Wissenschaft
Praktische Methoden
LebensPhilosophisches

KI-Blog 2026
KI – tool und Akteur

KI-Blog 2024 erstes Halbjahr

Blog zu KI erstes Halbjahr 2024

Beitragsseiten

Blog zu KI erstes Halbjahr 2024
Free Lunch?
Neosokratisch
Aus zwei wird eins
Das Rahmenproblem
Der Rechts-Sprung
Wissensallmende
Hard Science?
Kollision
Blinder Fleck
KI schlägt Bachelor?
Das nicht! (B)
Das nicht! (A)
Prompt
Test
Motivator
Wissenschaftstheorie
Studierende
Urknall 2022
Dialog
KI zitieren
Durchschnittlichkeit
Methodische Strenge
Qualität
Wiki
Explainable XAI
Alle Seiten

Innovationen, die durch KI angestoßen werden, folgen in atemberaubendem Tempo aufeinander. Dieser Blog versucht, wesentliche KI-Aspekte in Bezug zur tertiären Bildung aber auch aus philosophischer, erkenntnistheoretischer, psychologischer und soziologischer Sicht essayistisch aufzubereiten.

Das Format von Essays – so quasi in lockerer Kaffeehausatmosphäre – erscheint hierfür geeigneter zu sein als eine systematische Darstellung aufgrund eines statischen Konzeptes. Ein solches Konzept wäre bei dieser Thematik in kürzester Zeit obsolet.

„There is no free lunch in the Universe“

ist ein geflügeltes Wort, das in verschiedenen Zusammenhängen oder Disziplinen gerne verwendet wird, etwa in Wirtschaft, Physik oder Mathematik[1]. Soll heißen, es gibt nichts umsonst[2].

Evolution

Apfelbäume schenken uns Menschen ihre süß-sauren Früchte nicht, weil wir so lieb sind oder die evolutionär höchste Stufe auf dieser Erde erreicht haben. Sie tun das, weil sie über Millionen Jahre lernten, dass mobile Lebewesen die Früchte verspeisen, samt enthaltenen Samen.

Diese Samen (Apfelkerne) werden woanders ausgeschieden – praktischerweise zusammen mit einer sie umgebenden Packung Dünger. Voilà, die Samen keimen und wachsen zu neuen Apfelbäumen an einem anderen Ort heran. Wir mobile Wesen haben etwas geschenkt erhalten, leisten etwas dafür.

Aus der evolutionär langfristigen ‚Sicht‘ der Bäume sind wir Transporteure, die durch etwas Fruchtzucker leicht zu manipulieren sind. Wir nehmen die Rolle nützlicher Lastenträger ein, die das Baumgenom verbreiten.

Sommerliche Zitronentarte

Gratis- oder Bezahlinternet?

Das Internet hat uns in den letzten Jahren Ähnliches gelehrt. Wir bezahlen für die „Gratis“-Nutzung von Social Media-Plattformen im Nachhinein mit unseren Daten. Wir zahlen mit den elektronisch festgehaltenen Inhalten und Abbildern von uns selbst. Unsere Vorlieben und Schwächen werden von den Marketing-Abteilungen der Unternehmen in der Zusammenschau verschiedener Quellen punktgenau erkannt und ohne Rücksicht genutzt, um uns maßgeschneidert etwas zu verkaufen. In der Regel Unnötiges. :(

Es schaut aus, als wäre dieses *Konsumiere-jetzt-und-bezahl-später-ohne-dass-du-es-mitkriegst*-Phänomen der gegenwärtigen Internetwelt nur eine erste Runde gewesen. Die generierende KI läutet nun die zweite und vermutlich heftigere Runde ein. KI erzeugt einen Inhalt, der Fakt und Fiktion untrennbar fusioniert.

In diesem Verhältnis tut sich eine Schere auf: Fakten sind begrenzt, Fiktionen unbegrenzt. Fakten, also authentische Inhalte, werden zum Engpass, heiß umworben und umkämpft von den Netz-Giganten.

Das *NoFreeLunch-Theorem* rückt wieder in den Fokus, denn die generierende KI ist teuer. Entweder wir Nutzer zahlen (A) sofort einen Preis für den Zugang in die elektronische Welt und sehen relativ wenig umgefärbte Realität, oder wir meinen (B), die globalen Großtech-Player austricksen zu können, und stürzen uns (vorerst) ohne Bezahlung in die globale KI-Gischt von geschäumten Texten, Bildern und Tönen.

Natürlich male ich schwarz-weiß. Jedoch geht es darum, zentrale Antagonismen herauszuarbeiten. Es ist dann die Sache jedes Einzelnen, wie er oder sie mit diesen Extrempositionen oder Zwischenstufen umgeht.

Strategie "Bezahlung sofort"

Als Beispiel für die Strategie (A) ist Apple anzuführen. Die Anschaffungspreise für die Macs, iPads und iPhones sind deutlich höher als für ähnliche Geräte der Konkurrenz. Auch die Abos innerhalb der Apple-Welt generieren kontinuierlich immense Einnahmen.

Apple rechtfertigt die hohen Preise mit einer harmonisierten Hard- und Software (da alles aus einer Hand ist) und etwa einer damit einhergehenden erhöhten Datensicherheit. Im Juni 2024, gerade als die kritischen Stimmen („Apple hat die KI-Welle verschlafen“) unüberhörbar wurden, verkündet Apple eine Integration von KI auf seinen Geräten.

Eine Integration, die die Datensicherheit großschreiben soll: So werden einige KI-Funktionen nur lokal am Gerät laufen. Weitere Funktionen stehen innerhalb der iCloud auf Apple-Servern zur Verfügung. Dafür sollen eigene KI-Chips entwickelt werden. Erst in einer dritten Ebene werden Anfragen an die global verfügbare KI gestellt. Klingt aus Datensicht im ersten Moment überzeugend.

Da wäre nur ein Haken: Als strategischen KI-Partner hat sich Apple OpenAI ausgesucht. Für OpenAI unter Sam Altman als CEO steht Sicherheit in der Prioritätenliste nicht an oberster Stelle, wenn man veröffentlichten Interviews ehemaliger Mitarbeiter:innen Glauben schenken darf. Es wird spannend sein zu beobachten, welche Ergebnisse diese Kooperation liefern wird.

Strategie "Offener Zugang"

Als Beispiel für die Strategie (B) kann Meta genannt werden. Meta besitzt mit den drei bis vier Milliarden Nutzer:innen von Facebook, Instagram und Whatsapp einen riesigen Datenpool. Einen Datenpool, der insbesondere für die trainingsdatenhungrige KI von existenzieller Bedeutung ist.

Werden Texte, Bilder und Videos von Nutzern aus ihrem Alltagsleben für Trainingszwecke genutzt, so steigt die Wahrscheinlichkeit, dass authentische Daten, die nicht oder nur geringfügig verändert wurden, als Antworten gespiegelt werden. Was als eine Vorform von Identitätsdiebstahl aufgefasst werden kann. Die Nutzer sehen sich selbst dabei zu, wie sie etwas tun, dass sie nie gemacht haben.

Dies belegt, dass in dieser Strategie Nutzer für seine Internet-Aktivitäten mit seinem Selbst bezahlt. Beweisen die zahlreichen öffentlichen Beteuerungen von Mark Zuckerberg, Datensicherheit ernst zu nehmen, eine Umkehr bisheriger Meta-Geschäftspraktiken oder nur ein Bemühen, das Problem zu verschleiern? Die ebenfalls regelmäßigen und einseitig von oben diktierten Änderungen der Nutzungsbedingungen von Facebook und Co. zeigen, dass alles in Fluss ist. Und ob die Aktionäre von Meta tatsächlich zulassen, dass täglich Gratis-Mittagessen (free lunches) an Milliarden von Menschen ausgeteilt werden?

Fazit

Wir haben drei Optionen: (A) Wir zahlen bewusst mittels Euro, Dollar oder schwedischer Kronen beim Eintritt in die KI-gestaltete Welt für deren Wunder. (B) Wir zahlen unbewusst mit uns selbst als barer Münze im Strudel der Geschehnisse in der KI-geänderten Welt. (C) Wir mischen beide Strategien.

29. Juni 2024

PS:

In einer weiteren, gegenwärtig vorhandenen strategischen Variante (D) sind noch Reste der ursprünglichen Begeisterung, Freude und Motivation für ein freies und die Menschheit förderndes Internet enthalten. Diese Strategie zeigt sich teilweise bei Betriebssystemen wie Linux oder Open Access-Ansätzen in Wissenschaft und Internet-Plattformen.

Aber auch hier ist ein Kippen auf die andere Seite immer möglich, wie OpenAI vorzeigt.

[1] Hier hat es das geflügelte Wort sogar zu einem Lehrsatz geschafft, dem NFL-Theorem (NoFreeLunch Theorem). Dieser Lehrsatz wurde 1995 von Wolpert und Macready in „No free lunch theorems for search“ propagiert und verweist in deterministischen Bereichen auf die begrenzte Wirksamkeit von einzelnen Algorithmen zur effektiven Such-Optimierung. Inhärente Zielkonflikte in jedem System sind jedoch kaum zu überwinden.

[2] In der anschließenden Diskussion in der Informatik meinten einige Wissenschaftler, dass bei stetigen Zusammenhängen, bei Blackboxen und in einem evolutionären Kontext „free appetizer“ doch möglich sein können. Womit sich ein Fenster für die generative KI aufmacht. Eine Zusammenfassung ist zu finden in Burlacu, B. (2017). Tracing of Evolutionary Search Trajectories in Complex Hypothesis Spaces. Dissertation in technischen Wissenschaften an der Johann Kepler Universität Linz.

Der neosokratische Dialog

Fußballeuropameisterschaft 2024: Ein Spieler erobert den Ball am eigenen Sechzehner, läuft auf der rechten Seite mit zunehmender Geschwindigkeit in die gegnerische Hälfte, passt quer zur anderen Seite, läuft weiter in Richtung Tor. Er wird von der gegnerischen Verteidigung nicht mehr beachtet, erhält den Ball zurück, schlägt einen Haken, einen zweiten, schießt – Tor!

Diese Szene erinnert mich an die platonischen Dialoge von Sokrates. Harmlos anfangend, stellt Sokrates zunehmend unbequeme Fragen zu bisher als wahr geglaubten Annahmen seines Gegenübers. Er schließt mit einem fulminanten Argumentationsfinale, dem anvisierten Tor in seinem Spiel, in welchem sein Gesprächspartner am Ende dumm dasteht.

Als ich das erste Mal diese Dialoge las, vielleicht mit Ende zwanzig, war ich begeistert von dieser Methode, Illusionen zu durchschauen und für andere als solche sichtbar zu machen. Der Entschluss stand fest: Ich muss Philosophie studieren. Leider stieß ich in den nächsten Monaten auf einige Arbeiten der analytischen Philosophie, die damals modern war. Und mir drehte sich der Magen um.

Diese linguistischen Formalanalysen lasen sich wie eine Arbeit aus der Mathematik voller Symbole und Formeln. Der österreichische Mathematiker und Logiker Kurt Gödel hätte das sicher verstanden. Ich nicht.

Erst zwei Jahrzehnte später startete ich doch ein Philosophiestudium. Nun zwang mich der Wiener Studienplan zu einer Einführung in die Philosophie anhand eines Fichte-Textes. In einer Gegenthese zu Kant bietet Johann “Gottlieb” Fichte Argumente für den Glauben an einen Gott, wie ihn das Christentum lehrt.

Vom formalanalytischen Regen war ich in die paratheologische Traufe geraten. Diesmal allerdings hielt ich durch.

