

Einfach mal so tun, als ob – Dominik Lukeš über KI in der Hochschullehre

(Dr. Birgit Hawelka im Gespräch mit Dominik Lukeš)

Birgit Hawelka:

Hallo zusammen und herzlich willkommen zu dieser Podcast-Folge. Mein Name ist Birgit Hawelka. Und das heutige Thema hat unter Mitgliedern von Hochschulen auf der ganzen Welt eine intensive Diskussion ausgelöst. Wir werden uns auf generative künstliche Intelligenz konzentrieren, insbesondere auf ChatGPT und ähnliche Anwendungen, deren Auswirkungen auf die akademische Arbeit und die Hochschullehre. Wir wollen die potenziellen Vorteile und Herausforderungen für die Wissenschaft beleuchten. Und ich freue mich, dass ich mit einem sehr kompetenten Partner über dieses Thema diskutieren kann. Zu Gast ist Dominik Lukeš. Dominik arbeitet am Zentrum für Lehren und Lernen an der Universität von Oxford. Er erforscht und lehrt unter anderem, wie Technologie die Produktivität beim akademischen Lesen, Schreiben und Organisieren verbessern kann. Dominik, herzlich willkommen und vielen Dank, dass du dir die Zeit nimmst und dein Fachwissen mit uns teilst.

Dominik Lukeš:

Vielen Dank, Birgit.

Birgit Hawelka:

Also lass uns gleich anfangen. Wie wir alle wissen, hat OpenAI im letzten Herbst ChatGPT 3.5 veröffentlicht, und zum ersten Mal war ein leistungsstarkes KI-Tool für die breite Masse verfügbar. Doch bevor wir die Auswirkungen auf die akademische Arbeit diskutieren, sollten wir einen Schritt zurückgehen. Wie intelligent ist diese künstliche Intelligenz eigentlich?

Dominik Lukeš:

Das ist sowohl eine sehr gute Frage als auch eine sehr schlechte Frage. Denn man muss sich auch fragen, wie intelligent die Menschen wirklich sind? Richtig? Was bedeutet also Intelligenz? Wenn man die Geschichte der künstlichen Intelligenz bis in die 1950er Jahre zurückverfolgt, dann haben die Menschen über Intelligenz vor allem in Bezug auf die Dinge nachgedacht, die intelligente Menschen tun, richtig? Zum Beispiel Schach spielen, Mathematik, Wissenschaft, Fremdsprachen und so weiter. Und Computer waren sehr aufregend. Es sah so aus, als könnten sie all diese Dinge sehr schnell, sehr zuverlässig und viel kompetenter

als Menschen erledigen. Es schien so, als ob es nur noch ein paar kurze Schritte bis zur Nachahmung der Intelligenz von Menschen sein würden. Der erste Vorschlag hierfür lautete: Geben wir uns einen Sommer in Dartmouth, und dann werden ein paar Expert:innen auf dem Gebiet den Sommer nutzen und es herausfinden. Wir werden einen großen Fortschritt machen. Und das war in den 50er Jahren. Nun, in den 70er Jahren, nach vielen, vielen Höhen und Tiefen, prägte ein KI-Forscher den Satz, der oft wiederholt wird. Sein Name war Moravec, und er prägte den Satz „Leichte Dinge sind schwer und schwere Dinge sind leicht“. Und wenn es um künstliche Intelligenz geht, sind all die Dinge, die uns schwerfallen, wie das Addieren und Multiplizieren von Zahlen und das zuverlässige Speichern vieler Informationen, das Abrufen von Informationen, für uns schwer, aber für Computer leicht.

Aber die einfachen Dinge, wie zum Beispiel das Gesicht von jemandem zu sehen, wie die Dinge, über die wir nicht einmal nachdenken, einfach heranzugehen und ein Fenster zu öffnen oder einen Apfel zu betrachten und zu wissen, dass es ein Apfel ist und ihn einfach zu erfassen und ein Wort in einem Wörterbuch oder in einem Text zu betrachten und alle seine Feinheiten sofort zu verstehen, ohne zu denken, ohne etwas abzurufen. Das sind Dinge, die für uns einfach sind. Wir denken nicht darüber nach, und jedes Kind kann sie tun. Oder jemand mit schwerwiegenden geistigen Defiziten kann das, jemand, der aus irgendeinem Grund einen niedrigen IQ hat, kann das tun, jeder kann das tun, aber hey, KI kann das nicht. Das sind also die leichten Dinge für uns oder die schweren. Und plötzlich hat sich das alles geändert. Als die neueste Generation der künstlichen Intelligenz aufkam, sagtest du, dass dies sehr plötzlich geschah, aber in gewisser Weise war es gar nicht so plötzlich. Die große Veränderung fand etwa 2010 oder 2012 statt, als die Menschen begannen, effizientere Rechenmethoden zu verwenden. So wurden zum Beispiel Grafikkarten für Berechnungen mit sogenannten neuronalen Netzen verwendet, die es schon seit den 1940er Jahren gab. Die Idee der künstlichen Neuronen stammt aus den 40er Jahren.

