schüler:innen die Texte schreiben lassen und die Lehrkräfte die Benotung mit dem System machen, dann liest niemand mehr etwas und schreibt niemand mehr etwas, und dann fragt man sich, wofür diese lernenden, bildenden Praktiken da sind.

(Folie 8)

Ich habe hier zwei leere Stellen eingefügt. Sie können kurz überlegen, was Sie denken, was für eine Prozentzahl da erscheinen wird.

Es gab eine Umfrage im März 2023 in den USA, wie viele Schüler:innen Chatbots für schulische Aktivitäten genutzt haben, und es gab eine Umfrage in Deutschland mit einer anderen Peergruppe: Schüler:innen, Azubis und Studierende, wie viele ChatGPT und Co. nutzen, vom November 2023.

Sie haben Folgendes gesagt, was sie damit machen: Sie nutzen sie als Hilfsmittel bei Aufgaben und Arbeiten (zum Beispiel Erstellung einer Aufsatzstruktur, Gedankenimpulse, Umformulierungen, Korrekturlesen usw.), um Inhalte zu erstellen (Präsentationen, Bilder und Kunstwerke); als Suchmaschine (auch hier wird es als Suchmaschine benutzt, um Themen und Konzepte zu erklären, also als Wissensquelle statt als Sprachquelle), und nicht so viele haben angegeben, dass sie Texte verfasst haben, die sie dann als eigene Leistung eingereicht haben. Aber das wird besprochen.

(Folie 9)

So, jetzt haben Sie Zeit zum Überlegen gehabt: 47 Prozent der Schüler:innen haben im März gesagt, dass sie es für schulische Aktivitäten nutzen, und

68 Prozent der Schüler:innen, Azubis und Studierenden in Deutschland nutzen ChatGPT und Co., so eine aktuellere Umfrage.

Hier ist wieder die Suchmaschine eine vielleicht besonders problematische Art und Weise der Nutzung, wenn wir das so sehen wollen.

(Folie 10)

Zu Frage 2: Was sind die problematischen gesellschaftlichen Aspekte?

(Folie 11)

Zwei Aspekte von den vier Aspekten hier werden viel diskutiert: die Ausbeutung global betrachtet und die Desinformation, wenn es als Wissensquelle benutzt wird, und welche Auswirkungen das haben kann.

Ich möchte auf zwei andere Punkte eingehen: die generische Sprache und die Anthropomorphisierung.

Jürgen Geuter hat das beschrieben als Homogenisierung. Die Homogenisierung der Sprache und auch der Bildsprache (so wie wir das vorhin in dem einen Bild gesehen haben) stärkt die Dominanzgesellschaft. Das ist eine nicht so offensichtliche Version von Bias, aber wenn die Sprache in der Schule immer nur die Dominanzgesellschaft und Mehrheitspositionen zurückgibt, dann beobachten wir, dass die Gesellschaft einen Rechtsruck hat und das mehr in die Schule kommt. Diskriminierungen zirkulieren mehr. Minorisierte Positionen verschwinden von der Oberfläche, wenn diese Texte in die Schule kommen, und das kann als eine Entdemokratisierung oder Entpluralisierung von Wissenspositionen in der Gesellschaft betrachtet werden.

Zur Anthropomorphisierung: Ich habe hier unten ein Beispielbild. Es gibt ein Tool, einen Chatbot, der von Khan Academy explizit für den Bildungsbereich entwickelt wurde. Sarah Khan berichtet

hier von einer Schülerin, die einen Bericht über The Great Gatsby schreibt und sich fragt, warum er immer in das grüne Licht schaut. Sie unterhält sich mit diesem Chatbot, den Sie auf dem Foto sehen, mit Gatsby selbst, so: „Ich möchte jetzt mit Gatsby sprechen“, und dann hat sie eine Kommunikation mit ihm. Sie fragt ihn, was das grüne Licht ist, und er antwortet in dem Stil, den man von Gatsby kennt, und geht darauf ein, warum das grüne Licht so wichtig ist. Khan berichtet am Ende, dass die Schülerin, nachdem sie ein relativ langes Gespräch mit Gatsby hatte, sich entschuldigt, dass sie so viel von seiner Zeit in Anspruch genommen hat.

Khan erzählt das als eine witzige Geschichte, und so wird es auch vom Publikum aufgenommen. Aber wenn wir darüber nachdenken, welches Vertrauen sie zu diesem System aufgebaut hat und wie das als Intentional Agent, also als intentionaler Akteur in Interaktion mit ihr auftaucht, dann können wir uns fragen, was das ausmacht, wie sie später mit anderen automatisierten Systemen umgehen wird und ob sie dann ein übersteigertes Vertrauen in dieses System haben wird, das dazu führt, dass sie auch auf fehlerhafte Empfehlungen reagieren wird (wir wissen, dass viele Empfehlungen fehlerhaft sind).

Im Bildungsbereich grundsätzlich: Wenn eine Entscheidung vom System gemacht wird und die Verantwortung trotzdem noch bei der Person liegt, ist das manchmal in High-Stakes-Positionen, High-Stakes-Entscheidungen sehr schwierig für eine Person, gegen ein automatisiertes System zu entscheiden. Denn wenn man als Person falsch entscheidet, kann man dafür kritisiert werden, dass man etwas anderes empfohlen bekommen hat. Das sind große Fragen darüber, welche Entscheidungen getroffen werden und auf welcher Basis Entscheidungen getroffen werden.

(Folie 12)

Die dritte Frage: Was ist jetzt zu tun?

(Folie 13)

Es gibt viele Arten und Weisen, wie wir ChatGPT und Co. anders nutzen können, jenseits von Verboten oder Prompt-Kompetenz, was gerade viel besprochen wird. Was wir beobachten, ist eine manuelle Nutzungsweise zwischen Verzicht (wo will ich das nicht benutzen?) und Imagination, es auf kreative Art und Weise zu nutzen. Wir beobachten bei Nutzer:innen eine kritische, kreative Auseinandersetzung mit ChatGPT & Co., dass sie es manuell bearbeiten, also nicht automatisiert. Viele berichten, dass sie noch mal reingehen, noch mal korrigieren und noch mal überprüfen. Das ist ein Aspekt, der gestärkt werden müsste.