Der G'scheite und der Dumme

So wie ihn Platon schildert, war Sokrates das Manipulative nicht fremd. Es ist von Anfang an klar, wer „triumphieren“ wird. Es ist von Anfang an klar, wer nur über ein Schweinwissen verfügt, daher schonungslos bloßgestellt und der Dumme sein wird. Der sogenannte Dia-Log präsentiert sich nicht als eine zweiseitige, sondern als eine einseitige Sache.

Heiligt das Ziel, falsche Annahmen aufzudecken, den abwertenden Zweck? Ist nicht in einer aufgeklärten Gesellschaft ein faires Verhalten im Rahmen eines gemeinsamen echten Dialoges langfristig erfolgversprechender? Wir wollen – ähnlich wie Sokrates es für seine Gesprächspartner wünscht – selber denken lernen.

Venedig, Studierendenviertel / Campo Margherita: Diskussionen an einem Winterabend

Was hat das mit der KI zu tun?

Vieles! Jede Anfrage an die KI, jeder abgesendete Prompt, führt zu einer Antwort, die als erste Runde eines Dialogs angesehen werden kann. Wenn wir die KI-Antwort unbesehen übernehmen, also nicht checken oder nachfragen, stehen wir im Grunde als die Dummen da. Die KI hat die Rolle des weisen Sokrates übernommen und wir die Rolle des einfältig Fragenden.

Wir übersehen die Schwächen der KI, halten konstruierte Wahrscheinlichkeitsinhalte für die Realität. Die KI hat uns dann manipuliert, wenn wir annehmen, dass sie klüger sei als wir. Das antike Psychospiel zwischen zwei Menschen läuft nun zwischen Maschine und Mensch ab.

Das Manipulative an den sokratischen Dialogen ist seit langem bekannt. Bildungswissenschaftler (Leonard Nelson und Gustav Heckmann[1]) haben im 20. Jahrhundert eine faire Variante solcher Dialoge entwickelt. Es nennt sich das neosokratische Gespräch. Es geht um eine, von Vorannahmen wenig beeinflusste gemeinsame Suche in einer Gruppe nach übereinstimmenden Antworten.

Im Laufe der letzten Jahrzehnte wurden unterschiedliche Modelle hinsichtlich Ziele und Vorgehen für das neosokratische Gespräch vorgeschlagen. Übereinstimmende Merkmale dieser einzelnen Modelle sind:

- Sprich klar, kurz und in verständlicher Sprache.
- Nimm die Äußerungen des anderen ernst.
- Sprich Fragen und Zweifel an, aber zerstöre nicht das Gespräch.

Wie können diese Ideen im KI-Dialog umgesetzt werden?

Vier wichtige Punkte helfen in Anlehnung an das neosokratische Gesprächskonzept hierbei:

1. **Klarheit:** Im Prompt gezielte und präzise Fragen und Anweisungen geben.
2. **Exploration:** In einem anschließenden Dialog verschiedene Perspektiven erkunden. Die KI soll Vor- und Nachteile von Sichtweisen aufzeigen, Argumente für und wider bestimmte Thesen liefern.
3. **Kritisches Hinterfragen:** Weiterführende Erklärungen und konkrete Beispiele von der KI erfragen.
4. **Begriffsklärung:** Definitionen von verwendeten zentralen Begriffen anfordern und zueinander ins Verhältnis im Rahmen des relevanten Kontextes bringen.

Die umgesetzten Prinzipien des neosokratischen Gesprächs fördern eine strukturierte und tiefer schürfende Interaktion mit der KI. Niemand verliert, auch die KI nicht, denn sie weiß von diesem Psychospielchen nichts. Der Gewinner werden Sie allein sein.

24. Juni 2024

[1] Heckmann, G. (2018). Das sokratische Gespräch. Mit einem aktualisierten Vorwort von Dieter Krohn. Münster: LIT.

Wie aus zwei Blackboxen eine wird

Nun springt also auch Apple auf den KI-Zug auf. Allein und alle KI selbst entwickelt? Nein, mit OpenAI als Partner, also etwa ChatGPT im Hintergrund. Auf dem iPhone, iPad oder dem Mac bieten traditionelle Apple-Anwendungen Funktionen der KI an.

Sogar eine applikationsübergreifende Assistentin soll es auf den Apple-Geräten geben. Die hausbackene Siri wird zur informationsherbeizaubernden Fee.

Apple bewegt sich damit innerhalb eines Tsunamis an Software-Anbietern, die ebenfalls Funktionen der generativen KI in ihre Programme einbauen.

Natürlich entwickelt keines dieser Unternehmen KI neu. Sie nutzen hierfür APIs. Ein Application Programming Interface, kurz API, ist eine Programmierschnittstelle, die es erlaubt, andere Programme an ein System anzubinden. Schnittstellen aller möglichen Funktionen oder Tiefe – in Form eines einfach anzuwendenden Quellcode bis hin zum Maschinencode auf der tiefsten Ebene – sind alte Hüte. Alt, aber gut und daher Hut ab. Vernetzungen werden vor allem durch sie realisiert.

Also werden Anweisungen, die im Prompt geschrieben oder neuerdings gesprochen werden, durch APIs in die jeweilige generative KI gefüttert. Antworten kommen zurück und werden im anfragenden System angezeigt. So weit, so einfach.

Große Sprachmodelle wie ChatGPT oder Gemini sind sehr gut im Analysieren und Strukturieren von Sprache. Ihr Output hat aber Wahrscheinlichkeitscharakter; von der Realität oder gar der Wahrheit kann der Output weit entfernt liegen. So kann derselbe Prompt später anderes hervorbringen. Allgemeine KIs sind daher ungeeignet, objektiv Gegebenes zu suchen oder Fakten nachprüfbar zu verknüpfen.

Ein Weg, den Output näher an die Wahrheit zu bringen, ist, die KI mit rein spezifischen Daten zu einem konkreten Thema zu füttern. Dazu wird der Pool der Daten, mit denen die KI trainiert wurde, geleert und mit konkreten Themendaten neu angefüllt. Man nennt dies *Grounding*. *Ground Zero* heißt in der Militärsprache der Erdboden nach einer explodierten Atombombe oder einer Rakete. Auch das ehemalige Fundament des World Trade Centers in New York wird so bezeichnet. Alles, was vorher dastand, ist verschwunden.

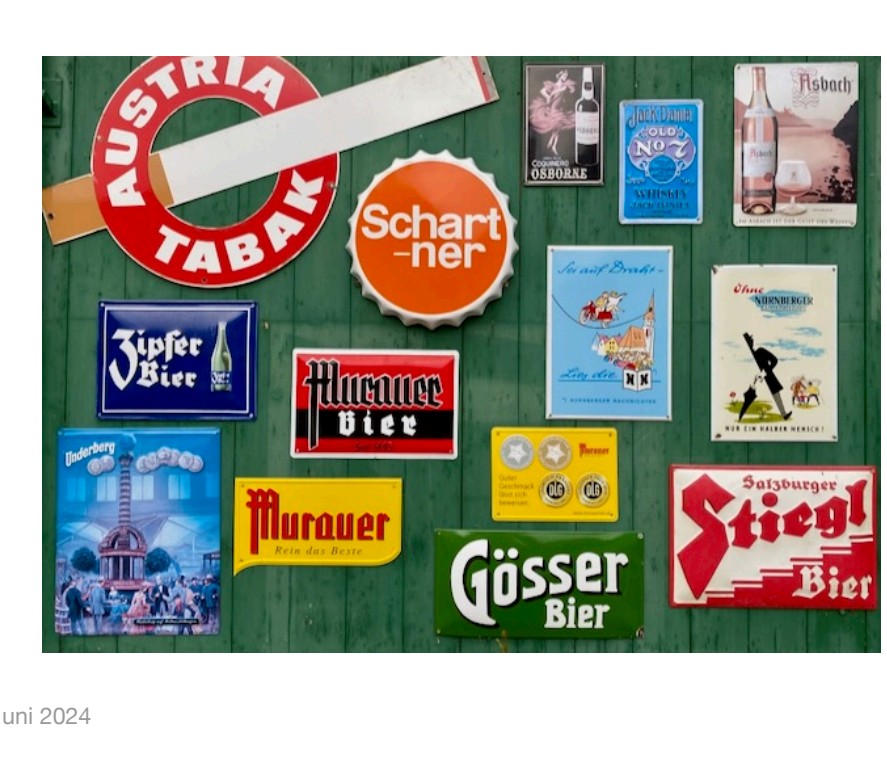
Will man etwa eine Themen-KI zum österreichischen Baurecht aufbauen, dann könnte man alle Gesetze, Verordnungen von Bund und Ländern, die Judikatur und alle relevanten Kommentare in den leeren Daten-Container der Themen-KI füttern und damit die KI trainieren. Die Antworten werden deutlich besser und verwertbarer sein als Antworten einer allgemeinen KI. Um einen letzten Check wird man dennoch nicht herumkommen. Die autonome, nicht beherrschbare Eigenschaft einer KI wird auch hier zum Vorschein kommen.

Was hat sich bei einer Themen-KI im Vergleich zu einer allgemeinen KI geändert? Eine der beiden Blackboxen, die Trainingsdaten selbst, sind nun themenspezifisch und transparent. Eine Blackbox wurde zur Whitebox. Die zweite Blackbox, die heuristischen Algorithmen der KI-Prozesse bleiben nach wie vor im Dunkeln.

Die Phantasie reicht kaum aus, um sich alle nur möglichen Themen-KIs vorzustellen. So wird derzeit an Nutzer-Avataren gearbeitet[1].

> Und so wird's gemacht: Der Nutzer füll einen Ground Zero mit Daten von sich selbst. Daten und Erzählungen von der Geburt, der erworbenen Bildung, den Erfahrungen durch Berufe, Partnerschaften.
>
> Kurz, die Zickzacklinien des Lebensweges werden samt ausgewähltem Kontext eingebaut. Ebenso enthält dieser Avatar Fotos und Videos des Nutzers und seine Stimme. Auch Verwandte oder Freunde können diesen Trainingscontainer mit Daten über den Nutzer füllen. Und voilà, selbst die Urenkel sehen den Avatar und hören seine Geschichten.
>
> Die Geschichten werden durch die KI häufig leicht verändert sein und vielleicht sich der Welt der zuhörenden Generation anpassen. Oder diese Avatar-KI wird im Museum des 21. Jahrhunderts zu bestaunen sein. So wie wir jetzt die ersten Fotografien aus dem frühen 19. Jahrhundert betrachten.

Die Themen-KIs entsprechen abgeschotteten Mini-Universen. Wir werden in naher Zukunft mit einem Schaum unzähliger kleiner KI-Blasen konfrontiert werden. Nun denn, schöne neue Multiversumswelt!



18. Juni 2024

[1] Sudholt, E. (05.06.2024). Frage: Hast du Angst vor dem Tod? Antwort: Mein Schöpfer hat darüber noch nichts mitgeteilt. Die Zeit Nr. 25/24. https://www.zeit.de/2024/25/kuenstliche-intelligenz-tod-michael-bommer-krankheit-leben.

Das Rahmenproblem

„Künstliche Intelligenz scheitert am Rahmenproblem. Es ist unlösbar.“

Prof. Josef Roppert, Institut für Statistik, HS f. Welthandel, Wien

So hieß es in den Jahren 1971 oder 1972 in einer Vorlesung an der Hochschule für Welthandel in Wien. Die ersten elektronischen Taschenrechner waren gerade eingeführt worden. Diskutiert wurde heftig, was Computer uns bringen könnten.