Aber plötzlich hatten wir all die Rechenleistung und einige nette kleine Tricks und viele Datensätze. Plötzlich hatten wir zum Beispiel die berühmten Bilddaten, die etwa zu dieser Zeit herauskamen. Und innerhalb weniger Jahre konnte man mit dem Bildnetz-Datensatz und den GPUs, den grafischen Verarbeitungseinheiten, den Grafikkarten, die die Leute für Spiele verwenden, plötzlich ein Bild erkennen, was man für unmöglich gehalten hatte, oder? Noch ein paar Jahre zuvor konnte niemand sagen: Zeige ein Foto von einer Katze, und der Computer konnte sagen, dass darauf eine Katze zu sehen ist. Aber jetzt wurde es auf einmal viel besser. Und es war nur diese kleine Veränderung. Aber es gibt keine Intelligenz. Wir haben einfach angefangen, diese einfachen Dinge für uns zu machen, und die schwierigen Dinge für einen Computer sind nun einfacher für Computer. Und dann haben wir diese Revolution einfach übernommen. Und seitdem gab es noch ein paar weitere Änderungen, und die letzte kam 2017 mit der Einführung von etwas, das Transformers genannt wird, und noch effizienteren Trainingsmethoden und größeren Datensätzen und mehr Überlegungen dazu. Und seither haben wir diese Revolution erlebt. Wir hatten GPT 2 herausgebracht, und das hat eine Menge Leute erschreckt. OpenAI, die Macher:innen von GPT und ChatGPT und der GPT-Reihe, weigerten sich sogar, es zu veröffentlichen, obwohl es „offen“ heißt. Früher haben sie alles für die Öffentlichkeit freigegeben. Weil sie Angst hatten, dass es für schlechte Dinge verwendet werden könnte.

Und es stellte sich heraus - und das ist jetzt etwa vier Jahre her -, dass es so schlecht und nutzlos war, dass gar nichts passierte und so ließen sie es einfach gut sein. Und dann haben sie schließlich GPT 3 herausgebracht und es war viel weniger zugänglich, und das gibt es jetzt auch schon seit über zwei Jahren. Die Leute konnten das benutzen. Und der große Durchbruch kam erst vor eineinhalb Jahren, als sie herausfanden, dass, wenn sie diese Modelle nur ein wenig trainieren und für dieses Training menschliche Bewerter:innen einsetzen, diese sagen, ob dies die richtige Art von Frage ist, und ihnen beibringen, wie sie im Wesentlichen zu Chatbots werden. Dieses große Upgrade der Chatschnittstelle, das herauskam, wurde der Öffentlichkeit zugänglich gemacht und sorgte im Dezember letzten Jahres für großes Aufsehen. Aber das hat den ganzen Unterschied ausgemacht. Plötzlich kamen all diese Dinge zusammen: eine einfache Schnittstelle, einfache Fragen und Antworten, die echte Antworten auf echte Fragen sind. Und das Internet explodierte. Die Leute fingen einfach an, damit herumzuspielen. Und wir sehen viele Verbesserungen, viele Veränderungen. Seitdem sind andere Konkurrent:innen aufgetaucht, die KI anbieten, oder Konkurrent:innen, die in diesem Bereich tätig sind, andere Alternativen, über die wir sprechen können. Das ist der Weg, den wir eingeschlagen haben. Zu keinem Zeitpunkt haben wir Intelligenz in den Mix eingebracht. Du hast gefragt: „Wie intelligent ist sie?“ Im Wesentlichen haben wir einfach größere Datensätze und effizientere Methoden, um Informationen aus diesen Datensätzen so zu extrahieren, dass wir auf der Grundlage unserer Eingaben etwas generieren können.

Und das gilt auch für Bilder und Texte. Das ist also in vielerlei Hinsicht eine große Revolution, aber es ist keine Revolution der Intelligenz. Es ist eine Revolution in dem Sinn, dass einfache Dinge, die für Computer schwierig waren, plötzlich viel einfacher für Computer geworden sind. Aber es ist keine Intelligenz in dem Sinne entstanden, wie wir Intelligenz verstehen würden. Auch wenn die Maschinen jetzt bei einem IQ-Test, bei einem traditionellen IQ-Test, eine wirklich hohe Punktzahl erreichen können. Aber das liegt nur daran, dass die einfachen Dinge einfach geworden sind und nicht die schweren. Das ist der Grund. Diesem System ist jedoch kein Intelligenzmodul hinzugefügt worden. Es macht einfach immer noch das Gleiche wie vorher, nur dass wir jetzt größere Datensätze und effizientere Berechnungsmethoden zur Verfügung haben.

Birgit Hawelka:

Ich verstehe. Es ist also nicht intelligent im menschlichen Sinne. Allerdings scheinen viele, vor allem Akademiker:innen, recht vorsichtig zu sein, mit diesem Programm zu chatten. Natürlich benutzen es viele, aber andere sind ziemlich zurückhaltend. Neulich habe ich mich zum Beispiel mit einem Professor unterhalten, der ChatGPT nur ungern benutzt, weil er Angst hat, dass das Programm aus jeder seiner qualitativ hochwertigen Eingaben lernt und er so dazu beitragen könnte, das System noch leistungsfähiger zu machen. Hat er Recht?

Dominik Lukeš:

Nein, aber ein bisschen, ja, aber eher nein. Die Funktionsweise dieser Systeme besteht im Wesentlichen darin, dass mehrere Module übereinander gestapelt sind. Und an der Basis all dieser Systeme sitzt etwas, das als großes Sprachmodell bezeichnet wird. Und das enthält den gesamten Code, die Intelligenz, das gesamte Modell der Welt, das weiß, was als nächstes

kommt. Und diese Modelle sind unglaublich schwer zu trainieren, unglaublich teuer, und sie brauchen wirklich gute Daten, die zuverlässig sind, d. h. riesige Mengen und einigermaßen saubere Daten. Wir wissen es nicht genau, aber es wird geschätzt, dass das Trainieren der großen Sprache, die GPT 4 zugrunde liegt, Millionen und Abermillionen oder vielleicht sogar 100 Millionen Dollar kostet und Wochen bis Monate dauert. Man hat also einen Computer, der all diese Daten immer und immer wieder auf so komplizierte Weise durchforstet, und das bedeutet, dass man es nicht live tun kann, um ihm neues Wissen zu vermitteln. Es gibt kein Online-Lernen. Es ist also im Grunde nicht so, wie man es sich bei Menschen vorstellt. Und das ist ein weiterer Unterschied zwischen Menschen und ChatGPT. Die Menschen lernen jedes Mal, wenn ich etwas sage, ich lerne auch etwas. Ich verändere mich, aber ChatGPT verändert sich nicht. Es ändert sich nicht, egal ob ich es frage. Es gibt jedoch noch eine andere Ebene, die man Verstärkungslernen nennt. Und diese Schicht besteht aus menschlichen Bewerter:innen, die wirklich hochqualitative, kuratierte Beispiele geben und dann die Antworten bewerten. Das ist die dritte Ebene.