Wenn ich das mit Schüler:innen oder Studierenden durchgehe, war immer auch ein Thema, wo sie selbst die Expertise schon haben. Ich habe neulich viel über Harry Potter gelernt. Denn wenn sie dort Fragen stellen, wo sie sich auskennen, dann können sie viel schneller erkennen, wo Lücken sind, wo Fehler sind und warum das dann statistisch plausibel erscheint.

Wir müssen aber weiter gucken als nur auf die Nutzungsweise und darauf, die Verantwortung den Nutzer:innen zu geben. Es gibt Ideen, wie man es anders entwickelt (ich habe hier eine Quelle eingefügt): nicht die großen Sprachmodelle, sondern kleinere, kuratierte Sprachmodelle. Anders regulieren: nicht nur auf individuelle Menschenrechte und individuelle Rechte eingehen, sondern stärker auf gesellschaftliche Ungerechtigkeiten eingehen; und anders finanzieren. Es gibt Vorschläge, dass, wenn diese Sprachmodelle zur notwendigen öffentlichen Infrastruktur für das Gemeinwohl werden, sie dann nicht mehr in privater Hand liegen dürfen.

(Folie 14)

Ich komme zu meiner ersten Folie zurück: „KI“ ist kein unkontroverses Ding, sondern „KI“ ist ein Narrativ.

(Folie 15)

Ich habe einige Lektürehinweise. Vielen Dank für die Aufmerksamkeit.

Elisabeth Gräß-Schmidt

Vielen Dank für Ihren Impuls.

Diskussion

Moderation: Elisabeth Gräß-Schmidt · Mitglied des Deutschen Ethikrates (51:14)

Nun möchte ich für die Diskussion auch Frau Weßels und Judith [Simon] in die Runde bitten.

Frau Weßels, Sie haben gezeigt, wie die Verbesserungen dieser generativen künstlichen Intelligenz exponentiell voranschreiten. Da stellt sich die Frage: Inwiefern ist dies auch eine exponentielle Steigerung der möglichen Falschnachrichten, die dann im Internet kursieren? Oder verbessert sich dann letztlich der Ertrag der Ergebnisse in Richtung Faktentreue, also dass man es tatsächlich, wie Frau Macgilchrist gesagt hat, als Suchmaschine benutzt, also Schüler es als Suchmaschine benutzen? Da ist aber die Gefahr, dass es keine Fakten sind, sondern, weil es nur Statistik ist, im Grunde ein geratenes Ergebnis.

Wie ist diese Verbesserung hin auf Faktenkonformität zu sehen und damit auch als qualitativer Wissenszugang zu bewerten?

Doris Weßels (52:41)

Bei Ihrer Frage möchte ich bei drei Punkten ansetzen. Der erste Punkt ist: Wie entsteht das überhaupt, dass Schüler und Schülerinnen das als

Suchmaschine fehlinterpretieren? Und wenn ich mich mit Studierenden und manchmal auch mit Lehrern und Lehrerinnen unterhalte, sagen die: „Das wusste ich gar nicht.“ Es scheint also (das war mir nicht bewusst): Wenn der normale User ohne Verständnis für die Hintergründe solch eine Benutzeroberfläche mit so einem Eingabefeld sieht, dann interpretiert und nutzt er das selbstverständlich wie eine Suchmaschine. Das heißt, wenn man Studierende darauf aufmerksam macht, dass sie dort bitte nicht Literaturquellen suchen sollen, sagen sie: „Das hab ich nicht gewusst. Uns sagt das ja keiner.“ Das Versagen der Lehrenden wird häufig von den Lernenden zurückgespielt, die sagen: „Wieso sind wir nicht aufgeklärt worden?“

Oder noch schlimmer: Wir haben unsere Lehrenden gefragt. Die haben es genauso falsch verstanden. Deshalb ist es so wichtig, dass wir ein Grundverständnis über die Arbeitsweise, die Struktur, die Mechanismen einer solchen Software in die Breite der Gesellschaft, insbesondere in die Gruppe der Lehrenden bringen. Denn wenn die es gewusst hätten, hätten sie es auch rechtzeitig an die Lernenden weitergegeben. Aber das ist passiert, weil wir es nicht geschafft haben, die Lehrenden rechtzeitig upzudaten oder ins Boot zu holen.

Das Zweite, konkret aus technischer Sicht: Die weiteren Entwicklungen der Systeme – es gibt aktuelle Studien, wonach wir im März bei der Version 4 eine Faktentreue von ungefähr 80 Prozent hatten. Das bedeutet im Umkehrschluss eine Halluzinationsrate von ca. 20 Prozent. Das scheint sich aktuell auf eine Halluzinationsrate von ungefähr 4 Prozent reduziert zu haben. Es ist immer die Frage, nach welchen Testszenarien das ermittelt wurde, aber wir können davon ausgehen, dass wir bei der Halluzinationsrate bei den hochentwi-

ckelten Sprachmodellen heute im einstelligen Bereich sind. Das glaube ich schon. Das heißt, die Gefahr der Halluzinationen ist deutlich reduziert worden.

Der zweite Mechanismus ist (den habe ich in meinem Impuls schon angedeutet), dass man auf die Kombination unterschiedlicher Techniken setzt. Das heißt, man benutzt eine Suchmaschine in Kombination mit generativer KI. Dieses Zusammenspiel ist sehr schön. Wenn man zum Beispiel bei Bing, bei der Suchmaschine von Microsoft, eine Frage eingibt und im Hintergrund GPT 4 benutzt, oder auch bei Perplexity AI, dann bekommt man in der Antwort kleine Hyperlinks, kleine Fußnoten. Da geht man dann drauf und kann sich dann vergewissern, ob dieser Hyperlink, diese Quelle wirklich zu der Aussage des Textes passt. Das heißt, ich als Mensch kann das relativ einfach prüfen: Stimmt das oder stimmt das nicht? Ist die Quelle eine valide Quelle? Passt sie auch zu dem Text, der mir dort angegeben wird?