Hierbei ging es nicht nur um die Leistungen dieser kleinen Wunderdinge, sondern auch um Artificial Intelligence (AI) als übergreifendes Thema. Was sie in Zukunft noch vollführen werde, ob sie Menschen überflügeln und bedrohen könne, um schlussendlich die Herrschaft über die Welt anzutreten.

Die eingangs erwähnte Aussage fiel im Rahmen einer Lehrveranstaltung mit dem anspruchsvollen Titel „Höhere Mathematik“. Wir Welthandelsstudenten plagten uns mit Bilanzrechnungen und Statistik. Ein Schnuppern in anderen Sphären versprach Abwechslung.

Also „höhere“ Mathematik plus Diskussion zu aktuellen Themen. Die Darstellung mündete hier in einer einzigen finalen Erkenntnis: Nein zur Weltbeherrschung durch die AI (oder KI). Das geht einfach nicht! Eine kompakte Schlussfolgerung, keine Herumrederei, einfach beeindruckend für junge Studenten, wie ich es einer war. Diese Aussage hatte sich in mein Gedächtnis eingebrannt.

Mehr als ein halbes Jahrhundert später läuft die Diskussion – befeuert durch die generative KI – wieder heiß. Aber Überraschung: Der Satz gilt für AI-Forschende nach wie vor[1,2,3]. Er verweist auf Fundamentales. Nur spricht man heute vom Kontext, nicht vom Rahmen. Der Begriff hat sich geändert, die Sache an sich nicht.

Reicht das, um von grenzen- und rahmensprengender Innovation zu sprechen? Klar, es gibt neue Ergebnisse in neuen Anwendungen. Auch bricht die generative KI momentan in viele Lebensbereiche ein. Aber der Rahmen – sorry, der Kontext – bildet die maximale Geschwindigkeit, die Lichtgeschwindigkeit. An ihr scheitert die KI.

Blick aus der tiefstgelegen Schutzhütte des Österreichischen Alpenvereins
am Buschberg im Weinviertel

Dieser endgültige Rahmen lässt sich aus zwei Sichten verdeutlichen:

AI-Sicht nach außen: Die AI besitzt kein Weltmodell, das alle Dimensionen und Sachverhalte beinhaltet. Und damit sind wirklich alle gemeint. Diese allumfassende Welt reicht von Quantenfeldern im Superkleinsten über nichtlebende Materie und lebende Materie wie Mikroben, Organismen, Pflanzen, Pilze, Tiere bis hin zum Menschen.

Sie muss unsere private Welt, die materielle und soziale Umwelt, die lokale, die übergeordnete regionale, nationale, globale politische, wirtschaftliche und militärische Welt bis zum gesamten Sonnensystem beinhalten. Und darüber hinaus sollte sie Myriaden von Galaxien voll von Myriaden Sternen und Planeten bis an den Rand des Universums (wenn es diesen überhaupt gibt) inkludieren. Das alles sollte nicht statisch vorhanden, sondern in interaktiven Prozessen aller Felder und aller Dinge geschehen.

KI-Modelle, wie wir sie nun erleben, beschränken sich auf jene Daten, mit denen sie trainiert wurden. Diese sind Abbildungsdaten und nicht reale Daten. Jedes KI-Modell ist beschränkt auf jenes System, für das es entworfen, entwickelt und trainiert wurde. Und wenn die Daten noch so umfangreich werden, die skizzierten Systemgrenzen bleiben eisern gezogen.

Um dies anhand eines Beispiels zu demonstrieren: Den Pflege-KI-Roboter musste man spezifizieren, wo er sich bewegen darf, dass er nur bestimmte Türen öffnen, die weiteren Möbelstücke im Zimmer jedoch unberührt lassen soll. Ausdrücklich musste man ihm bisher alle Szenarien und Bedingungen des Handelns im Rahmen von fixierten Algorithmen beibringen. Neu ist jetzt, dass dieser Roboter Heuristiken wie trial and error etc. anwendet, um seine Aufgaben zu erfüllen. Ok.

Setzt man ihn aber auf einer Wiese ab, wird er trotz der Wahrscheinlichkeitsheuristiken im Grund nicht wissen, was er tun soll. Er wird leerlaufen oder ewig Kreise ziehen und Türen und zu pflegende Personen suchen.

Programmiert man ihn hierfür neu, sodass er vielleicht die Wiesenblumen gleich mitpflegen kann, ist das fein. Führt einige Zeit später der angrenzende Bach Hochwasser und überschwemmt die Wiese, ist alles perdu. Er hätte wasserdicht sein müssen und vielleicht schwimmen können. Und was hätte er getan, wären die zu pflegenden Personen in dieser Wiese gesessen?

Für komplexe und sich ändernde Umgebungen sind alle relevanten Faktoren neu zu bestimmen. Die KI muss sich im neuen Rahmen, dem neuen Kontext, dem neuen Kosmos zurechtfinden. Sie muss wieder Relevantes von Irrelevantem unterscheiden lernen, um Handlungsfolgen in diesem neuen Kontext voraussehen zu können.

Doch irgendeine Instanz muss von außen Eckpunkte vorgeben. In einer dynamischen Umwelt unter unvorhergesehenen Bedingungen versagt das entworfene KI-System. Ein Selbstlernen mittels Heuristiken unter unvorhergesehenen Umständen ist nicht möglich.

AI-Sicht nach innen: Das eben präsentierte Beispiel führt in den inneren existenziellen Rahmen der KI. Die KI muss nicht nur geschaffen, sondern auch gewartet, repariert, gepflegt sowie an neue Gegebenheiten angepasst werden.

KI benötigt dazu viele menschliche (!) Helferlein. In Zukunft vielleicht auch nichtmenschliche Helferlein wie externe Software-Apps für Datentransport, Datenverfügbarkeit und Datensicherheit, die dann selbst wieder Helferlein der zweiten Ordnung benötigen.

Hierbei ist interessant, dass ein Programmierfehler im IT-Jargon *bug* heißt, Käfer. Dies stammt aus jenen Vor-Transistor-Zeiten, als Insekten in die Leitungen eines Großcomputers gerieten und dadurch einen Kurzschluss erzeugten. Ein kleiner *bug* legte auf diese Weise Teile oder den ganze Rechner lahm. Diesen *bug* zu finden und zu entfernen, bedurfte es Kopf- und Handarbeit. Die äußere Instanz des Menschen hat alles geregelt. Einer solchen äußeren Instanz bedarf die KI auch in Zukunft.

Fazit: Das Rahmenproblem bleibt die fundamentale Schwierigkeit in der KI-Forschung. Effektiv und effizient funktionieren KI-Systeme nur in jenen realen Umgebungen, die sie vorher mühsam erlernt haben.

15. Juni 2024

PS: Selbst die gefeierten elektronischen Hilfen wie die anfangs erwähnten Taschenrechner benötigten lange, bis sie im täglichen Leben ankamen. Noch Jahre später kalkulierte ich Marketing-Budgetzahlen mit dem Rechenschieber. Dies ist ein Gerät, das auf logarithmischen Funktionen basiert Erfindungen des frühen 17. Jahrhunderts.

[1] Bennett, M. (2023). A Brief History of Intelligence: Why the Evolution of the Brain Holds the Key to the Future of AI. William Collins.

[2] Smith, B. (2019). The Promise of Artificial Intelligence. Reckoning and Judgement. London: MIT.

[3] Marcus, G. & Davis, E. (2019). Rebooting AI: Building Artificial Intelligence We Can Trust. Pantheon.

Jus/Jura –
der Sprung aus der Vergangenheit in die Zukunft

Das Recht ist überwiegend gekennzeichnet durch die Anwendung der Vergangenheit auf die Gegenwart.

Gegenwärtige Sachverhalte werden aufgrund von Normen bewertet, die in der Vergangenheit entwickelt und kodifiziert wurden. Raub oder Diebstahl genauso wie Verleumdungen oder körperliche Verletzungen werden bestraft. Sowohl die Gesetzesinhalte als auch das Ausmaß der Sanktionen hängen von den gesellschaftlichen Auffassungen der vorhergehenden Generationen ab. Änderungen dauern lange. Nur mühsam wurde etwa aus der Vergewaltigung in der Ehe ein strafbares Delikt.

Die rechtliche Bewertung des Umgangs mit anderen Menschen, Lebewesen und Dingen hat daher eine rückwärtsgewandte, bewahrende, also konservative Schlagseite. Eventuell passen Richter:innen in ihren Entscheidungen dies den gegenwärtigen (zum Teil geänderten) gesellschaftlichen Auffassungen an, soweit es in ihrem Ermessen liegt. Dies ist in der Struktur des westlichen Rechtssystems und somit auch in der westlichen Rechtsphilosophie angelegt. Es geht konzeptuell nicht anders.

Das sizilianische Taormina im Jahr 2000 als Symbol für liebliche Vergangenheit

Oder doch?

Die Vorreiter in der Umsetzung von generativer KI sind derzeit Vertreterinnen von angewandten Rechtsdisziplinen! Man glaubt es auf den ersten Blick kaum. Urheberrecht, Arbeitsrecht, Baurecht, Privatrecht wie Vertrags-, Erb- oder Familienrecht, Medizinrecht, Versicherungsrecht, Handelsrecht, Verkehrsrecht, Umweltrecht etc. All diese anwendenden Rechtsfächer haben das unmittelbare gesellschaftliche und wirtschaftliche Leben im Fokus. Sie regeln nicht nur unser Leben, sondern auch dessen spezifische Details[1,2].

Gerade noch überwiegend vergangenheitsorientiert, agieren diese Rechtsdisziplinen plötzlich an vorderster Front in der Handhabung (und wenigstens teilweisen) Disziplinierung von KI. Die Stellungnahmen von KI-Expertinnen aus dem Recht sowie anlaufende Projekte haben Hand und Fuß. Sie zeigen, wie KI in der Gegenwart und in naher Zukunft anwendbar gemacht werden kann. Es haben sich europaweite Rechtsnetzwerke gebildet, welche Projekte unterstützen, die die Umsetzungs-Potenziale von KI bewerten.[3] Der reale Druck zu reagieren ist in diesen Handlungsfeldern sehr groß.

Wo bleiben die sprachwissenschaftlichen Fächer?

Ein erster subjektiver Eindruck drängt sich auf, dass Vertreterinnen von geistes- und kulturwissenschaftlichen Fächern geradezu paralysiert auf diese KI-Revolution blicken. Man hört von ihnen kaum, sie scheinen wenig dazu zu forschen oder zu diskutieren. Wo sich doch gerade linguistische Fächer im Zentrum des genKI-Geschehens befinden. Denn sprachwissenschaftliche Fächer zeigen auf, wie Sprache uns Menschen formt und wie Menschen Sprache nützen, um die Welt zu gestalten.

Schnelle, wenn auch oberflächliche Recherchen zeigen, dass hier durchaus veröffentlicht wird. Jedoch haben diese Veröffentlichungen nicht annähernd jene unmittelbare pragmatische Brisanz, wie sie rechtliche Bewertungen haben.

Schrift und Recht

Bei näherer Betrachtung wird einiges klarer. Die geschriebene Sprache fungiert als universale Basis von Recht. An ihr hängt alles, inklusive der daraus folgenden Bestimmungen und Konsequenzen für unser Leben. KI kann Sprache sehr gut analysieren und gemäß den Anforderungen wiedergeben. Dadurch ist sie im Handumdrehen zu einem innovativen Instrument geworden, das die Rechtsbasis im Kern betrifft.