Auch das geschieht nicht live. Es passiert also nicht, während du dich im Chat befindest. Aber ChatGPT sagt in seinen Nutzungsbedingungen „Wir können einige der Daten, die Sie verwenden, nutzen, um uns zu helfen, das Modell in Zukunft zu trainieren.“ Aber es würde nicht das zugrundeliegende große Sprachmodell trainieren, aber es könnte einen Beitrag leisten und wir wissen nicht genau wie. Es besteht also eine gewisse berechtigte Besorgnis, aber nicht in dem Sinne, dass sich jemand einbildet, dass seine Eingaben einen so großen Unterschied machen werden. Und es ist unwahrscheinlich, dass eine einzelne Person tatsächlich einen so großen Unterschied macht.

Birgit Hawelka:

Okay, ich verstehe. Was auch immer wir von ChatGPT halten, es sieht so aus, als ob es dauerhaft da ist. Und Studierende werden es zur Unterstützung des Studiums nutzen. Und wenn wir an akademische Arbeit im Allgemeinen denken, denken wir vor allem an drei Kernbereiche: Informationen nachschlagen, lesen und schreiben. Das ist das, was Akademiker:innen normalerweise tun. Schauen wir uns also diese akademischen Aufgaben einmal genauer an. Und vielleicht sollten wir mit der Informationssuche beginnen. Natürlich können wir ChatGPT oder einer anderen KI eine beliebige Frage stellen und eine Antwort erhalten. Aber wie sehr können und sollten wir den Informationen, die ChatGPT liefert, vertrauen?

Dominik Lukeš:

Wir sollten ihm 0 % vertrauen. Wir sollten eigentlich aufhören, es als Schnittstelle zur Informationsbeschaffung zu betrachten. Es ist sehr verlockend, das zu tun, weil man ein Gespräch wie mit einem Menschen führen kann. Und sehr oft, in den allermeisten Fällen, wird es dir vielleicht sogar die richtigen Informationen geben. Aber das ist nicht das, wofür man es benutzen sollte wegen des berühmten Problems der Halluzination. Die Funktionsweise von ChatGPT oder diesen großen Sprachmodellen besteht darin, dass sie lediglich das nächste Wort vorhersagen, und das ist nicht einmal das nächste Wort. Es ist nur ein sogenanntes Token. Dabei handelt es sich um eine Art Teil eines Wortes, der durch einen Prozess abgeleitet wird, der im Wesentlichen die Daten in etwas Handhabbares komprimiert. Und jedes Mal,

wenn es ein Token hinzufügt, sagt es das nächste voraus und das nächste, das nächste. Es gibt eine Logik, die im Grunde nur sagt: Okay, das könnte als nächstes kommen und das erzeugt das. Im Grunde heißt es immer, okay, das könnte als nächstes kommen, das sieht gut aus. Das bedeutet im Wesentlichen, dass die Leute sagen: „Manchmal halluziniert es.“ Aber eigentlich halluziniert es die ganze Zeit. Es halluziniert nur dieses nächste Stück, das nächste Stück, das nächste Stück oder was auch immer, und es kommt immer etwas, das Sinn macht. Denn das ist zwar plausibel, aber es hat keinen Sinn für Wahrheit in sich, oder?

Du hast vorhin die Frage nach der Intelligenz gestellt. Eines der Merkmale unserer Erfahrung mit menschlicher Intelligenz ist, dass wir diese Art von Pausen haben. Wir können sagen, okay, warte mal: Macht das Sinn, oder kann ich das in Schritte aufteilen, oder kann ich irgendwo etwas nachschlagen? ChatGPT hat keinen Nachschlagemodus. Es hat keine Datenbank. Es kann nicht nachschauen, es hat nicht einmal Zugriff auf seine eigenen Trainingsdaten. Es gibt also nichts, was in diese Richtung gehen könnte. Es sagt buchstäblich nur: „Okay, hier ist eine Reihe von Token, lass uns ein paar mehr generieren. Für dich sieht das wie eine vernünftige Sache aus, wie etwas, das Sinn macht, wie eine Frage.“ Und dann erhalten wir so etwas, das wie eine Antwort aussieht. Aber ChatGPT unter der Oberfläche sieht nur aus wie eine Reihe von Token, die alle Zahlen sind. Es sind nicht einmal Buchstaben dafür. Es sind einfach nur Zahlen. Ihnen sind diese wirklich komplizierten Vektoren zugewiesen, die die Beziehung zwischen ihnen angeben, und es spuckt sie einfach aus. Und es sieht für uns so aus, als ob es irgendeinen Sinn ergibt, aber das tut es nicht. Es gibt keine Logik. Es gibt nur diese Vorhersage. Und es stellt sich einfach heraus, dass die Anzahl der Parameter, die es für diese Vorhersagen verwendet, jetzt in die Hunderte von Milliarden geht. Plötzlich kann es dies in mehreren Sprachen tun, kann Computercode schreiben, es kann diese vernünftigen Dinge machen, die so oft, wahrscheinlich häufiger als nicht, tatsächlich sowohl korrekt als auch plausibel sind. Aber wir sollten ihm niemals vertrauen. Dies ist ein ungelöstes Problem, das Problem der Halluzinationen. Und selbst wenn es oft sagt, zum Beispiel, ich habe diese Daten durch einen Prozess berechnet, sagt es nur, dass dies die Art von Dingen ist, die hier erscheinen könnten. Es tut das nie wirklich. Es hat niemals Zugang zu seinen internen Prozessen, so wie wir es vielleicht denken. Das ist also das Wichtigste. Wenn du also ChatGPT als Ersatz für Google ansiehst, dann benutzt du es im Grunde genommen falsch.