Das heißt, da haben wir schon gute Wege gefunden, um die Gefahr der Falschinformation deutlich zu reduzieren.

Elisabeth Gräß-Schmidt (55:50)

Vielen Dank. Also die Task-Force KI-Bildung, die Sie einfordern, wäre hier wichtig.

Frau Macgilchrist, Sie haben in Richtung der Notwendigkeit, dass Schüler und Schülerinnen aufgeklärt werden müssen, auf das Konzept einer Narrativität von KI hingewiesen. Das Narrativ des KI deutet darauf hin, dass man sehen muss: Was steckt dahinter? Das betrifft auch die Frage des Verstehens. Müssen nicht in der Schule vor allem jüngere Schülerinnen und Schüler darauf hingewiesen werden, dass KI eben nicht verstehen kann? Das wird ja gerade bei ChatGPT im Grunde verdunkelt, wenn man den Eindruck hat, hier wird

tatsächlich verstanden, weil auf meine Frage geantwortet wird, und insofern die Gefahr des Anthropomorphismus gegeben ist und dass gerade für jüngere Schülerinnen und Schüler diese Gefahr hoch ist. Daher die Frage: Ab welchem Alter ist in der Schule die Anwendung von Tablets oder überhaupt von digitalen Tools eigentlich sinnvoll? Ist es nicht für jüngere Schülerinnen und Schüler kontraproduktiv?

Felicitas Macgilchrist (57:28)

Ja, diese Frage wird heiß diskutiert. Es gibt die Position: lieber nicht, lieber später. Ich vertrete eher die Position: Die jungen Menschen spielen sowieso damit. Und wenn die Schule sie da nicht begleitet, dann haben einige eine Begleitung, die ihnen hilft, das kritisch einzusetzen, und andere nicht. Die Schule ist ja dafür da, Gerechtigkeit zu ermöglichen, damit alle die gleiche Möglichkeit haben, sich kritisch damit auseinanderzusetzen.

Dabei gibt es viele Probleme. Eine Sache bei OpenAI und diesen verschiedenen Produkten ist, dass man eigentlich, wenn man die Nutzungsbedingungen, die Terms and Conditions liest, unter 18 Jahren gar keine Lizenz haben darf. Es wird viel diskutiert, und das finde ich auch eine gute Idee, wenn die Lehrkraft das an der Tafel macht und man gemeinsam diskutiert. Wunderbar sind hier Fragen zu Harry Potter, zu Fußball oder was auch immer die Kids fragen wollen. Wenn die jungen Menschen dann eine Frage dazu stellen, was sie gerade interessiert, wo *sie* die Expertise haben und auch geweckt werden (auch Zehnjährige haben Expertise), dann sehen sie gleich: „Okay. Aber Moment, da ist ein Fehler drin, wieso ...“ Oder: „Da fehlt ja meine Lieblingsposition“ oder was auch immer. Dann kann man darüber sprechen, nicht nur über die Fehlinformationen, die wichtig sind (Halluzination ist wieder so ein anthropomorphisierender Begriff), sondern dann kann auch die Mehrheitsposition gezeigt

werden, und die jungen Menschen wissen sehr wohl, wie sie das dann dekonstruieren können, wenn sie die Expertise haben.

Ein Beispiel, das ich letztens hatte (aber nicht mit jungen Menschen): Ein Student hatte nach der Geschichte von Geometrie gefragt. Die Antwort war sehr ausführlich. Alles war da, aber wir haben noch mal gefragt: Was fehlt hier? Alles war korrekt, es waren keine Fehler drin. Was fehlt denn hier? Es waren nur männliche Mathematiker, die da genannt wurden. Also kann man dann noch mal nachhaken: Gab es auch Frauen, die in der Vergangenheit von Geometrie wichtig waren? Dann kam die nächste Antwort, dass die Frauen da waren. Diese Informationen sind da, aber sie sind nicht in der ersten Antwort, und zwar wegen der Vergangenheit von Texten, die in den Datensätzen sind.

Das kann alles ganz einfach und schlicht gemacht werden mit jungen Menschen. Es ist nicht gut, das verpflichtend zu machen. Also es gibt einige Lehrkräfte, die geben verpflichtende Aufgaben: „Guck das mal nach“ oder ein Thema zu recherchieren und dann eine kritische Auseinandersetzung mit dem Output von ChatGPT. Das ist eine sehr produktive Aufgabe, aber ein Problem, wenn sich Schüler:innen für die Lizenz per E-Mail anmelden müssen, um das zu tun. Das kann aber mit einem zentralen Konto gemacht werden, sodass die jungen Menschen das ausprobieren können und gerade die kritische Auseinandersetzung mit dem Output von ChatGPT einüben.

Elisabeth Gräß-Schmidt (01:00:16)

Vielen Dank.

Judith, dein wichtiger Impuls war das Aufmerksam machen auf die dreifache Täuschung. Wie wird in der Schule darauf vorbereitet, dass das hier auch erkannt wird? Wichtig ist vor allem die mittlere Täuschung der Halluzinationen und die

letzte Täuschung der Fake News, also der Manipulation, dass das erkannt wird.

Meine Frage: Wie ist das zu verstehen, dass den Schülern das Problem der unterschiedlichen Einschätzungen, also der Umgangsmöglichkeiten mit diesen Ergebnissen vermittelt wird? Ist nicht Voraussetzung, um das beurteilen zu können, dass schon ein bestimmtes Wissen vorhanden ist? Meine Frage zielt auf das Verhältnis von Wissen, von Wissensreservoir der Schüler und die Notwendigkeit des Umgangs. Oft wird gesagt: Wir müssen nur noch den Umgang mit Wissen erlernen, das Wissen selbst kommt von der KI. Es muss also nicht mehr gelernt werden, nichts mehr gepaukt werden, sondern nur noch der Umgang mit Wissen gepflegt werden. Meine Frage: Funktioniert das? Oder müssen wir uns nicht doch auch die herkömmlichen Wissensressourcen aneignen?