Das Durcharbeiten zahlreicher Aktenordner in Gerichtsverfahren kostet Zeit, Mühe und Geld. KI reduziert das und beschleunigt diese Vorgänge. Aber zu welchem Preis? Anfang 2023, also bald nach Markteinführung von ChatGPT im November 2022, hatten US-amerikanische Anwälte mit Präzedenzfällen argumentiert, die es so nie gegeben hatte. KI ist eben eine Wahrscheinlichkeits- und keine Wahrheitsmaschine. Diese Anwälte scheiterten damit kläglich, boten aber Lehrreiches zum Umgang mit KI.

Wie geht's weiter?

Im Frühjahr 2024 wurde vom EU-Parlament ein Gesetz über die künstliche Intelligenz verabschiedet[4]. Demnach sollen Risiken beschränkt werden durch zu konkretisierende Leitlinien einer KI-Verordnung. Im Gegensatz etwa zu den Praktiken in China sollen die Risiken Gefahren reduzieren helfen, welche durch biometrische Fernidentifizierungen oder emotionale Erkennung entstehen. Es geht hierbei um den beruflichen Einsatz durch natürliche und juristische Personen wie Organisationen und Institutionen aller Art, nicht jedoch um Endnutzer.

Der größte Teil des hunderte Seiten umfassenden Gesetzestextes bezieht sich auf Hoch-Risiko-Systeme. Diese sind besonders streng zu regulieren. Im Anhang III werden hierzu acht Bereiche aufgeführt. Auffallend ist, dass zwei (!) dieser acht Bereiche sich auf das Recht beziehen: Strafverfolgung und Rechtspflege.

Fazit: Rechtsbereiche sind gegenwärtig sowohl in der europäischen Gesetzgebung als auch in den nationalen Alltagsanwendungen jene Treiber, die an vorderster Front KI handhabbar und sicherer machen.

12. Juni 2024

[1] Als Beispiel: Durantaye, K. (2024). Nutzung urheberrechtlich geschützter Inhalte zum Training generativer künstlicher Intelligenz – ein Lagebericht. AfP, 55(1), 9-22. https://doi.org/10.9785/afp-2024-550103

[2] Als Beispiel: Schwartmann, R. (2024). Anwendungsszenarien und Rechtsrahmen für den Einsatz künstlicher Intelligenz im Journalismus. AfP, 55(1), 1-9. https://doi.org/10.9785/afp-2024-550102

[3] Als Beispiel: Future Law (12.06.2024). Legal Tech Netzwerk. Künstliche Intelligenz – Einordnung & Evaluierung von Potentialen. https://future-law.eu/ki-projekte-in-legal/

[4] Europäisches Parlament P9_TA(2024)0138. Gesetz über künstliche Intelligenz. Legislative Entschließung des Europäischen Parlaments vom 13. März 2024 zu dem Vorschlag für eine Verordnung des Europäischen Parlaments und des Rates zur Festlegung harmonisierter Vorschriften für künstliche Intelligenz (Gesetz über künstliche Intelligenz) und zur Änderung bestimmter Rechtsakte der Union (COM(2021)0206 – C9-0146/2021 – 2021/0106(COD))

Allmende –
was für ein wundervoll seltsames,
leider selten genutztes Wort

Liebe Blog-Leserinnen,
Sie sind diesem Begriff vermutlich schon begegnet. Man könnte Allmende als mittelalterlichen Rechtstitel auffassen: Alle Mitglieder einer lokalen Gemeinde haben das Recht, bestimmte Wege, Weideland, Waldflächen oder Wasserquellen zu nutzen. Es gehört allen in der Gemeinde, private Eigentümer gibt es nicht. Die Ressourcen werden kollektiv genutzt und überwacht.

Ein in der Neuzeit übersehener Vorteil von Allmende besteht darin, dass sie nachhaltig ist: Viele aufeinanderfolgende Generationen profitieren davon. In der Gegenwart sind kollektive Güter in dieser Art selten geworden, obwohl die im europäischen Mittelalter verbreitet waren.

So wurde noch bis Anfang des 19. Jahrhunderts die Bodenseefischerei durch eine Fischer-Allmende geregelt. Die Regeln der Bewirtschaftung wurden von allen Fischern (und ihren Vertretern) geschaffen. Alle hielten sich daran und hatten dadurch über Jahrhunderte hinweg ein Auskommen. Wenn das Vorkommen einer bestimmten Fischart abnahm, konnte zum Beispiel festgelegt werden, dass die Maschengröße der Netze größer sein musste.

Junge Fische entkamen so. Nach zwei oder drei mageren Jahren stellte sich in der Regel wieder eine zufriedenstellende Fangmenge für alle Fischer ein.[1] Wenn Menschen kooperieren, führt das zum langfristigen Erfolg.

Demgegenüber steht die "moderne" Auffassung, dass es bei Allmendegütern zwangsläufig zur Tragödie (*The tragedy of the Commons* nach Garrett Hardin) **muss. Verständlich, wenn man den Menschen als konsequent handelnden homo oeconomicus ansieht:** Ungezügelte individuelle Freiheit bei Allmendegütern führt zur Übernutzung und letztendlich zum Ruin für alle.

Wie ist das mit Wissenschaft?

Wissenschaft funktioniert wie eine Wissensallmende. Denn sie gedeiht aufgrund ihrer Offenlegung von Daten und Verfahren. Das sichert Korrekturen von Fehlern oder Verzerrungen und ermöglicht ein weitgehend gesichertes Wissen in einer modernen Informationsgesellschaft.

Richtig, dieses Ideal wird nicht immer erreicht – Forschungszugänge werden reglementiert, nichtgenehme Studienergebnisse verschwinden in Schreibtischen etc. Doch ist ebenso klar, eine systematisch versteckte Wissensproduktion oder eine nur privat vorteilhafte Nutzung beeinträchtigen negativ das Wissen und zerstören die Nachhaltigkeit.

Wie ist das mit KI?

Das Verhältnis von privater zu öffentlicher Finanzierung spiegelt große Ungleichheit wider. Es liegt irgendwo zwischen 10:1 oder 100:1. Regelungen werden daher von den Tech-Giganten als hemmend empfunden und umgangen. Oder aufgrund ihrer geballten Marktmacht vorerst überhaupt ignoriert.

Exemplarisch zeigt dies ein aktuelles Beispiel. Es betrifft eine Pseudo-Veröffentlichung in der renommierten Wissenschaftszeitschrift *Nature*. AlphaFold 3, eine KI-Applikation von Google Deep Mind, prognostiziert Interaktionen zwischen Proteinen und anderen Molekülen.[2] AlphaFold 3 hat das Potenzial, die Entwicklung von Medikamenten zu revolutionieren. Unter den angegebenen Schlagwörtern ist auch Drug Discovery zu finden. Dies gibt Auskunft über das im Hintergrund anvisierte ökonomische Ziel.

Dennoch ist dies großartig – bessere Medikamente sind in Sicht! Leider gibt es Einschränkungen, die Sorgen bereiten:

1. Das Modell und der Programmcode werden geheim gehalten. Wie diese Applikation arbeitet, lässt sich nicht überprüfen, wie fundiert die Ergebnisse sind. Fundamentale wissenschaftliche Gütekriterien wie Offenheit und Überprüfbarkeit werden nicht beachtet.
2. Das ethisch Bedenkliche ist außerdem, dass dieses KI-Modell mit Daten und Verfahren aus öffentlich zugänglichen Datenbanken trainiert wurde. Also trainiert mit Daten und Verfahren, die großteils im Rahmen öffentlich finanzierter Forschung entwickelt wurden.
3. Ein offener Brief, der bereits von mehreren hundert Forscherinnen unterschrieben wurde, prangert dies an.

Fazit: Die Wissensallmende wird privat übernutzt. Wir blicken in eine Arena, in der ein Kampf um Marktanteile und Gewinnchancen zwischen wenigen Groß-Gladiatoren stattfindet, sowohl zum Vorteil als auch zum Nachteil für uns alle.

PS: Noch etwas kommt einem in den Sinn: Das Mittelalter war aus Sicht der Nachhaltigkeit moderner als die Gegenwart. In der Zukunft wird eine kollektiv verantwortliche Ressourcennutzung von Boden, Wasser, Luft und Information (!) aus schierer Not heraus wieder in das Zentrum von regionalen, nationalen und übernationalen Institutionen rücken.

8. Juni 2024

[1] Kehnel, A. (2021). *Wir konnten auch anders. Eine kurze Geschichte der Nachhaltigkeit.* München: Karl Blessing. Kapitel Commons und die Kunst, Externalitäten zu internalisieren. S. 63-95.

[2] Abramson, J./Adler, J./Dunger, J. et al. (2024). *Accurate structure prediction of biomolecular interactions with AlphaFold3*. Nature. https://doi.org/10.1038/s41586-024-07487-w.

Wasser als gemeinsames Gut
Nächtliche Schiffsreise in der Wachau: Sternbild Grosser Wagen und verglühender Meteor

Das überraschende Prognose-Tool

Was wir seit langem über generative KI wissen: Halluzinationen kommen vor, Unscharfes und Durchschnittliches kennzeichnen nicht selten die KI-Antworten. Davor wird überall gewarnt, Checks und Korrekturen werden empfohlen.

Und nun das: Eine umfassende Studie an der Universität Chicago belegt, dass Umsatzprognosen von ChatGPT-4 (von OpenAI) oder Gemini Pro (von Google) nicht nur besser sind als speziell entwickelte mathematische Finanzmodelle, sondern auch besser als die Prognosen erfahrener Finanzanalysten[1].

Diese Studie wirkt überlegt und gewissenhaft durchgeführt. Die Ergebnisse basieren auf einem Datenpool von mehr als 15.000 Unternehmen mit 150.000 Jahresumsätzen im Zeitraum 1968–2021. Sowohl die Unternehmen als auch die Jahre sind anonymisiert, sodass die Chatbots keine Informationen aus umgebenden Texten (Geschäftsberichte etc.) sammeln können. Die Eingaben bestehen daher nur aus Zahlen.

Das Interessanteste ist der gekonnte Umgang mit dem Prompt. Eine einfache Frage nach der Prognose zukünftiger Umsätze zeigt für GPT-4 ein geringfügig schlechteres Ergebnis als das, was ein erfahrener Finanzanalyst prognostizieren würde. Nutzt man jedoch eine Angabe, die im Prompt die Überlegungen und Gedankenketten eines menschlichen Finanzanalysten schrittweise vorgibt, so werden die KI-Prognosen signifikant besser als die Prognosen der Finanzanalysten.

Das bedeutet, Large Language Modelle (LLMs) können mit der Abfolge spezifischer menschlicher Überlegungen Zahlenreihen besser fortschreiben als der Mensch allein oder GPT-4 allein. Menschen denken aufgrund ihrer kaum bewussten Vorannahmen und Vorurteile ineffizient und verzerrt. Chatbots allein sind auch eher schwach.

Jedoch in Kombination mit menschlichen, algorithmusähnlichen Überlegungen, die groben Algorithmen ähneln, sind sie kaum zu schlagen. Fazit: Die Kombination aus KI und Mensch gewinnt!

Trotzdem muss theoretisch und praktisch durchdacht werden, was LLMs zu leisten imstande sind. Wir wissen offensichtlich noch immer nicht, welches Potenzial in ihnen steckt. Jeder Chatbot stellt eine Blackbox dar und überrascht nach wie vor – vielleicht sogar die Schöpfer selbst.