Birgit Hawelka:

Es ist also kein guter Rat, mit ChatGPT nach Informationen zu suchen.

Dominik Lukeš:

Nun, ja und nein. Es ersetzt nicht einmal Wikipedia. Weder Wikipedia, noch eine Enzyklopädie. Dafür ist es nicht gedacht, auch wenn es oft die richtige Antwort gibt. Man verlässt sich darauf. Man muss es immer nachprüfen. Aber es ist sehr nützlich, um die Informationen, die man findet, zu verarbeiten. Es stellt sich heraus, dass die Genauigkeit der Informationen manchmal nicht so wichtig ist wie die Fähigkeit, die Informationen zu verarbeiten und neue Denkstrukturen zu entwickeln. Und genau das macht ChatGPT so mächtig, weshalb es manchmal tatsächlich Sinn macht, es um etwas zu bitten, das nicht unbedingt Fakten sind.

So macht es zum Beispiel keinen Sinn zu fragen, in welchem Jahr jemand geboren wurde. Es mag die richtige Antwort geben oder auch nicht, aber warum eigentlich? Man kann das googeln und bekommt sehr schnell die richtige Antwort. Aber man könnte ihm zum Beispiel eine Frage über die Struktur eines Konzepts stellen. Man könnte also fragen, welches die wichtigen Konzepte im Bereich des Arbeitsgedächtnisses sind. Ich habe das zum Beispiel gefragt, und es wird dir eine Liste von Konzepten geben, und dann kannst du diese Liste weiter verwenden, um das zu entwickeln, was ich kognitives Gerüst nenne. Damit du besser in die Materie eindringen kannst. Du kannst es also bitten, ob es die Liste in Unterkategorien unterteilen, sie verlängern, verkürzen oder eine Tabelle daraus machen kann. Das sind die wesentlichen Elemente.

Wenn sich zum Beispiel Studierende auf eine Vorlesung vorbereiten oder ein neues Buch über ein bestimmtes Thema lesen wollen, lohnt es sich wirklich sehr. Bitte ChatGPT, dir eine Liste von Schlüsselkonzepten und schwierigen Begriffen zu geben. Bitte ChatGPT, diese auf verschiedene Arten zu erläutern, damit du dich darauf vorbereiten kannst, und nutze dann dein kognitives Gerüst, um dieses neue schwierige Gebiet zu verstehen. Das ist wirklich eine sehr nützliche Art, es zu nutzen. Aber mein Rat an alle: Nutze immer das, was ChatGPT dir gibt, als Hypothese zum Testen. ChatGPT wird dir also in vielen Bereichen sehr gute Antworten geben, aber es ist nicht wichtig, ob sie richtig sind oder nicht. Was zählt ist, dass es dir dabei hilft, diese Strukturen aufzubauen, sodass du dann diese Erfahrung nutzen kannst, um besser zu lernen, indem du bestätigst, ob das, was ChatGPT dir gesagt hat, richtig oder falsch ist. Und das ist der Punkt, bei dem ich denke, dass dies ein nützlicher Weg ist, es zu nutzen.

Birgit Hawelka:

Das wäre der nächste Punkt gewesen. Wenn ich es richtig verstehe, kann es uns beim Lesen von Texten helfen. Es kann einen Text zum Lesen vorbereiten, es kann uns helfen, die richtigen Fragen zu einem Text zu stellen. Habe ich es richtig verstanden?

Dominik Lukeš:

Ja. Es gibt also eine Reihe von Dingen, die es sehr gut kann. So kann man zum Beispiel einen Text einfügen und sagen „Erstelle mir Verständnisfragen, Multiple-Choice-Fragen zum Verständnis des Textes.“. Und auch hier gilt: Wenn du um zehn Verständnisfragen bittest, ist die Wahrscheinlichkeit groß, dass zwei von diesen zehn Fragen nicht wirklich gut sind, weil sie vielleicht schlecht formuliert sind, weil die Optionen schlecht sind oder weil es sich um eine Frage zu etwas handelt, das nicht im Text steht. Aber auch hier gilt: Solange man darüber nachdenkt, hilft es mir, den Text besser zu verstehen. Die Tatsache, dass die Frage falsch ist, spielt also keine Rolle, solange man weiß, dass sie falsch sein könnte, denn man kann sie nachlesen und bestätigen: Ah, okay, eigentlich lag ChatGPT hier falsch. Er hat sich nur eine Frage eingebildet. Aber das macht nichts. Es hat mir geholfen, den Text zu verstehen. Das ist also eine sehr wirkungsvolle Art, es zu nutzen.