Judith Simon (01:02:03)

Die kurze Antwort wäre: Ja, das müssen wir weiterhin. Die längere würde in die Richtung der unterschiedlichen Aspekte der Täuschung gehen. Der Fokus in den Medien und in der Politik lag vor allem auf dem dritten Aspekt. Die Frage von Fake News und falschen und generierten Inhalten war das, was als Erstes diskutiert wurde. Denn das war ein Thema, das für die Medien und für die Politik relevant war.

Dieser zweite Aspekt, die Anthropomorphisierung, die dahintersteckt, diese Täuschung darüber, was die Systeme können – man darf sich da nichts vormachen, das ist auch ein Stück weit ein gepushtes Narrativ gewesen. Die ganzen Annahmen darüber kommen ja nicht von ungefähr.

Ich fand es auch ganz interessant (wenn ich nochmal den Bezug zu Frau Weßels nehmen darf), dass das als Suchmaschine wahrgenommen wird. Das liegt einerseits sicherlich am Interface, aber es wurde auch von Anfang an als die neue Version

der Suchmaschine gepusht. Das heißt, es wurde verkauft als die neue Art, Informationen zu suchen. Deswegen muss man verstehen, wie die Systeme funktionieren: nämlich wahrscheinlichkeitstheoretisch und statistisch, und wenn man das weiß, dann kann man diese Systeme auch sinnvoll einsetzen.

Natürlich hilft einem das immer noch nicht, Fake News oder getäuschte Bilder zu erkennen. Das Problem dabei ist: Wir wissen aus der Geschichte, dass wir gar nicht so wahnsinnig gute KI brauchen, um uns täuschen zu lassen, wenn das, was uns vorgelegt wird, unseren Interessen entspricht. Wir lassen uns alle nur zu gern täuschen von den Informationen, die unsere Vorannahmen treffen. Da spielt Bildung eine Rolle, da spielt auch das technische Erkennen von Fake News oder Bildern eine Rolle. Aber das ist nicht alles, sondern es geht auch um die Frage: Wie gehe ich mit scheinbar verführerischen emotionalen Informationen um, die mir genau das transportieren, was ich gerne möchte?

Wenn ich als letzten Kommentar noch auf diese Halluzinationen eingehen darf, die wir hatten: Ich finde den Begriff auch schwierig. Das erste Mal wurde er diskutiert in Bezug auf die fabrizierten Quellen von gefakten Journalartikeln, wo gesagt wurde: Da wurden Quellen angegeben von wissenschaftlichen Publikationen, und die waren generiert. Wenn man weiß, dass das einfach eine Wahrscheinlichkeitstheoretische Textproduktion ist, dann sind Quellen und Referenzen zu Studien auch nur Wortmuster. Und dann muss man sich nicht wundern, wenn man sich ein wissenschaftliches Paper generieren lässt, das auch diese Quellen – und das ist natürlich hochgradig problematisch.

Durch diese Kopplung von Suche und generativer KI können wir im Idealfall (wie Frau Weßels gesagt hatte) das Beste von zwei Welten haben, aber

im schlimmsten Fall das schlechteste von zwei Welten. Meine Sorge ist, dass es zu einer permanenten Vermischung kommt zwischen dem, was eine Referenz auf tatsächliche Quellen von außerhalb ist, und den wahrscheinlichkeitstheoretischen Inhalten. Ich glaube, es ist eine Bildungsfrage: Wie kriegen wir es hin, dass die Leute das richtig einsetzen können? Aber es ist auch nicht so trivial, wenn das in einem Interface vermischt wird.

Elisabeth Gräß-Schmidt (01:05:05)

Nun haben wir viele Fragen seitens des Publikums. Daher würde ich jetzt das Wort dem Publikumsanwalt Mark Schweda geben für die Fragen, die vom Publikum kommen.

Mark Schweda (01:05:33)

Moin aus dem schönen Oldenburg. Der Slido ist voll mit Fragen und Kommentaren, und ich will versuchen, sie ein bisschen zu gruppieren. Wir haben konkrete Nachfragen; dann gibt es Fragen rund um den konkreten Einsatz von KI im Unterricht und die Bedeutung von KI für die verschiedenen am Unterrichtsgeschehen beteiligten Parteien, und dann gibt es Fragen, die eher auf einer politisch-regulatorischen Ebene angesiedelt sind: Was muss man machen? Wie muss im Bereich Bildung, Erziehung reguliert werden?

Eine konkrete Rückfrage (ich glaube, an Judith Simon) war, noch mal die These zu erläutern, warum und inwiefern es eine Täuschung sei, mit KI zu interagieren. Wo liegt da die Täuschung?

Judith Simon (01:06:36)

Ich habe versucht deutlich zu machen, dass zum Beispiel in dem KI-Act eine Kennzeichnungspflicht für KI oder Chatbots gefordert wird. Da geht es darum, dass es eine Täuschung geben kann, wenn ich zum Beispiel mit einem Kundenberater oder einer Kundenberaterin interagiere

und denke, das ist ein Mensch, und das aber ein Chatbot ist. Da würde für diese sehr offensichtliche Form, dass ich einfach nicht weiß, ob ich gerade mit einem Menschen oder einem Chatbot interagiere, eine Kennzeichnungspflicht helfen.

Davon wollte ich die zwei anderen Formen von Täuschung unterscheiden: einerseits die Täuschung über Inhalte, über die auch viel diskutiert wird (Fake News und die generierten Bilder vom Papst), aber auch politisch brisante Bilder, wo gefälschte Bilder von Trump oder was auch immer produziert werden können.

Drittens gibt es auch eine Täuschung darüber (und die kann bewusst gemacht werden) von den Unternehmen: wenn Entwickler dieser Unternehmen auf einmal davon sprechen, dass ihre KI Bewusstsein oder Verständnis habe. Einerseits geht es darum, dass das auch so signalisiert wird, und zum anderen erscheint es Nutzerinnen und Nutzern durch diese Sprachform der Interaktion mit den Chatbots manchmal so, dass man den Eindruck hat, dass diese Systeme verstehen, und dass es durch diese Art der Interaktion auch zu einer Täuschung kommen kann. Diese unterschiedlichen Aspekte wollte ich auseinanderhalten.