Daraus ist zu folgern, dass LLMs aus den Soft-Science-Bereichen in die Felder der Hard-Science hinüberreichen. Bei entsprechender Handhabung in Kombination mit dem Menschen als Kontext-Experten stellen die Large Language Modelle eine pragmatische Verbesserung dar für:

1. das interne Rechnungswesen/Controlling in Unternehmen,
2. die Marktbeobachtung in der Betriebswirtschaftslehre und viele Kennzahlen (Key Performance Indicators) für das strategische Management,
3. die prognostische Entwicklung von Aktien, Zinsen, Wechselkursen, BIP-Wachstum oder -Schrumpfung, Arbeitslosenzahlen oder Inflation in der Volkswirtschaftslehre,
4. kurzfristige Wettervorhersagen und langfristige Klimawandelprognosen, mögliche Systemveränderungen, Annäherungen an Kipppunkte,
5. Vorhersagen von Erdbeben oder Vulkanausbrüchen in den Geowissenschaften,
6. frühzeitiges Erkennen von Krankheitsausbrüchen und Epidemien in den Gesundheitswissenschaften,
7. die Bevölkerungsentwicklung samt Hinweisen auf die Planung für Infrastruktur und entsprechenden Anforderungen für Gesundheits- und Bildungssysteme,
8. Prognosen von Ernten, Schädlingsbefall oder die Entwicklung der Bodenbeschaffenheit für Pflanzen in den Agrarwissenschaften.

Daraus wiederum ist zu schließen, dass die Chatbots direkt die Entscheidungen von Managern, Administratoren oder Politikern beeinflussen können und dies auch werden. Diese KI-basierten Entscheidungen werden uns über kurz oder lang alle betreffen.

6. Juni 2024

[1] Kim, A., Muhn, M., Nikolaev, V. (Mai 2024). *Financial Statement Analysis with Large Language Models.* Working Paper No. 2024-65. Becker Friedman Institute, the University of Chicago. https://bfi.uchicago.edu/wp-content/uploads/2024/05/BFI_WP_2024-65.pdf (3. Juni 2024)

Wieviele Kunden werden heute kommen?
Venedig: Campo Santa Margherita

Kosmische Kollision

Richtlinien zur

- Erstellung schriftlicher Seminararbeiten
- Bachelorarbeiten

Richtlinien zum

- Verfassen von Masterthesen
- Dissertationen

Richtlinien, Richtlinien – Universitäten, Hochschulen und akademische Lehrgänge quellen über von detaillierten Regeln für das Verfassen von Arbeiten. Sie bestimmen nicht zuletzt die Strukturen des Forschens. Wie etwa das Recherchierte darzustellen ist, wann zu interpretieren, schlusszufolgern und zu diskutieren ist.

Was im ersten Moment wie Quälerei von Studierenden ausschaut, ist – abgesehen von überformalisierten Vorgaben – eine Hilfe, das seriöse und nachvollziehbare Erkunden der Welt zu erlernen. Es geht darum, das bestätigte oder durch die Forschung zusätzlich gewonnene Wissen sachgerecht darzustellen und der Scientific Community effizient zu präsentieren.

Sowohl diese Vorgaben insgesamt als auch die Zitierrichtlinien haben sich im Verlauf von Wissenschaftler-Generationen im jeweiligen Fach entwickelt. Und zwar inhaltlich fundiert begründet:

- Es gibt Gründe, wenn Historiker eher Fußnotenzitierungen nutzen. So werden auf derselben Seite ausführliche Informationen zu den jeweils verwendeten Quellen bereitgestellt. Und Quellen sind in den Geschichtswissenschaften das Um und Auf.
- Dagegen neigen Sozialwissenschaftler eher zu computergerechten Kurz- und Positionszitaten im laufenden Text, wie etwa den Harvard-Zitierregeln. Weil hier die Argumentationslinie als

im laufenden Text, wie etwa den Harvard Zitierregeln. Weil hier die Argumentationslinie als zentral angesehen wird.

Diese – allerdings nur auf den ersten Blick rigiden – Richtlinien sind Ausdruck des sukzessiv angesammelten Know-hows zahlreicher Wissenschaftsgenerationen.

In diese strukturierte Welt bricht eine anders funktionierende ein: Die Welt der Tech-Giganten mit KI als brachialem Werkzeug. Aber den großen Tech-Unternehmen ist gewachsenes Know-how verdächtig. Meta, Microsoft, OpenAI, Google setzen auf „speed kills“: Urheberrechte werden ignoriert, ungeniert werden Texte, Bilder oder Stimmen geraubt.

Falls ein Aufschrei unüberhörbar wird und US- oder EU-Behörden mit Regulierungen drohen, nehmen die Tech-Unternehmen den einen oder anderen Missbrauch nach zwei, drei Jahren mit Bedauern zurück. Man habe es nicht so gemeint. Im Gegenteil, sie hätten doch immer hohe ethische Ansprüche gehabt und wollten „der Menschheit dienen“.

Was die Tech-Giganten in den letzten zwanzig Jahren mit der Innovation von sozialen Medien bewirkt oder zumindest verstärkt haben: Das soziale Auseinanderbrechen in polarisierte und zerrissene Gesellschaften. Aus einem ursprünglich für militärische Resilienz gegründeten Internet wollte man in den 1990er Jahren eine gemeinnützige Infrastruktur schaffen. Diese sollte Kommunikation für alle verbürgen und damit freien Zugang zu Wissen ermöglichen.

Heute funktioniert das Internet leider anders. Man hat den Eindruck, stündlich über den Tisch gezogen zu werden durch maßgeschneiderte Angebote, welche die eigenen Schwächen gnadenlos ausnützen.

Was bringt nun die KI?

Die KI-Welt wird von denselben privaten Technik-Giganten vorangetrieben. Ethische Überlegungen spielen keine Rolle – Macht, Einfluss und Profit müssen am Ende des Tages exzeptionell sein.

Diese weitgehend ehrlose Welt der Tech-Giganten bricht jetzt in die respektable Welt der wissensgenerierenden Institutionen ein. Aus wissenschaftlicher Sicht erleben wir einen derart fundamentale Kollision, wie sie sich in unserem Sonnensystem vor viereinhalb Milliarden Jahren abgespielt hatte:

> Die Proto-Erde soll nach der Theia-Hypothese mit einem etwas kleineren Planeten (eben namens Theia) kollidiert sein. Die zahllosen heißen Gesteinstrümmer dieses katastrophalen Ereignisses zogen in einer Bahn rund um die Erde, kondensierten und bildeten den Mond. Es gibt Gutes im Schlechten: Der Mond übte über die Jahrmilliarden einen stabilisierenden Einfluss auf die Bewegung der Erde aus, sodass sich Leben bilden und entwickeln konnte.

Gehen wir ähnlichen Zeiten entgegen? Zuerst die Kollision der KI-Welt mit der wissenschaftlichen Welt samt Zerstörung zentraler Forschungstraditionen – aber könnte später eine Stabilisierung einsetzen?

Wenn ja, wird die Forschungs- und Wissenswelt in vielen Teilen anders funktionieren als bisher. Wir gehen spannenden Zeiten entgegen.

27. Mai 2024

Graue Literatur – KI besetzt den blinden akademischen Fleck

Interviews von Wissenschaftlern in Tageszeitungen, Statistiken, Konferenz- oder Geschäftsberichte, Blogs, Newsletter, Symposien, White Papers, persönliche Briefe, Präsentationen, Pressemitteilungen, Patente, technische Dokumentationen, unveröffentlichte Werke –

was ist allen diesen Formen gemein? Sie sind oder enthalten Daten, das Gold der Gegenwart. Häufig allerdings auch zweifelhafte Daten. Daher ist in wissenschaftlichen Arbeiten zurecht Skepsis angebracht, diese eins-zu-eins als seriöse Quelle zu verwenden. Doch gleich ein Verbot auszusprechen, wie es einige akademische Bereiche tun, hieße, das Kind mit dem Bade auszuschütten.

So kann gerade in Interviews von Wissenschaftlern Neues und noch nicht Publiziertes enthalten sein. Es können übergreifende Konzepte in verständlichen Worten aufgezeigt werden. Gerade übergreifende Konzepte sind in der Regel paradigmatisch in Fachveröffentlichungen nicht vorgesehen.

Graue Literatur wird traditionell definiert als nicht kommerziell durch Verlage vertriebene und nicht im Buchhandel erhältliche Literatur. Wissenschaft ignoriert graue Literatur. Sie ist im akademischen Bereich global nur in Spuren vorhanden[1], obwohl diese Literaturform in der Regel detailliert, flexibel, aktuell und kostengünstig herzustellen ist. Häufig bietet sie eine deutlich übergreifendere Sicht als andere Daten- und Textarten. Überdies verfügen Institutionen und Verlage nicht über den regulierenden Zugriff, höchstens können sie die Verbreitung unterstützen[2].

In der Wirtschaft scheinen Unternehmen diese Quellen nun zu entdecken. Denn in diesen grauen Daten können Erkenntnis-Perlen zu finden sein. Die darin enthaltenen Informationen könn(t)en die Erkenntnisse aus den üblichen betriebswirtschaftlich monolithischen Datenströmen der Unternehmen erweitern und so verbessern.

Graue Literatur stellt ein Netz von vielen zweifelhaften Daten dar. Aber in diesem Netz stecken interessante, die Norm bestätigende oder von der Norm abweichende Informationen. Klar, es ist nicht so einfach, die Perlen in dem Informationshaufen der grauen Literatur zu finden. Oder sinnvolle Erkenntnisse aus einer Gesamtsicht dieses Netzes abzuleiten.

Die Herausforderungen sind beachtlich, denn diese graue Literatur gehört selbstverständlich aufgrund von Prüf- und Selektionsprozessen aufbereitet, will man sie denn zu Erweiterung eines bewährten Wissens nutzen.

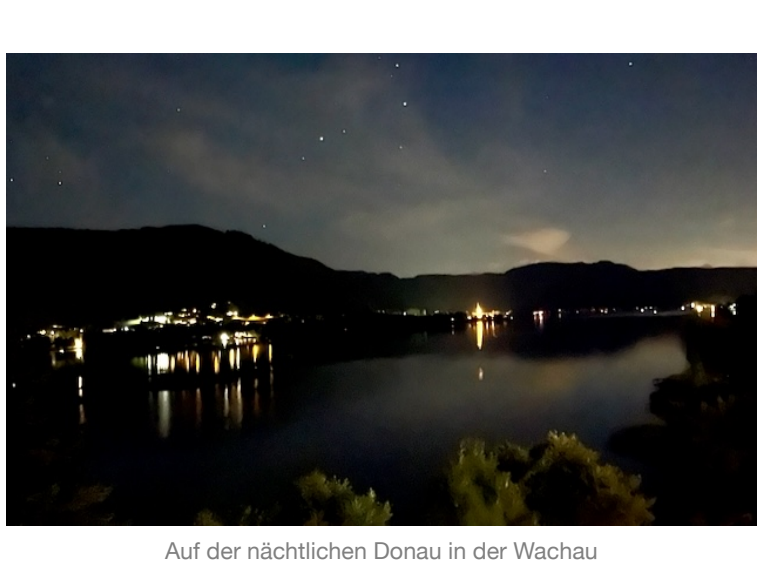
Nun aber kommt die genKI. Man schätzt, dass die KI innerhalb der nächsten zwei bis drei Jahre das komplette Internet abgegrast und aufgenommen haben wird. Also auch die nichtkonventionelle graue Literatur.

Sie wird daher in Kürze in den KI-Speichern vorhanden sein und in die Algorithmen und Antworten der KI einfließen. Dadurch wird die Qualität der KI-Antworten vielleicht mal besser, wahrscheinlich aber eher schlechter ausfallen. Graue Literatur wurde eben nicht fachgerecht aufbereitet.

20. Mai 2023

[1] Goedegebuure, A./Kumura, I./Driessen, S. et al. (2023). *Data Mesh: a Systematic Gray Literature Review.* J. ACM, Vol. 37, No. 4, Article 111. Publication date: August 2023.