Aber man kann es bitten, alle im Text erwähnten Personen usw. nachzuschlagen und eine Tabelle zu erstellen, oder verschiedene im Text erwähnte Begriffe. Auch hier muss man mit Ungenauigkeiten rechnen. Aber auch hier geht es darum, ein kognitives Gerüst aufzubauen. Das ist es, worum es geht. Ich habe das Programm zum Beispiel gebeten, eine Kurzgeschichte

auf Russisch zu erstellen, weil das auch gut zum Sprachenlernen geeignet ist. Sie handelte von einem Mädchen, das in eine Apfelplantage ging und stattdessen aus Versehen eine Birne fand. Es war eine nette kleine, lustige Geschichte, und ich bat das Programm, einige Änderungen daran vorzunehmen. Aber dann habe ich gesagt, okay, mach mir eine Liste mit Vokabeln, die mir beim Lesen dieses Textes helfen werden. Und es war wirklich gut. Alle Vokabeln waren sehr gut, bis auf eine, die statt einer Apfelplantage stand. Da war aus irgendeinem Grund eine Nussbaumplantage drin. Was nicht im Text stand. Aber für mich als Leser ist das wieder so eine Sache, die für Computer schwer ist, für mich leicht. Und natürlich weiß ich, dass es um Äpfel geht. Selbst wenn ich kein Russisch als Muttersprache habe, kann ich sagen: Okay, das steht natürlich nicht im Text. Aber all die anderen Dinge, die es für mich getan hat, waren für mich sehr nützlich. Das ist, glaube ich, wichtig. Solange man also daran denkt, dass man sich nie zu 100 % auf einen Text verlassen sollte, sondern ihn als Anregung für das eigene Denken nutzt, damit man sich dann mit dem Text und dem Thema, das man studiert, auseinandersetzen kann, ist es definitiv ein hilfreiches Werkzeug.

Birgit Hawelka:

Das klingt nach einer fantastischen Möglichkeit, KI in der akademischen Arbeit einzusetzen. Der Punkt, der in den letzten Monaten am intensivsten diskutiert wurde, ist jedoch die Frage, ob KI in der Lage ist, akademische Texte zu produzieren. Viele Dozent:innen befürchten zum Beispiel, dass Student:innen ihre Aufsätze von ChatGPT schreiben lassen können, was bedeuten würde, dass das Schreiben von Aufsätzen keine angemessene Methode mehr wäre, um Prüfungen abzulegen. Kannst du dir vorstellen oder glaubst du, dass ChatGPT in der Lage sein wird, Aufsätze in jedem Fach zu schreiben?

Dominik Lukeš:

Ja. Es ist tatsächlich eines der Dinge, die die Leute diesen ChatGPT-Moment genannt haben: das Ende der Hausaufgaben. Und plötzlich gibt es kein einziges Fach mehr, in dem ChatGPT nicht einen sinnvollen Beitrag zu deinem Ergebnis leisten könnte. Zum Beispiel kann ChatGPT selbst keinen ganzen Aufsatz mit mehr als 2000 oder 3000 Wörtern schreiben. Aber wie gesagt, so sollte man es nicht verwenden, wenn man versucht, etwas damit zu schreiben. Du kannst es benutzen, um eine Gliederung zu erstellen und dann die Gliederung ausfüllen. Ich habe zum Beispiel eine unserer Prüfungen hier in Oxford getestet, und ich habe einfach den Prompt gemacht, ich sagte „Okay, hier ist die Aufgabe für die Prüfung. Mach mir eine Liste mit den wichtigsten Punkten, die ich ansprechen möchte.“. Und dann habe ich wörtlich jeden Punkt eingefügt, um zu sagen „Erläutere das nur für diese fünf Punkte.“. Und um ehrlich zu sein, habe ich mich nicht um die Formatierung gekümmert, also habe ich keinen guten Aufsatz geschrieben, aber ich habe ihn eingeschickt und gesagt: „Wisst ihr was, das ist nicht sehr gut, aber ich würde es schaffen.“ Das ist es also, oder? Und schon sind wir wieder bei der Frage, was leicht und was schwer ist, nicht wahr?

Ich habe mir also meine ersten Tests dazu angesehen. Bevor ChatGPT veröffentlicht wurde, habe ich tatsächlich GPT 3 für diesen Zweck getestet. Und ich habe ihm eine Aufgabe gestellt. Schreibe mir etwas über die Konversationsimplikatur von Grice - das ist ein Philosoph. Und das tat es. Ich bat um drei verschiedene Versionen, nur um zu sehen, dass es sich immer

ändert. Und jedes Mal, bei allen drei Versionen, war der Vorname von Grice falsch. Mal war es John, mal George. Aber es passte alles. Also sagte ich: Okay, wen kümmert's? Und dann war es auch noch eine Arbeit, die er in den 1950er Jahren geschrieben hat, und die war richtig. Das heißt, er hat diese Arbeit tatsächlich geschrieben, aber es war nicht die Arbeit, in der es um die Implikatur von Gesprächen ging. Also sagte ich, okay, das ist offensichtlich nutzlos. Und dann habe ich ein paar Minuten darüber nachgedacht. Ich habe tatsächlich überlegt: Nein, wenn ein Student das schreiben würde, würde ich ihn durchlassen, weil ich denke, dass er sich mit den Konzepten auseinandersetzt. Denn wen interessiert schon der Vorname dieses nicht ganz obskuren, aber irgendwie nicht sehr berühmten Philosophen aus Oxford. Und wen interessiert es, ob er in den 50er oder 60er Jahren publiziert hat. Es ist mir also egal, oder? Mich interessiert es, wenn sich die Studierenden mit dem Thema auseinandersetzen. Und dann stellt sich plötzlich heraus, dass es eigentlich ganz einfach ist, diese Auseinandersetzung mit einem Thema vorzutäuschen.