Mark Schweda (01:07:53)

Fragen, die das konkrete Unterrichtsgeschehen betreffen, beziehen sich zum Beispiel darauf, wie KI soziale Interaktionen im Unterricht in der Schule, in der Klasse möglicherweise beeinflusst. Eine Frage, die ich da gleich mit anschließen möchte, ist die Frage (das geht in die Richtung von Felicitas Macgilchrist), wie wir verhindern können, dass der Einsatz von KI kritisches Denken und autonomes, selbstständiges Denken untergräbt.

Felicitas Macgilchrist (01:08:37)

Das sind zwei wichtige Fragen, die zwei unterschiedliche Antworten in mir aufrufen. Ich gehe

zuerst auf die zweite Frage ein, kritisches Denken. Mir scheint auch in diesem Fall der Weg eine klassische Bildungsaufgabe zu sein: Man nimmt einen Text und nimmt ihn auseinander. Dann ist der Text ein Prompt von den Schüler:innen auf dem Board und der Output davon. Deswegen habe ich vorgeschlagen, dass man mit der Expertise der Schüler:innen anfängt. Denn dann fällt ihnen viel schneller auf, was die Probleme sind und was sie für eine kritische Auseinandersetzung brauchen, als wenn man nach der Beschreibung der zehn wichtigsten Flüsse in der Welt oder so fragt. Also deren eigene Expertise, und dadurch kann man erklären, wie es funktioniert, dass das statistische Probabilitäten sind und auf welchem Trainingsdatensatz das beruht, und dann ist ein kritisches Bewusstsein da.

Eine Studentin von mir hatte vorgeschlagen – wir haben die Kennzeichenpflicht diskutiert, weil ich das als eine gute Idee begrüßt habe, und sie meinte: Ja, aber dann ist ja die Täuschungsmöglichkeit noch größer, wenn man weiß, wie man das hacken und etwas produzieren kann, was nicht da ist. Auch diese Fragen kann man diskutieren anhand der Beispiele, die man hat, ob Bild oder Wort, und auch die einfache Frage: Wessen Perspektive fehlt hier im Diskussionskontext?

Das geht dann auch zu der Frage der sozialen Interaktion: Wenn KI etwas ist, was jeder einzeln für sich benutzt, dann stört das eine soziale Interaktion. Wenn der KI-Output oder das KI-System hingegen zum Unterrichtsobjekt wird, dann bringt es eine soziale Interaktion wieder zusammen und ermöglicht jede Art von Interaktion.

Es gibt ganz andere KI-Systeme, die Wahrscheinlichkeiten von Erfolg und von Risiko auswerten usw., die die soziale Interaktion von Studierenden, von Schüler:innen in die Berechnungen einfließen lassen. Da zählt aber nur die soziale Interaktion, die in dem System gemacht wird, also

wenn sie in Chats miteinander sprechen und aufeinander reagieren. Ausgewertet wird nur das, was im System ist, und nicht das, was im Unterricht passiert. Das kann die Aufmerksamkeit der Lehrkräfte verschieben, denn diese Systeme aggregieren die Ergebnisse und verschieben sie zur Schulleitung. Die Schulleitung sieht dann aber nur die Interaktion im System und nicht die Interaktion im Unterricht.

Das wäre ein Problem, wenn nicht alle Lehrkräfte (und da habe ich volles Vertrauen in die meisten Lehrkräfte) wissen, dass das nicht das ist, was sie beobachten wollen, was sie lehren wollen und was sie gelernt haben wollen. Deswegen sehen wir in den Umfragen, dass viele Lehrkräfte auch manuell mit diesem System interagieren und der automatischen Auswertung nicht vertrauen.

Mark Schweda (01:11:42)

Vielen Dank. Gibt es dazu Ergänzungen aus der Runde? –

Wenn das nicht der Fall ist, würde ich eine Reihe von Fragen bündeln, die sich auf die Rolle der Lehrenden beziehen und ganz pragmatisch fragen: Was kann KI für die Lehrenden bringen? Und fokussiert auf den Bereich Prüfung: Was bedeutet es, Prüfungen, Leistungsüberprüfungen zu stellen mit ChatGPT, aber auch ohne ChatGPT? Wie geht das?

Doris Weßels (01:12:29)

Das ist eigentlich das spannendste Thema überhaupt: Wie wirkt sich das ganz praktisch auf Prüfungen aus? Im letzten Wintersemester, als das Thema akut wurde mit ChatGPT, hatte ich ein Modul genau zu diesem Thema. Im Sommersemester hatte ich die Idee, dass man mal generative KI in einem Modul zu KI und Ethik als Thema platzieren könnte, und dann kam es dazu, dass ChatGPT am 30.11., also in unserem laufenden Wintersemester veröffentlicht wurde.

Dann haben wir diskutiert: Wie gehen wir jetzt damit um? Wir wollten eigentlich am Ende eine klassische schriftliche Ausarbeitung schreiben. Dann habe ich das mit den Studierenden diskutiert und wir haben das ganze Verfahren modifiziert, und zwar habe ich drei Ps als Bausteine, die ich empfehlen kann. Das erste P steht für den Prozess, das zweite P für das Produkt, das am Ende entsteht, das Elaborat, die schriftliche Ausarbeitung, und das dritte P steht für die Präsentation, das heißt in der Regel ein mündliches Prüfungsgespräch oder Präsentation.