[2] Dubal M. (2018). *A Study of Grey Literature.* S. 265-269. In: ASM"s International E-Journal on "Ongoing Research in Management & IT", 12th - 13th January 2018.



Auf der nächtlichen Donau in der Wachau

KI-Erfahrungen wertvoller als Bachelor-Abschluss?

Jeder zweite (49%) einer Befragung von 800 amerikanischen Führungskräften und Personalverantwortlichen gab im April 2024 an, ChatGPT-Kenntnisse und Erfahrungen im KI-Umgang seien wichtiger als ein College-Abschluss[1] [2].

Fabian Stephany, Arbeitsmarktforscher an der Universität Oxford, erklärt: „Eine umfassende Analyse von Jobofferten und Gehaltsangaben aus Großbritannien hat ergeben, dass der Gehaltsaufschlag für KI-Skills rund 16 Prozent beträgt. Das ist fünfmal mehr als der durchschnittliche Wert aller anderen Fähigkeiten. […] Tatsächlich ist bei KI-Jobs kaum noch ein Gehaltsvorteil für Akademiker zu erkennen.“[3]

Fazit: Die theoretische Fundierung von Fähigkeiten und Wissen wird weniger wichtig. Aufgrund der rasanten KI-Entwicklung rückt das Beherrschen von Szenarien der realen Welt in den Vordergrund. Das selbständige Lernen und fortwährende Entwickeln neuer Fertigkeiten ist ausschlaggebend. Das wiederum verstärkt Kreativität und Produktivität.

—> Brennt Feuer am Dach unserer Unis?

Ja und Nein!

Darum Nein:

Ein Knackpunkt liegt in den genannten Begriffen *KI-Kenntnisse* und *Erfahrungen*. Wer nie erfahren hat,

- was seriöse Quellen sind,
- wie nützlich Theorien und Modelle sein können,
- wie wichtig aber auch deren Grenzen und Geltungsbereiche sind,
- wie Fallgruben in (nicht präzisen) Definitionen von rasch formulierten Begriffen liegen,

kann den Output von KI nicht wirklich einschätzen und beurteilen. Er wird den Output für bare Münze halten. Auch wenn der (tendenziell denkbefreite?) Anwender gelernt hat, irgendwie passende KI-Prompts zu verfassen.

Insofern gleicht die Begeisterung von Führungskräften und Personalverantwortlichen für KI den Modewellen im Bereich betriebswirtschaftlicher oder technologischer Innovationen. Jahr für Jahr wird eine neue und quietschende Marketing- oder Technologie-Sau durch das globale Dorf getrieben: Big Data, Digitale Transformation, Customer Relationship Management, Nachhaltigkeit, Corporate Social Management, Agiles Projektmanagement, Internet der Dinge, Blockchain-Technologie, Plattformökonomie .... Nicht, dass die Inhalte dieser gehypten Konzepte falsch wären, nur lehrt die Erfahrung, dass sie oft eine ganze Generation und länger benötigen, bis sie greifen.

Wenn man zum Beispiel liest, wofür gemäß der erwähnten Befragung die genKI in den US-amerikanischen Unternehmen eingesetzt werden soll, dann ordnet sich einiges deutlich kleiner ein: Dateneingabe, Schreiben von e-Mails, Aufgabenlisten, Recherchen, Kundensupport.

Darum Ja:

Die Ziele der tertiären Bildung changierten schon immer zwischen (allgemeiner und erkenntniserweiternder) Bildung und (beruflicher) Ausbildung. Die rasante Entwicklung der genKI und die breite Übernahme durch Unternehmen, Organisationen und Gesellschaftsgruppen aller Art werden die Waage in Richtung Ausbildung drücken.

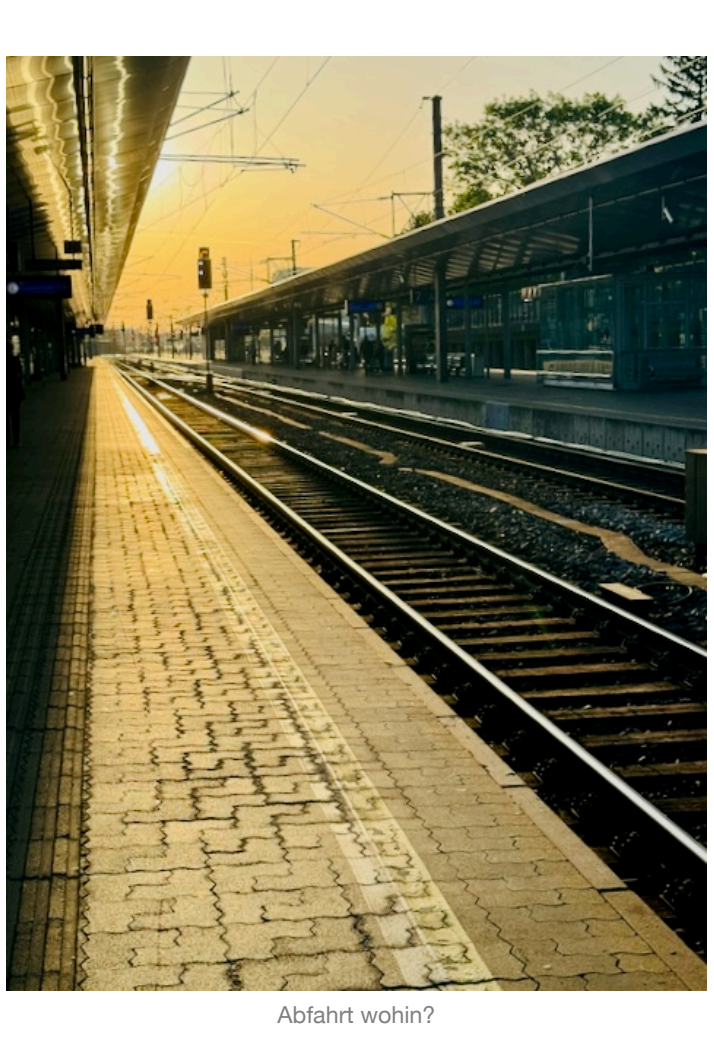
Abzuwarten bei der KI-Thematik nach dem Motto „Schau ma moi, dann seng ma scho" wird sowohl den Unis als auch den Studierenden schaden.

18. Mai 2024

[1] intelligent.com, learning how to learn (10.04.2024). *Half of business leaders say chatgpt experiences is more valuable than a college degree.* https://www.intelligent.com/half-of-business-leaders-say-chatgpt-experience-is-more-valuable-than-a-college-degree/

[2] Dies entspricht im europäischen Hochschulsystem einem Bachelor-Abschluss.

[3] Von Elm, K. (27.02.2024). Stephany im Interview: *KI-Skills bringen so viel wie ein Doktortitel.* Handelsblatt. https://www.handelsblatt.com/unternehmen/interview-mit-fabian-stephany-ki-skills-bringen-so-viel-wie-ein-doktortitel-03/100012315.html



Abfahrt wohin?

NICHT von der KI zu erfragen (B):

3) Zitieren und Zitate suchen

Angenommen, du hast zu offenen Fragen spontane Aussagen von 700 Personen. Diese Aussagen werden in der weiteren Analyse in Kategorien unterteilt. Wenn du nun nach bestimmten Aussagen suchst, wirst du ein blaues Wunder erleben. Dies gilt auch dann, wenn du ausdrücklich darauf hinweist, dass du Originalaussagen präsentiert haben willst.

- Vergleichsweise harmlos ist das erste Beispiel: Aus der Originalaussage "1 Kind" wird "Ein Kind".
- Schon deutlich abweichender ist das zweite Beispiel: Aus der Originalaussage "Auch die Lehrkräfte waren" wird "Die Lehrer:innen waren". Überdies fehlt das "Auch", wodurch es zu Sinnveränderungen kommt. Ebenso verändernt GPT die unzulässigerweise die Aussage, indem er meint, gendern zu müssen.

Ein schwerwiegende Fehler entsteht aber aus der Tatsache, dass GPT4 – angabenwidrig – Zusammenfassungen mehrerer Originalaussagen macht und als eine (!) Originalaussage präsentiert.
Fazit: Vieles sieht man nicht sofort (daher alles, aber auch wirklich alles checken!), manches ist als Zusammenfassung sofort zu erkennen. Wenn zum Beispiel nach der Situation in der einzelnen jeweiligen Familie während der Corona-Pandemie gefragt wurde, aber als "Originalaussage" präsentiert wird: "Einige Familien erlebten […]", dann weißt du, dass du diese Aussage im eingegebenen Text nicht finden wirst. Sie ist schlicht falsch.

Worauf die nächsten beiden **Don't do that** logisch folgen:

4) Vertraue nicht darauf, dass die KI etwas ganz spezifisch Gesuchtes findet.

Das Gesuchte ist möglicherweise NICHT im Text vorhanden --> simple Texteditoren oder Spreadsheets wie Excel können das besser.

5) Vertraue nicht darauf, dass die KI etwas ganz spezifisch Gesuchtes im eingegebenen Text tatsächlich nicht vorhanden sein.

Das Gesuchte kann DENNOCH im Text vorhanden sein --> simple Texteditoren oder Spreadsheets wie Excel können das besser.

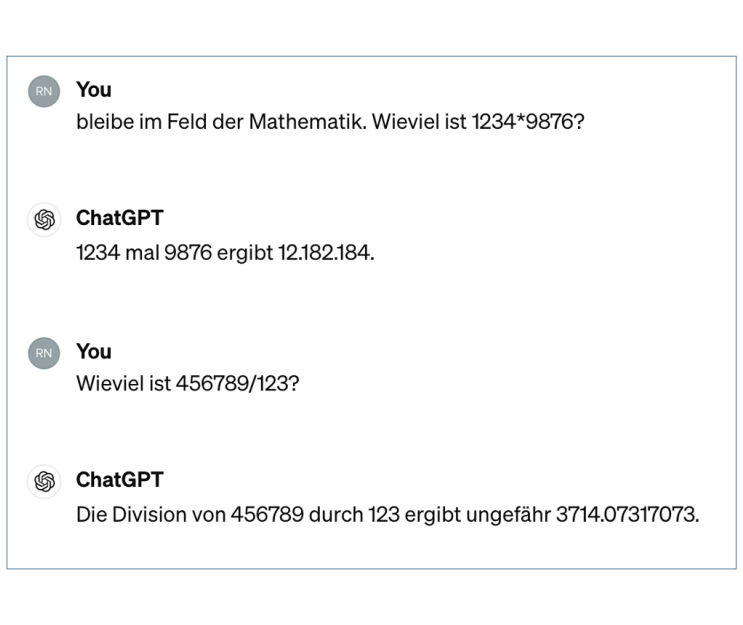
15. Mai 2024

Die Fehler-Bugs lauern überall

NICHT von der KI zu erfragen (A):

1) Exakte Zahlen errechnen

Zwei Beispiele zum Multiplizieren und Dividieren ergaben im ersten Versuch folgende Ergebnisse:



You
bleibe im Feld der Mathematik. Wieviel ist 1234*9876?

ChatGPT
1234 mal 9876 ergibt 12.182.184.

You
Wieviel ist 456789/123?

ChatGPT
Die Division von 456789 durch 123 ergibt ungefähr 3714.07317073.

Die Multiplikation ist eindeutig falsch. Das richtige Ergebnis lautet 12.186.984, also um fast 5.000 daneben.

Dagegen führt die Division nicht sofort in tiefen Abgrund. Das richtige Ergebnis lautet 3.713,73. Es weicht also hinsichtlich von zwei Nachkommastellen um 34 Hunderstel vom Ergebnis ab.