Das machte mich nicht nur misstrauisch gegenüber der Art von ChatGPT, sondern auch gegenüber meinem früheren Einsatz von Aufsätzen. Ich habe schon früher darüber nachgedacht, aber es ist für Studierende manchmal ziemlich einfach, eine Fähigkeit zu entwickeln, die es so aussehen lässt, als würden sie sich intensiv mit etwas beschäftigen, obwohl sie das gar nicht tun. Wir alle haben schon einmal erlebt, dass wir aus einer Prüfung herausgegangen sind und dachten: „Na, das habe ich doch geschafft, oder? Ich habe mich in etwas hineingeredet, von dem ich wirklich keine Ahnung hatte.“ Und dann haben wir manchmal das Gegenteil erlebt. Ich habe wirklich fleißig gelernt und denke, dass ich es verstanden habe, aber ich konnte es einfach nicht in dem Aufsatz oder der Prüfung oder dem Gespräch mit dem Prüfer/der Prüferin zeigen. Wir wissen also, dass es das gibt, oder? Wir haben es alle schon erlebt. Aber wir vertrauen irgendwie blind darauf, dass, wenn wir sehen, dass Studierende diesen flüssigen Text über etwas schreiben, dies tatsächlich auf einen gewissen Lernerfolg hindeutet. Aber mit ChatGPT wissen wir jetzt, dass man das auch ohne Lernen, ohne Wissen, ohne Intelligenz oder tiefes Verständnis tun kann, indem man einfach Token für Token erzeugt. Ich würde also sagen, dass die Essays in gewisser Weise noch nie eine sehr gute Methode waren, um Studierende zu bewerten, und jetzt sind sie sogar noch schlechter. Denn wir wissen, dass sie mit diesem Instrument leicht erstellt werden können.

Das bedeutet nicht unbedingt, dass der Essay, wenn man ihn mit den richtigen Absichten schreibt, nicht auch Verständnis ausdrückt. Aber es stellt sich heraus, dass er auch ohne all diese harte Arbeit geschrieben werden kann. Wir werden nur einige raffinierte Methoden anwenden, um die Dinge zu umgehen, die man nicht tun will. Meine Antwort darauf wäre also: Nun ja, vielleicht sollten wir aufhören, Aufsätze zu verwenden. Man kann sie immer noch in einer Art individueller Prüfung durchführen, bei der die Studierenden einfach nur schreiben. Aber waren sie jemals sehr gut?

Aber die andere Frage, die größere Frage ist eigentlich: Was wird in zwei oder drei oder fünf Jahren sein? Was wird mit der ganzen Idee des akademischen Schreibens passieren? Im Moment haben wir dieses Modell des akademischen Schreibens: Jemand setzt sich hin und tippt Buchstabe für Buchstabe, aber das ist eine Art idealisiertes Modell, das wir haben, aber es ist nicht wirklich richtig, oder? Wenn man sich die Publikationslisten der Wissenschaftler:innen anschaut, ist man heutzutage eine(r) von 12 oder 10 oder 50 oder sogar 200 Autor:innen. Und

die meisten dieser Autor:innen haben nie etwas damit zu tun gehabt. Um ehrlich zu sein, steht mein Name unter einer Arbeit, an der ich nicht ein einziges Wort geschrieben habe. Und auch das kann passieren. Wir machen das also nicht immer, aber wir haben dieses Modell, dass jemand das Wort für Wort abtippen muss. Heutzutage muss das aber nicht mehr der Fall sein. Wir können ChatGPT oder diese generative KI verwenden, um eine Liste von Ergebnissen in Text zu verwandeln. Und dann muss der Mensch den Text ändern. Auch hier testen wir ständig Hypothesen. Wir können uns nie darauf verlassen, dass etwas genau ist. Aber ganz plötzlich... Wir haben die Eintönigkeit des Schreibens und wie viele Akademiker:innen das Schreiben einfach verachten. Ich habe Präsentationen gehalten, und die Leute sind zu mir gekommen und haben gefragt: „Oh, kannst du nicht einfach meine Arbeit für mich schreiben? Weil ich das Schreiben wirklich hasse.“ Also erfolgreiche, publizierte, vollendete Akademiker:innen. Einige von ihnen lieben das Schreiben, andere hassen es. Und sehen wir uns doch einmal an, welche akademischen Arbeiten produziert werden. Einiges davon ist furchtbar. Einiges davon ist schwer zu lesen. Es ist kompliziert, weil die Leute einfach viel zu tun haben oder nicht sehr gut darin sind. Warum also nicht plötzlich ein Werkzeug einsetzen, das die Lesbarkeit für alle erleichtert und auch den Akademiker:innen die Arbeit erleichtert? Und so stelle ich mir vor, dass in zwei oder drei Jahren, denn ich bekomme die Fragen schon jetzt, diese Art des Schreibens viel akzeptabler sein wird. Und da ich gerne schreibe, habe ich noch keine gute Möglichkeit gefunden, ChatGPT zum Schreiben zu nutzen. Aber ich kenne einige Leute, die sich mit akademischem Englisch oder Englisch im Allgemeinen schwer tun, vor allem Akademiker:innen aus aller Welt. Ich habe mit Leuten gesprochen, die sagten: „Okay, das spart uns so viel Geld, weil wir Korrekturleser bezahlen, richtig? Wir bezahlen Lektoren für diese Arbeit und plötzlich kann ChatGPT uns helfen, viel bessere englische Texte zu produzieren, als wir es vorher konnten.“ Und wir werden irgendwie auf einer einheitlicheren Ebene sein. Ich hoffe also, dass dies eine positive Veränderung für die Zukunft des akademischen Schreibens ist, denn es ist nicht unbedingt so einfach, wie die Leute denken.

Birgit Hawelka:

Ja, ich bin mir nicht ganz sicher, ob das passieren wird. Ich hoffe es aber. Als deutsche Muttersprachlerin benutze ich natürlich KI, um englische Texte zu schreiben, weil mein Englisch manchmal etwas unbeholfen klingt. Ich habe zum Beispiel ChatGPT benutzt, ihm einen Text gegeben und die Aufforderung gegeben: „Schreibe diesen Text wie ein Akademiker.“ Ich habe die Ausgabe einer englischen Muttersprachlerin gegeben und sie hat darüber gelacht und mir gesagt, dass niemand jemals so schreiben würde. Aber vielleicht gilt auch hier: Müll rein, Müll raus. Vielleicht war der von mir gewählte Prompt nicht gut.