Das erste P ist der spannendste Punkt und ist auch das Novum: der Prozess. Ich hatte folgendes Bild im Kopf: Ich habe gedacht, das ist so, als seien wir alle unverhört und unangekündigt Leiter und Leiterin einer Schreibfabrik geworden. Wir haben aber nicht nur eine Schreibfabrik, sondern auch eine Bilderfabrik, eine Musikfabrik, und wir haben alle diese Fabriken sogar auf unseren mobilen Endgeräten. Wenn ich jetzt Leiterin einer Schreibfabrik bin, was ist dort meine Rolle? Meine Rolle ist eine verantwortliche Rolle, denn ich bin für den Produktionsprozess verantwortlich. Bei dieser Rolle, die man dort im Produktionsprozess einnimmt, muss ich gar nicht mehr als Mensch Hand anlegen, sondern ich delegiere das an diesen Automatismus dort und schaue quasi durch das Fenster in meine Schreibfabrik hinein. Aber die Art und Weise, wie ich diese Fabrik aufbaue, wie ich den Prozess gestalte, das ist jetzt meine Rolle, die ich als Mensch einnehme. Das heißt, mein Bausteinprozess bedeutet: Wie ist das ganze Forschungsdesign, das Methodische gestaltet? Das war früher auch schon drin, aber jetzt kommt noch die Auswahl der Werkzeuge, der Tools. Das heißt, wie digital kompetent habe ich die Werkzeuge für diesen Prozess ausgewählt? Das dürfen dann auch KI-gestützte Werkzeuge sein.

Das heißt, es war ein neuer Prüfungsbaustein da. Der bezog sich darauf, wie sinnstiftend und klug, mit welcher Kompetenz Tools und Werkzeuge ausgewählt wurden, die man dort in diesen Prozess integriert hat. Das war sozusagen das neue P.

Dann war die Frage: Wie macht man das jetzt? Profitiert derjenige, der die tollen Tools findet? Da hab ich gesagt: Ja, der oder die profitiert. Denn diese Tools zu finden, mit den Tools kompetent umzugehen und zu zeigen, dass das zu guten Ergebnissen führt und sinnstiftend und zielführend ist, das würde ich als eine hochwertige Kompetenz ansehen, die man hinterher – wir wollen ja auch Employability, wir wollen den Nachwuchs für die Zukunft qualifizieren, und von daher ist das eine wichtige Kompetenz. Aber das erfordert von uns Lehrenden, dass wir diese Kompetenz auch haben. Das heißt, wir sind täglich gefordert, uns mit all diesen Werkzeugen, diesen Tools und dieser unglaublichen Dynamik zu beschäftigen, um das auch bewerten zu können.

Das ist die größte Herausforderung, die ich in diesem Jahr immer wieder wahrgenommen habe, auch in vielen Gesprächen, weil mir immer wieder zurückgespiegelt wurde: Ich komme nicht mehr mit. Ich kann gar nicht mehr mithalten. Jeden Tag neue Tools, neue Werkzeuge. Wie soll ich die ausprobieren, wie soll ich mir da eine Meinung bilden? Ich muss das bewerten können. Das geht mir genauso. Wir alle spüren, glaube ich, so eine Atemlosigkeit, eine Überforderung angesichts der unglaublichen Dynamik. Aber mir hat damals, als die Frage, wie man damit umgeht, ein akutes Problem wurde, das Bild der drei Ps geholfen und insbesondere dieses Bild: Auf einmal bin ich Leiterin einer Schreibfabrik, und wie bewerte ich jetzt meinen Produktionsprozess? Da gibt es aus dem Qualitätsmanagement klare Anforderungen: Es zählt nicht nur das Produkt am Ende. Es

zählt auch die Qualität des Prozesses der Produktion.

Mark Schweda (01:17:06)

Dann hole ich gleich noch die dritte Partei, die hier sehr interessiert ist, mit ins Boot: Das sind die Eltern. Es gibt Fragen im Slido, die darauf abzielen, wie die Rolle der Eltern zu verstehen ist, wie die Eltern ihre Kinder begleiten können und was sie selbst dafür wissen und können müssen, um ihre Kinder beim Umgang mit KI im Unterricht und in der Schule, im Bildungswesen kompetent begleiten zu können.

Felicitas Macgilchrist (01:17:48)

Ich würde es nicht anders sagen als zu den Lehrkräften, denn das Schwierigste wird es sein, es zu verbieten. Es gibt viel Beratung für Eltern, die auf die Bildschirmzeit eingeht, und die Tendenz ist: verbieten oder minimieren. Es gibt eine alternative Position, die empfiehlt, in einen Austausch zu treten und mit den Kindern darüber zu sprechen. Wenn die Kinder darüber sprechen können, was sie spielen, was sie nutzen, was sie schauen und welche Tools sie nutzen, ob sie ChatGPT und Co. nutzen, dann können sie ein vertrauensvolles Gespräch mit den Eltern führen. Wenn die Eltern das nur reduzieren und verbieten wollen, dann können die Kinder nicht darüber sprechen, aus der Gefahr heraus, dass noch mehr verboten wird. Also das Gespräch zu suchen ist der erste Punkt.

Die Eltern müssen nicht viel mehr wissen, als dass das ein statistisches Modell ist. Sie können vielleicht dieses Ding ausprobieren, wenn sie selbst ein Konto haben, und wenn es nur das kostenlose ist. Zu sagen, dass die Eltern wissen, wo die Expertise der eigenen Kinder liegt, und sie einladen: „Lass uns mal kurz ausprobieren“, ist ein sehr pädagogischer Ansatz für Eltern. Aber grundsätzlich in einen Austausch zu treten und darüber zu

sprechen: „Was haltet ihr davon und was wisst ihr darüber?“ usw., ist die zentrale Empfehlung.

Mark Schweda (01:19:12)

Danke, das ist sehr konkret. Das kostenlose Konto reicht.

Der Sinn von Verboten leitet über zur Frage der Regulation. Was muss, was kann und was sollte auf einer politischen Ebene getan werden, um zu regulieren? Wie sind möglicherweise schon vorhandene Ansätze einzuschätzen? Hier wurde insbesondere auf den Europarat eingegangen, der sich neben dem KI-Act der EU speziell mit dem Bereich Bildung beschäftigt. Wie sind die laufenden Regulierungsbemühungen einzuschätzen, gerade mit Blick auf Bildung, und was täte not? Was bräuchte es?