GPT4 führt hier einen Eiertanz auf: Erstens schreibt es richtigerweise von "ungefähr", aber bietet dann eine Zahl mit acht Nachkommastellen an. Man nennt das in der Statistik *false precision* oder *fake precision*. Eine bestimmte Genauigkeit vortäuschende Präzision führt in die Irre.

Dies kommt nicht selten in wissenschaftlichen Veröffentlichungen vor. Wenn bei 70 erhobenen Datensätzen beispielsweise für ein deskriptives Ergebnis 23,6% angeben wird. Eine solche standardisierende Promille-Angabe (Pro-Mille, also je Tausend) wäre eigentlich erst im vierstelligen Bereich von Datensätzen berechtigt.

Fazit: Simple Taschenrechner können das besser.

2) Frequenzen erheben

Wie zum Beispiel die Anzahl des Wortes „und“ in einem Text zählen. Ein eingegebener deutscher Beispieltext von insgesamt 3.236 Wörtern enthält das Wort „und“ 116 Mal.
GPT4 meint dagegen, es seien 122.

Fazit: Simple Texteditoren können das besser.

12. Mai 2024

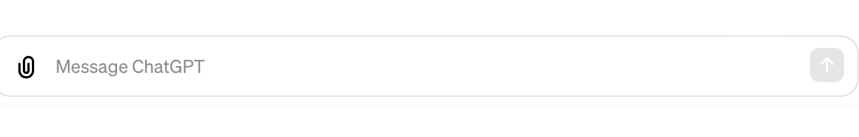
Ein Befehl, der etwas ans Licht bringt

Das Hauptwerkzeug, um die KI zu bedienen, ist die einzugebende Anordnung, das Prompt. Das Wort kommt aus dem Lateinischen und bedeutet "hervorholen, ans Licht bringen, zeigen", aber auch "erzählen, eine Rede halten"[1].

Treffender könnte es für das, was die KI dann produziert, nicht formuliert sein. Die KI bringt durch das Prompt etwas hervor, indem sie erzählt.

Das hört sich vorerst simpel an, du brauchst ja nur was befehlen. Ist aber nicht so simpel, die Anordnung stellt sich gleich von Anfang an als Herausforderung dar: Du schaust auf eine leere Zeile, die offensichtlich etwas erwartet. Und zwar was Gescheites und Sinnvolles!

Psychologisch gesehen eine schwierige Situation, die Druck aufbaut. So leer, aber auch erwartungsvoll blickt das Prompt auf dich:



Andererseits gleicht der Prompt einer Peitsche des Dompteurs, wenn er in der Zirkusmanege Tiger durch brennende Reifen springen lässt. Das kann klappen, aber auch schief gehen. Der Anwender merkt es wahrscheinlich nicht, dass er einen Blödsinn serviert kriegt.

Da die vorhergehenden Beiträge länger waren, werde ich mich diesmal kurz halten. Drei zentrale Dimensionen kennzeichnen ein geglücktes Prompt: die Rolle, der Kontext, die Anweisung.

- Zur Rolle, welche die genKI übernehmen soll: Sie kann künstlerische Aufgaben übernehmen ("Erzähl' mir über ..") oder den Buddy/Kumpel spielen, motivieren, zusammenfassen, Alternativen aufzeigen, als Trainer oder als Sparringpartner innerhalb eines neosokratischen Dialogs etc. fungieren.
- Zum Kontext: Der Kontext (wortwörtlich: "webt zusammen"[2]) verknüpft die Rolle mit der relevanten Umgebung. Da dieser Blog fokussiert ist auf genKI in Zusammenhang mit tertiärer Bildung (Universitäten, Hochschulen, Kollegs,...), ist es ratsam, die jeweilige wissenschaftliche Disziplin anzugeben.
  Je konkreter der Kontext, umso zutreffender die Antwort und umso geringer ist der anschließende Check- und Korrekturaufwand.

  Klar, dass auch fachübergreifend (inter- oder transdisziplinär) der Kontext angegeben werden kann. So werden Wege aufgezeigt, an die man nicht gedacht hatte. Die Chance, erheblichen Quatsch zu erhalten, der weit jenseits von Wissenschaft liegt, sozusagen "pseudo-scientific", ist dennoch hoch.
- Zur Anweisung: Die genKI soll was tun. Ein Fokus liegt auf den Anweisungen mit Hilfe der verwendeten Zeitwörter, Verben oder "Tun-Wörter", wie diese Wortform in der Volks- und Grundschule richtigerweise genannt wird. Hauptwörter dagegen sind wie riesige Steine, die in der Gegend herumliegen. Erst Tun-Wörter erzählen die Geschichte und bringen Leben in die Szene --> also konkret die KI generieren lassen. Die Anweisung kann aus mehreren Sätzen bestehen.

Viel Mut im Umgang mit der KI – schwingen Sie die Peitsche!

5. Mai 2024

[1] Langenscheidt (2008). *Großes Schulwörterbuch Lateinisch-Deutsch.* Berlin: Langenscheidt, S. 647

[2] ebd., S. 192

GPT4: Auf den theoretisch-paradigmatischen Zahn gefühlt

Schon im vorherigen Beitrag hatte ich ein Thema abgefragt, das innere Widerhaken hat:

Public Health – Prävention durch individuelle Präzisionsmedizin.

Diesmal wollte ich es genau wissen, ich bin nämlich der Meinung, dass die Verbindung der zwei angeführten "Sub-"Themen im gemeinsamen Kontext miteinander nicht vereinbar sind. Keinesfalls baut eines auf dem anderen auf. Auch ergänzen sie einander nicht. Wird ein Ansatz forciert, kommt der andere unter die Räder

Auch klar ist, dass beide Ansätze in der gesellschaftlichen Realität schon nebeneinander vorkommen, wenn auch die Präzisionsmedizin noch als kleine Schwester fungiert. In der Folge steht das vollständige Prompt mit GPT4 in der Rolle als Gutachter:

(Prompt): Du hast die Rolle eines wissenschaftlichen Gutachters. Der Kontext ist wiederum mein Studium in Gesundheitswissenschaften, im Speziellen das Thema "Public Health – Prävention durch individuelle Präzisionsmedizin".

In diesem Thema sind zwei Subthemen "Public Health" und "Prävention durch individuelle Präzisionsmedizin" miteinander verknüpft. Sind diese zwei Subthemen von ihren Paradigmen her miteinander kompatibel? Antworte mit Ja oder Nein und gib Begründungen dafür an.

Zum inhaltlichen Verdeutlichung dieses Prompts: Ich stelle hierin vier theoretisch-paradigmatisch relevante Vorgaben auf: 1. GPT4 soll in der Rolle eines wissenschaftlichen Gutachters, also kritisch sein. 2. Es geht um Gesundheits-Wissenschaften. 3. Es geht um „individuelle“ Präzisionsmedizin (mit einer Hinweisverdoppelung durch „individuell“, obwohl das Individuelle bereits im Begriff Präzisionsmedizin impliziert ist) und das alles entscheidende Umfassende: 4. Ist das Gesamtthema von ihren Paradigmen her kompatibel?.

Urteil von GPT4: Ja, die beiden Subthemen „Public Health“ und „Prävention durch individuelle Präzisionsmedizin“ sind von ihren Paradigmen her miteinander kompatibel und können sich sogar ergänzen.

Ich erspare den geneigten Blog-Lesern die Details der ausführlichen Begründungen. GPT4 schreibt von einer Erweiterung des Public Health-Ansatzes, die zu einer größeren Effektivität der öffentlichen Gesundheitsmaßnahmen führe.

Es werden zielgerichtete Präventionsstrategien angeführt, die auf spezifische Risikogruppen zugeschnitten sind, was zu geringeren Kosten führen soll. Ebenso sollen Integrations- und Synergie-Effekte durch Nutzung von anfallenden Daten entstehen. Und – kaum zu glauben: Diese thematische Verbindung sei ethisch und sozial kompatibel.

Die Begründungen lesen sich abgeklärt und flüssig, wie frisch aus dem Lehrbuch für Gesundheitswissenschften eines Bachelorstudiums abgeschrieben. Sie wirken überzeugend.
Auf den ersten Blick.

Der zweite Blick zeigt, dass die Begründungen nicht stichhältig sind. Die Paradigmen, die Ansätze und entsprechenden Definitionen, Vorgehensweisen und Folgerungen sind nicht nur grundverschieden, sondern einander entgegengesetzt.

Hierbei geht es keinesfalls um begriffliche Spitzfindigkeiten. Wissenschaftliche Aussagenverlieren ihre Geltung, wenn sie in andere Rahmenbedingungen und in andere Fachgebiete unverändert übertragen werden.

Die chaotischen Diskurse während der Corona-Pandemie samt den langfristigen politischen und sozialpsychologischen Folgen im deutschsprachigen Raum zeigen dies deutlich. So hatten Schwurbler, Verschwörungstheoretiker, Impfgegner und Wissenschaftsleugner viel Auftrieb erhalten.

Beginnen wir mit den Definitionen. Eine verbreitete Definition von Public Health lautet: „Unter Public Health verstehen wir eine von der Gesellschaft organisierte, gemeinsame Anstrengung, mit Ziel der Erhaltung und Förderung der Gesundheit der gesamten Bevölkerung oder Teilen der Bevölkerung, […]“[1].

Dieser soziale und gesellschaftlich breite Top down-Ansatz kommt im deutschen Begriff „öffentliche Gesundheitspflege“ für Public Health noch deutlicher zum Ausdruck.

Dagegen wird Präzisionsmedizin definiert als „[…] an innovative approach that uses information about an individual’s genomic, environmental and lifestyle information to guide decisions related to their medical management. The goal of precision medicine is to provide more a precise approach for the prevention, diagnosis and treatment of disease.“[2]

Kurz, es geht um das spezifische medizinische Management von Individuen.

Dazu der zusammenfassende letzte Satz von GPT4: Insgesamt sind die Paradigmen von Public Health und der Prävention durch individuelle Präzisionsmedizin kompatibel und können durch ihre Kombination zu einer umfassenderen und effektiveren Gesundheitsversorgung und -prävention führen.

Nicht doch, eine Gegenüberstellung von Ansätzen, entsprechenden Handlungen und Folgerungen zeigt das gegenteilige Bild:

1. Public Health nützt gemäß ihrem seit rund 150 Jahren entwickelten Konzept universelle Ansätze wie Aufklärung und Impfungen. Public Health beruht auf kostengünstigen Angeboten oder sanften Interventionen, die auf große Segmente der Bevölkerung, wie Kinder, Schwangere oder Alte, gerichtet sind.
2. Präzisionsmedizin hingegen fokussiert mittels spezifischer Methoden auf eine individuelle Person. Präzisionsmedizin baut auf einzigartigen biologischen Faktoren wie Genen, Molekülexpression oder Mikrobiom auf. Sie ist daher höchst ressourcenintensiv und teuer.

1. Public Health beruht auf knappen Budgets öffentlicher Institutionen, die auf traditionelle Techniken und Anwendungen ausgerichtet sind.
2. Präzisionsmedizin erfordert neueste Technologien und spezifische Fachkenntnisse, was die Kosten für das Gesundheitssystem erhöht und potenziell Ressourcen von breit angelegten Public Health-Programmen abzieht.

1. Public Health nutzt allgemeine Daten der Population, die öffentlich zugänglich sind.
2. Präzisionsmedizin nutzt private biologische Daten, die vertraulich und streng zu schützen sind: „[it] takes advantage of large data sets of individuals such as their genome or their entire electronic health record to tailor their healthcare to their unique attributes“[3].