Dominik Lukeš:

Ja, das ist eine andere Frage. Wir entwickeln diesen neuen Bereich des Wissens, nämlich wie man die Prompts für ChatGPT und diese anderen Tools formuliert. Und dafür ist sogar ein neuer Begriff aufgetaucht, der Prompt Engineering heißt. Und es ist nicht Engineering, es ist Prompt-Formulierung. Aber ich ziehe es vor, es als Kontextgestaltung zu bezeichnen. Man muss einen guten Kontext schaffen, damit das Tool mehr generieren kann, richtig? Denk also

daran, dass es wirklich nur um den Kontext geht, bevor du neue Dinge verwendest. Wenn du also den Kontext so formulierst, dass er neue Dinge hervorbringt, wirst du bessere Ergebnisse erzielen. Nun, das ist eine nützliche Tatsache, die man wissen sollte. Eine berühmte Art der Eingabeaufforderung für ChatGPT ist zum Beispiel die Gedankenkette. Wenn man ChatGPT eine Aufgabe stellt, z.B. „Erkläre die Relativitätstheorie.“ oder etwas anderes und sagt „Denke Schritt für Schritt.“. Und nur weil man es bittet, Schritt für Schritt zu denken, sieht es plötzlich so aus, als würde es inneres Denken offenbaren, aber das tut es nicht. Es wird immer noch nur Text erzeugt. Aber dadurch, dass es diesen Token-Schritt nach und nach erzeugt, entsteht ein reichhaltigerer Kontext. Plötzlich stellt es also Kontext her. Das bedeutet, dass es mit größerer Wahrscheinlichkeit bessere Ergebnisse erzielen wird.

Das Gleiche gilt auch für Texte. Wenn du also versuchst, als Akademiker:in zu schreiben, kannst du vielleicht akademisch in den Naturwissenschaften bleiben, akademisch in deinem speziellen Bereich. Du kannst jetzt auch ein Beispiel geben. Du könntest ihm die Information geben: „Hier ist ein Absatz aus einer Zeitschriftenartikel, und hier ist das, was ich geschrieben habe. Kannst du das, was ich geschrieben habe, im Stil dieses Absatzes umschreiben?“

Und noch einmal: Es tut nichts unter der Oberfläche, was du in deinem Kopf tun würdest. In deinem Kopf würdest du das erste Stück lesen. Hier sind also einige Funktionen, kann ich sie replizieren? Nichts davon, nichts davon passiert. Alles, was passiert, ist das Generieren von Token für Token, Token für Token. Es wird einfach ein Token nach dem anderen hinzugefügt. Wohlgemerkt, das sind keine Wörter. Aber weil du diesen ausführlichen Kontext spezifizierst, führen die Vorhersagen, die die neuen Token generieren, mit größerer Wahrscheinlichkeit in die von dir gewünschte Richtung. Und das ist wirklich die Zukunft der Sache. Das ist die Art und Weise, wie wir anfangen müssen, über die Interaktion mit diesen Tools nachzudenken.

Birgit Hawelka:

Leider läuft uns die Zeit davon, aber ich möchte gerne eine letzte Frage stellen. Natürlich kann niemand, auch du nicht, die Zukunft vorhersagen. Aber welche Schritte könnten aus deiner Sicht Lehrende oder Institutionen unternehmen, um eine verantwortungsvolle und effektive Nutzung von ChatGPT in der Hochschullehre zu gewährleisten?

Dominik Lukeš:

Nun, zunächst möchte ich deiner Vermutung widersprechen. Ich glaube, ich kann die Zukunft vorhersagen. Und meine Vorhersage ist sehr einfach: Die Zukunft ist ungewiss. Wir leben in einer sehr aufregenden Zeit. Für manche Menschen ist das etwas sehr Negatives, für manche Menschen etwas Positives. Aber wir leben auch in einer Zeit des schnellen Wandels, und es kommen neue Dinge auf uns zu. Jede Woche gibt es neue Ansätze, jede Woche gibt es etwas Neues, Interessantes zu entdecken. Es gibt so viel zu lernen.

Und eines der Dinge, die dies interessant machen, ist, dass es so viel intellektuelle Begeisterung für diese Thematik gibt. Aber was die Sache für jemanden, der nicht intellektuell begeistert ist, schwierig macht, ist die Tatsache, dass es sich um eine komplette Flutwelle von Informationen und Krach und so weiter handelt. Ich würde also sagen: Das größte Problem für einzelne Hochschulangehörige, für Menschen innerhalb dieser Institutionen, ist es, einen

Weg zu finden, um zu lernen und eine eigene Praxis zu entwickeln. Man sollte also nicht unbedingt daran denken, wie die Studierenden es nutzen könnten, sondern es als ein Werkzeug betrachten, das man selbst nutzen könnte.

Und ich würde sogar fast sagen, dass es in einigen Bereichen fast unverantwortlich ist, es nicht zu nutzen. Denn stell dir vor: Wenn du produktiver sein kannst, warum nicht produktiver sein? Warum sollte man als Lehrender nicht [davon Gebrauch machen]. Ich habe das bereits erwähnt, ich habe dieses Beispiel der Texterstellung gebracht. Wir geben den Studierenden nie genug Beispiele, weil sie so schwer zu finden sind. ChatGPT oder diese Tools können wirklich gute Beispiele liefern. Und dann kann man sie als Experte bzw. Expertin schnell überarbeiten, um sicherzustellen, dass sie wirklich korrekt sind. Und es können Fragen dazu gestellt werden. Wie gesagt, wir geben den Studierenden nicht genug Fragen zum Verständnis, nicht genug Gelegenheiten zum Üben. Plötzlich kann man mehr und bessere Lernerfahrungen für die Studierenden machen. Also mach das!