Doris Weßels (01:20:16)

Aus meiner Sicht brauchen wir bei der Frage, wie wir mit dieser Technologie umgehen, eine andere Struktur des Umgangs als mit früheren Entwicklungen. Weil diese Entwicklung so schnelllebig ist und sich quasi täglich weiterentwickelt, hilft es nichts, wenn wir versuchen, in den klassischen Strukturen und Mustern zu regulieren, weil wir dieser Entwicklung immer hinterherhinken und immer weiter hinterherhinken werden, weil diese Entwicklung immer schneller wird. Das heißt, wir brauchen ein Konstrukt oder einen Mechanismus, um ein kontinuierliches Monitoring dieser Entwicklung zu garantieren, um dann rechtzeitig oder möglichst schnell Maßnahmen oder etwas Regulatorisches durchzuführen.

Erschwerend kommt hinzu: Neben der Geschwindigkeit ist diese Entwicklung auch nicht vorhersehbar. Wenn wir jetzt im Headquarter von OpenAI oder bei Google sitzen würden, hätten wir wahrscheinlich ein bisschen mehr Weitblick und könnten besser vorhersehen, was kommt. Aber ChatGPT ist ja wie ein Meteoriteneinschlag zu

uns gekommen. Es gab keine Vorwarnung, wir hatten keine Vorbereitungszeit. Die weitere Entwicklung wird genauso sein. Irgendwann wird Apple – meine Wette war: AppleGPT kommt noch in diesem Jahr, aber ich glaube, die Wette verliere ich, zumindest läuft die Zeit gegen mich; aber nehmen wir mal an, jetzt kommt AppleGPT. Wenn die den Markteintritt wagen, wird das sicherlich eine hochkarätige Variante sein, die vielleicht wieder ein Gamechanger ist, die uns wieder unvorbereitet trifft und uns wieder in helle Aufregung versetzt, und wenn wir dann schon irgendwas in der Tasche haben, was wir gern regulatorisch vorsehen würden oder schon eingeleitet haben, dann müssen wir das wieder überarbeiten.

Ich glaube, wir müssen uns den Realitäten dieses Multimilliardenmarktes, der von den Tech-Giganten der Welt bespielt wird, stellen. Ich habe in vielen Gesprächen mit Universitätspräsidenten und -präsidentinnen erlebt, dass die plötzlich in einer Rolle der Getriebenen waren und dass das eine kleine Ego-Problematik war, dass man auf einmal einer technologischen Entwicklung ausgesetzt ist, sich der ausgeliefert fühlt und nicht vorhersehen kann, wie es weitergeht. Dieses Gefühl, ein Getriebener zu sein in dem Prozess, dem man ausgeliefert ist, ist eine traurige Position, die wir derzeit haben. Deswegen ist es ja so wichtig, dass wir in eigene Sprachmodelle investieren und eigene Sprachmodelle so fit machen, dass wir mit denen gute Alternativen haben.

Felicitas Macgilchrist (01:23:20)

Ich finde das sehr wichtig, was Sie sagen, Frau Weßels, aber ich bin nicht sicher, ob das nicht auch das Narrativ der Entwickler:innen ist. Das Narrativ der Entwickler:innen ist: Das ist total gefährlich und wir kommen super schnell, und die Regulierung kann gar nicht so schnell sein, wie wir entwickeln.

Doris Weßels (01:23:39)

Nein, ich habe nicht gesagt, dass das nicht gehen kann. Ich habe nur gesagt, dass wir aus den bisherigen Prozessen und Vorgehensweisen, wie wir regulieren, rausmüssen. Das ist wirklich ein Paradigmenwechsel, den wir auch strukturell wagen müssen.

Felicitas Macgilchrist (01:23:55)

Überzeugend fand ich eine alternative Position auf Regulierung, die neulich auf Mastodon und so zirkulierte, wo ein paar Leute gesagt haben: Die Regulierung ist eigentlich gar nicht so schwierig und muss gar nicht immer aktualisiert werden, denn wenn wir regulieren, geht es um Ungerechtigkeit. Es geht um kollektive und gesellschaftliche Ungerechtigkeiten. Wenn benachteiligte Personen und Gruppen noch mehr benachteiligt werden, dann tritt die Regulierung ein. Und das sind Aspekte, die sich nicht geändert haben, auch wenn sich die Technik geändert hat. Ich bin keine Rechtswissenschaftlerin. Deswegen weiß ich nicht, wie das funktioniert. Aber das scheint mir als Leitplanke immer wieder einen sehr wichtigen Gedanken in das Narrativ einzubringen, wenn überlegt wird, wie jetzt reguliert werden sollte.

Es gibt zwei wunderbare Texte. Einer ist von Nathalie Smuha. Sie hat vor zwei Jahren die Entwürfe zum AI Act auseinandergenommen und gesagt: Es geht sehr viel auf individuelle Rechte ein und nicht unbedingt auf gesellschaftliche Benachteiligungen usw. Das fand ich eine überzeugende Position.

Dann gibt es von Salome Viljoen alternative Ideen, wie reguliert werden kann, wie die Governance funktionieren kann und wer wie über welche Daten, die dahinterliegen, verfügen kann. Das fand ich auch eine sehr wichtige Position.

Noch kurz zu den Eltern: Bei Eltern und Lehrkräften habe ich nur gesagt, dass sie ausprobieren sollten usw. Wichtig zu ergänzen ist, dass sie nach einem kurzen Ausprobieren gemeinsam entscheiden und die Kinder vielleicht auch selbst entscheiden: „Das will ich nicht, denn das finde ich zu spooky“ oder „das finde ich zu unplausibel und deswegen will ich darauf verzichten“. Das wäre eine sehr wichtige Entscheidung, die junge Menschen treffen können.

Elisabeth Gräß-Schmidt (01:25:54)

Zur Regulierungsfrage: Es ist ein großes Anliegen des AI Acts gewesen, in besonderer Weise zu regulieren, um die Vertrauenswürdigkeit der KI zu stärken, und das auch zum Markenzeichen der europäischen KI-Entwicklung zu nehmen. Da muss man auch schauen, und auch die Frage, wie diese Vertrauenswürdigkeit erreicht wird und wie sie sich von anderen Entwicklungssystemen absetzen kann, müsste man weiterdenken und einordnen in der Weise, wie Sie es gesagt haben, Frau Weßels.