Und plötzlich fängt man an, diese Themen zu lernen, und auch hier gilt: Man sollte es nicht übertreiben. Viele Leute beginnen ihre erste Abfrage, und das war auch meine erste Abfrage, indem sie versuchen, ChatGPT zu entlarven. Okay, zeigen wir mal, wie dumm es ist, wie groß die Halluzination ist, die ich daraus machen kann. Aber warum? Wir wissen bereits, dass er halluziniert. Wir sollten es nicht als etwas betrachten, das genau ist. Wir sollten es als etwas betrachten, das uns das Leben erleichtern kann, und herausfinden, auf welche Weise es das tun kann. Und wir sollten lernen, wo es ungenau oder begrenzt ist, und das dann in den Arbeitsprozess einbauen, oder? Wir haben das mit Taschenrechnern gemacht. Wir haben es mit allem gemacht, warum also nicht auch damit? Man sollte also diesen Prozess durchlaufen. Ich denke, das ist die Lektion für jeden Einzelnen, und das ist im Moment ein verwirrender Zeitpunkt. Wenn man zum Beispiel ChatGPT benutzt... Aber gerade letzte Woche hat ein Konkurrent von OpenAI, den Machern von ChatGPT, namens Anthropic, die zweite Version ihres Chatbots namens Claude veröffentlicht. Claude ist der kostenlosen Version von ChatGPT sehr ähnlich und ermöglicht es sogar, eine Datei hochzuladen. Aber er hat eine sehr interessante Funktion, die sich von ChatGPT unterscheidet. Es hat ein sehr langes Kontextfenster. Ich habe ja schon über den Kontext gesprochen und noch nicht erwähnt, dass er begrenzt ist. In ChatGPT sind es ein paar Tausend, 8000 Token oder 4000 in der kostenlosen Version, was ungefähr 75% der englischen Wörter entspricht, weniger in Deutsch und anderen Sprachen. Anthropic hat ein Kontextfenster mit 100.000 Token, was je nach Sprache 50.000 bis 75.000 Wörtern entsprechen kann. Plötzlich kann man also viel mehr damit machen. Es hat seine Vorzüge und Nachteile.

Man muss sich also Gedanken darüber machen, welche Werkzeuge man verwenden sollte. Wie lernt man am besten die Grundlagen des Prompting-Designs oder der Kontextgestaltung, die man vornehmen muss, um die besten Ergebnisse zu erzielen. Das sind also alles Dinge, mit denen man anfangen könnte. Aber die größte Gefahr hier, das höre ich ständig, und ich habe viel Zeit damit verbracht, mich mit Akademiker:innen und Studierenden darüber zu unterhalten, wie sie ChatGPT nutzen. Und ich habe keinen Unterschied in der Komplexität festgestellt. Ich glaube, wir hören immer wieder, dass wir den Studierenden beibringen müssen, wie sie es verantwortungsvoll nutzen können. Aber offen gesagt, glaube ich nicht, dass wir dazu bereit sind. Wir wissen nicht, wie wir es richtig nutzen können nur für uns selbst, oder?

Also lasst uns noch einmal sicherstellen, dass wir wissen, wie man es richtig macht. Wir können über Plagiate und Ethik reden, aber das ändert sich nicht wirklich, oder? Ich meine, wenn man behauptet, etwas geschrieben zu haben, was man gar nicht geschrieben hat, dann ist das Betrug. Und das war früher Betrug, und ist heute Betrug. Aber bei den subtileren Themen rund um große Sprachmodelle und diese generative KI sind wir noch nicht in der Lage, den Studierenden beizubringen, wie man es richtig macht, und die meisten Akademiker:innen wissen es auch nicht. Wir müssen sie also erst fortbilden.

Und das bringt mich zu der Frage, was die Institutionen tun sollten. Die Hochschulen sollten damit beginnen, intern Kapazitäten aufzubauen. Ich schlage also gerade die Idee eines KI-Labors für generative KI innerhalb der Universität vor. Wir können einige dieser Dinge testen, bei denen die Leute ihre Prüfungen mitbringen und sagen können, okay, wie würde das aussehen, wenn ich ChatGPT dafür verwende? Die Leute könnten anfangen, darüber nachzudenken, welche Möglichkeiten es gibt, Materialien für Studierende zu entwickeln, und welche Möglichkeiten wir in der Büroarbeit nutzen können, um die Produktivität zu steigern. Aber das wäre auch ein Ort, an dem wir all diese schnellen Entwicklungen verfolgen können, oder? Weil das so ist, gibt es keinen Ort, an den ich sie jetzt schicken kann, um ihnen alle Informationen zu geben, denn das alles passiert in Blogs, auf Twitter, in Preprint-Veröffentlichungen, auf all den vielen YouTube-Kanälen. Es passiert so viel in dieser multimodalen Welt der Information. Und ich würde nicht erwarten, dass ein Einzelner den Überblick darüber behält, so wie ich es getan habe. Deshalb denke ich, dass die Institutionen die Tatsache anerkennen sollten, dass man nicht einfach eine schnelle Checkliste geben und erwarten kann, dass jede(r) damit zurechtkommt. Das ist etwas ganz Neues. Das ist eine große Veränderung, und wir müssen diese Veränderung unterstützen und die sich schnell entwickelnde Situation im Auge behalten.

Birgit Hawelka:

Das ist einerseits fantastisch und andererseits auch ein bisschen beängstigend, muss ich zugeben. Und auch wenn es noch viele Punkte zu diskutieren gäbe, tut es mir leid, dass wir unser Gespräch an dieser Stelle beenden müssen. Nochmals vielen Dank, Dominik, dass du unser Gast warst. Ich bin mir wirklich sicher, dass deine Einblicke eine wertvolle Perspektive geboten haben und wesentlich zu einem besseren Verständnis von ChatGPT beigetragen haben. Herzlichen Dank dafür.

Dominik Lukeš:

Gerne, vielen Dank für die Einladung.