Judith Simon (01:26:38)

Es sind einige Punkte aufgemacht worden, die deutlich machen: Wenn wir über Regulierungsfragen reden (und nicht nur da), reden wir über Machtfragen. Das ist ganz klar geworden. Die Frage ist: Wer entscheidet? Das hatte Frau Weßels schon betont, was ja interessant ist, warum wir über – also ChatGPT und die Diskussion über generative KI haben uns gefühlt alle überrollt, und zwar auch Leute, die im KI-Bereich dazu geforscht haben, weil es nicht um die Entwicklung in der Technologie ging, sondern darum, dass auf einmal ein simples Tool frei verfügbar im Internet war. Das heißt, das war eine wirtschaftliche Entscheidung, etwas verfügbar zu machen, und das sind solche Wendepunkte, wo, sagen wir mal, Unternehmer Gesellschaftsexperimente machen und die Externalitäten, also die

Folgen outsourcen: Wir machen das jetzt einfach mal und gucken, was passiert.

Diese Akteure setzen aber auch die Narrative, die Erzählungen, wie wir über KI sprechen. Ein Bereich, den ich da beispielhaft finde, war der UK AI Safety Summit, der mich sehr geärgert hat aus dem schlichten Grund, dass Begriffe wie AI Safety, aber auch Frontier AI gepusht worden sind von OpenAI. Das heißt, auf einmal redete man nicht mehr über KI-Regulierung im Allgemeinen, sondern nur noch über Safety, was ja nur ein spezifischer Fokus ist, und zum anderen ging es nur noch um Frontier AI, das heißt um zukünftige, fiktive Technologien, und nicht mehr um die, die wir jetzt gerade verwenden. Das sind Diskursverschiebungen, und dann springen aber alle dorthin; auch alle Minister gehen nach England und nehmen daran teil.

Wir müssen also gucken, dass wir auf der einen Seite eine effektive Regulierung haben, für die man eine gewisse Kompetenz braucht, aber dass man sich auch nicht einwickeln lässt von bestimmten Erzählungen von mächtigen Akteuren. Das finde ich nicht trivial.

Zum anderen geht es um geopolitische Fragen: Wie kann man sich positionieren in einem Feld? Da haben Politikerinnen und Politiker natürlich unterschiedliche Anforderungen: einerseits an den Wirtschaftsstandort in einem bestimmten Land und andererseits an den Schutz von Bürgerinnen und Bürgern. All das auszutarieren mit dem Blick darauf, wer gewinnt und wer verliert, finde ich nicht trivial und das macht es nicht leichter. Die Erzählung um die vertrauenswürdige KI war der Entwurf, etwas entgegensetzen, eine bestimmte Art von KI-Entwicklung, die sich an bestimmten Werten orientieren wollte. Da wurde Vertrauenswürdigkeit als eine Art Metawert dargestellt, der unterschiedliche Komponenten hat wie Sicherheit, Zuverlässigkeit usw. Ich halte das

weiterhin für sinnvoll. Aber es ist alles andere als trivial und man muss sich das als einen großen Machtkampf vorstellen, und die Frage ist, wo man da aktiv werden kann.

Ich möchte noch etwas zu den Lehrer:innen und zu den Eltern sagen und etwas verstärken, was Frau Macgilchrist und Frau Weßels gesagt haben: Die Grundkompetenz, die wir brauchen, ist ein Grundverständnis von Statistik und Wahrscheinlichkeitstheorie. Ich glaube nicht, dass es um Programmierkenntnisse für alle gehen muss. Es muss auch nicht darum gehen, dass man diese Tools alle kennt, weil man sowieso überrollt wird, aber dass man im Bildungsbereich die Grundprinzipien kennt, dass man versteht, wie schließende Statistik funktioniert; nicht, dass jeder alle Tests (U-Test und T-Test und was auch immer) kennen muss, sondern dass man einfach verstanden hat, wie das Schließen funktioniert. Denn das spielt eine große Rolle in vielen Bereichen unseres Lebens, und wenn ich irgendeinen Bildungsauftrag an bestimmte Kernkompetenzen noch mitgeben würde, dann wäre das ein Bereich, den ich ganz wichtig finde.

Felicitas Macgilchrist (01:30:22)

Es wäre schön, wenn man das etwa politische Statistik nennen würde, um die Machtfrage in der Statistik in dieser zentralen Kompetenz mitzudenken.

Elisabeth Gräb-Schmidt (01:30:32)

Wir sind am Ende einer sehr reichen, zukunfts-trächtigen und zukunfts herausfordernden Diskussion angekommen. Ich denke, das wird uns in Zukunft begleiten und beschäftigen. Es sind radikale Veränderungen für den Bildungsbereich angesprochen worden, aber auch für die gesellschaftliche Entwicklung insgesamt. Wir alle sind gefordert. Das haben alle drei Referentinnen herausgestellt. Wir müssen uns einarbeiten in die neue

Thematik der künstlichen Intelligenz und dürfen uns da auch nicht scheuen. Wir müssen auf diesem Gebiet Urteils- und Verantwortungsfähigkeit gewinnen. Verantwortlichkeit bedeutet immer auch (was wir noch wenig angesprochen haben, aber was uns begleiten wird), die Diskriminierungsfragen gründlich zu überlegen und darauf zu achten, wie sie minimiert werden können.

Es geht in der Schule für die Lernenden und künftigen Erwachsenen auch darum, und das wird ein Hauptziel der Bildung bleiben (und da sollte KI auch unterstützen und es nicht verhindern), dass wir urteilsfähig und vor allem fit für die Demokratie-Erhaltung die Menschen heranziehen, damit sie mündig werden, die Gesellschaft mitzugestalten in einer Weise, die die Freiheit des Urteils nicht scheut.

In diesem Sinne entlasse ich Sie in den Abend und in weitere Diskussionen. Bleiben Sie wachsam. Bis zum nächsten Mal.