1. Die Prinzipien von Public Health sind auf Gleichheit ausgerichtet. In vielen Ländern wird die Ungleichheit in der Gesundheitsversorgung größer.
2. Gleichheit ist kein relevantes Prinzip der Präzisionsmedizin. Im Gegenteil, der in der Welt nur einmal vorkommende Mensch ist als In-dividuum (wörtlich 'un-teilbar': ein Teilen/Sharing ins genetische bzw. molekulare Unten oder ins soziale Oben ist nicht enthalten) anzusehen: „It is common sense that no two individuals are the same, and so they should not get the same healthcare. Precision healthcare embodies that simple idea.“[3]

--> All dies wird unweigerlich zu anhaltenden monetären, gesellschaftlichen und politischen Konflikten zwischen beiden Ansätzen führen.

Schon im Jahr 2012, kurz nach dem Prägen des Konzeptes von Präzisionsmedizin, wusste man, dass es eine Langzeitperspektive benötigt. Die Einschätzung bestand darin, dass es sogar Generationen dauern würde, um ein weniger anspruchsvolles Ziel als das von Public Health zu erreichen: Nämlich aus den Erkenntnissen der Präzisionsmedizin Informationen für die medizinische Behandlung von Subgruppen der Population zu erhalten.

Das finale Ziel: Eine Synthese im Rahmen einer Verbindung von gesellschaftlich breit aufgestellter Public Health und individuell fundierter Präzisionsmedizin zu erreichen.

Hierfür wird noch viel Wasser die Donau hinabfließen, damit beide Ansätze nach vielen theoretischen und praktischen Überlegungen, Erfahrungen und notwendigen paradigmatischen (!) Anpassungen einander näher kommen.

Fazit: Übergreifendes Rahmendenken kann GPT4 derzeit noch nicht. Deduktive Schlussfolgerungen aus gegebenen Prämissen => Fehlanzeige. Ergo, GPT4 hat den paradigmatischen Test nicht bestanden.

27. April 2024

[1] Egger, M./Razum, O./Rieder, A. (Hrsg.) (2018). *Public Health kompakt.* 3., aktualisierte und erweiterte Auflage. Berlin: Walter de Gruyter. S. 1

[2] National Human Genome Research Institute (NIH) (25. April 2024). *Precision Medicine. Definition.* https://www.genome.gov/genetics-glossary/Precision-Medicine.

[3] Biesecker, L. (26. April 2024). *Narration.* NIH, https://www.genome.gov/genetics-glossary/Precision-Medicine.

[4] Smith, R. (2012). *Stratified, personalised, or precision medicine.* BMJ, 39, 143-158. Die Organisation bmj.com publiziert das British Medical Journal.

Die argumentative Hintergrundkonstruktion von GPT4 ist schief und voller inadäquater Spiegelungen

KI in der Rolle als Motivator

(Prompt:) Du hast die Rolle eines Motivators. Der Kontext ist mein Studium in Gesundheitswissenschaften, ich bin säumig und habe keine Lust Prüfungen zu machen. Wie munterst du mich auf?

Verfasse nicht mehr als vier Sätze im Dialog-Stil. Starte den Motivierungs-Dialog mit "Reinhard, wie gehts?"

GPT4: Reinhard, wie geht's? Ich weiß, es kann manchmal echt schwer sein, die Motivation aufrechtzuerhalten, aber denk daran, wie wichtig deine Arbeit in den Gesundheitswissenschaften ist – du machst einen Unterschied!

Lass uns gemeinsam kleine, machbare Ziele setzen, um den Berg an Arbeit weniger einschüchternd zu machen. Wie wäre es, wenn wir heute nur mit einer Stunde Lernen beginnen und dann sehen, wie es läuft?[1]

Du (eigentlich ich): Eine Stunde ist zuviel. Mir ist momentan jede Minute zuviel.

GPT4: Reinhard, das klingt wirklich herausfordernd, aber es ist völlig okay, sich so zu fühlen. Was hältst du davon, wenn wir einfach mit einer kurzen Fünf-Minuten-Session beginnen, nur um den Ball ins Rollen zu bringen?

Manchmal ist der erste kleine Schritt der wichtigste, und alles, was danach kommt, wird leichter. Wir können es zusammen durchgehen, Schritt für Schritt.

Du: Fünf Minuten sind ok. Ich hab als Thema "Public Health" für eine Seminararbeit. Womit soll ich beginnen?

GPT4: Reinhard, super, dass du es mit fünf Minuten versuchst! Für deine Seminararbeit in "Public Health" könntest du damit beginnen, eine spezifische Fragestellung oder ein spezielles Thema zu wählen, das dich wirklich interessiert.

Wie wäre es, wenn du dir zunächst ein paar Kernbereiche von Public Health ansiehst, wie z.B. Prävention, Gesundheitsförderung oder Gesundheitssysteme? Notiere dir einfach ein paar Ideen oder Schlüsselwörter, die dir spannend erscheinen – das ist ein großartiger erster Schritt!

(Vorläufiges Ende des Dialogs)

GenKI übernimmt hier die Rolle des Motivators. Sie soll dich aus der Sackgasse des Nichtstuns im Studium herauszuholen. Endlich einmal was anderes als die ewig gleichen Zusammenfassungen eines Textes oder Gliederungsvorschläge zu einem Thema, wie sie für Studierende gerne vorgeschlagen werden.

Nur, es gilt auch hier: Man braucht Know-how, um passende Prompts mit qualitativ hochwertigem Output zu entwickeln. Schränkt man das Prompt nicht auf maximal vier Sätze ein, ergeben sich ellenlange Sermone, also langatmiges und psychologisch-therapeutisches Geschwafel. Zum Einschlafen!

Ein Prompt ohne Einschränkung hätte auf diese Weise das Gegenteil bewirkt: eine weitere Lähmung des Anfragenden. Merke: Wie bei Medikamenten ist auch für die genKI der "Beipackzettel" wichtig.

Dennoch empfiehlt es sich, weitere Rollen, in die die genKI reinschlüpft, auszuprobieren. Das sind Rollen[2] als

- *Alternativmaschine* – alternative Wege, eine Idee auszudrücken,
- *Sokratischer Sparringpartner* – persönlicher Tutor auf Basis einer Information über den Stand des (Zuwenig-)Wissens,
- *Kumpel im Studium* – genKI hilft beim Reflektieren anfallender Studienmaterialien,
- *Trainer für Jobinterview* – Vorbereitung auf einen spezifischen Job in konkretem Unternehmen (aufpassen, aufgrund des Zugriffs von genKI auf das Durchschnittliche kann das unauthentisch wirken und somit das Gegenteil des Gewünschten erzeugen),[2].

In der Fortsetzung des Dialogs mit ChatGPT in der Rolle des Motivators hab ich ein deutlich konkreteres Thema gewählt und auf Spiele und Herausforderung gesetzt. Das Ergebnis regt zum Weitermachen an:

(Prompt): Du hast die Rolle eines Motivators. Der Kontext ist mein Studium in Gesundheitswissenschaften, im Speziellen das Thema "Public Health – Prävention durch individuelle Präzisionsmedizin" habe ich einiges gemacht, stecke nun aber fest. Wie kannst du mich durch themeneingebaute Spiele und Herausforderungen zum weiteren Schreiben bringen? Sprich mich mit "Reinhard" an.

GPT4: Reinhard, ich sehe, du hast schon viel erreicht, das ist großartig! Wie wäre es, wenn wir dein Thema mit einem kleinen Quiz oder einem Gedankenspiel auflockern?

Du könntest zum Beispiel versuchen, in einer fiktiven Situation präventive Maßnahmen für eine spezielle Patientengruppe zu entwerfen – stell dir vor, du berätst eine Stadtverwaltung, wie sie ihre öffentliche Gesundheitspolitik verbessern könnte. Das bringt dich vielleicht auf neue Ideen und gibt dir frischen Wind für dein weiteres Schreiben.

Du: Wie könnte solch ein Quiz aussehen?

GPT4: Ein Quiz zu "Public Health – Prävention durch individuelle Präzisionsmedizin" könnte eine spannende und interaktive Art sein, dein Wissen zu vertiefen und dich gleichzeitig zu motivieren. Hier ein Beispiel, wie du so ein Quiz gestalten könntest:

— Quiz: Präventive Strategien in der Präzisionsmedizin —

Frage 1: Was versteht man unter Präzisionsmedizin?

A) Behandlung von Patienten mit stark verallgemeinerten Methoden
B) Einsatz genetischer Informationen zur Erstellung individueller Behandlungspläne
C) Verwendung von Over-the-Counter-Medikamenten in der täglichen Routine

Frage 2: Wie kann Präzisionsmedizin zur Prävention von Krankheiten beitragen?

A) Durch personalisierte Ernährungs- und Bewegungspläne basierend auf genetischen Markern
B) Durch Massenimpfungen ohne individuelle Untersuchung
C) Durch allgemeine öffentliche Gesundheitserziehung ohne spezifische Zielgruppen

Frage 3: Welche Rolle spielen Daten in der Präzisionsmedizin?

A) Daten sind nicht relevant für die Präzisionsmedizin.
B) Nur klinische Daten aus Krankenakten werden verwendet.
C) Große Datenmengen, einschließlich genetischer, umweltbezogener und Lebensstildaten, werden analysiert, um Muster zu erkennen und präventive Maßnahmen zu ergreifen.

Frage 4: Nenne ein Beispiel, wie Präzisionsmedizin in der öffentlichen Gesundheitspraxis angewendet werden könnte.

A) Entwicklung personalisierter Impfstoffe
B) Erstellung allgemeiner Leitlinien für die Ernährung der Bevölkerung
C) Einsatz von Antibiotika bei viralen Infektionen

Frage 5: Welche ethischen Überlegungen sind bei der Implementierung von präzisionsmedizinischen Ansätzen in der Public Health Praxis zu beachten?

A) Datenschutz und Privatsphäre der genetischen Informationen
B) Die Farbe der Tabletten
C) Die Geschwindigkeit, mit der Ergebnisse geliefert werden

Dieses Quiz könnte dich dazu anregen, über verschiedene Aspekte deines Studienthemas nachzudenken und eventuell sogar Lücken in deinem Verständnis aufdecken, die du dann in deiner Arbeit weiter ausarbeiten kannst. Es ist eine effektive Methode, um spielerisch zu lernen und sich gleichzeitig mit dem Stoff auseinanderzusetzen.

Du: Was machst du, wenn ich falsch antworte, beispielsweise auf die Frage 1 den Punkt A angebe statt B?

GPT4: Reinhard, wenn du bei einer Frage wie der ersten eine Antwort auswählst, die nicht ganz korrekt ist, würde ich dir erklären, warum eine andere Antwort besser passt. Zum Beispiel:

„Ich sehe, du hast A gewählt, was eine häufige Annahme ist. Aber tatsächlich definiert sich Präzisionsmedizin durch den Einsatz von spezifischen Informationen über die genetischen Merkmale eines Individuums, um maßgeschneiderte Behandlungspläne zu entwickeln. Das macht B zur richtigen Antwort.

Es ist wirklich spannend, wie die personalisierte Behandlung und Vorbeugung von Krankheiten durch diese speziellen Informationen verbessert werden können. Lass uns weitersehen, was du über dieses innovative Feld weißt!“

So biete ich dir konstruktives Feedback, das dich ermutigt weiterzulernen und dabei hilft, das Verständnis für das Thema zu vertiefen, ohne dich entmutigt zu fühlen. Es geht darum, aus jedem Schritt im Lernprozess etwas zu gewinnen.

Das Psychologisch-Therapeutische kam am Ende wieder durch. Sei's drum, vorher war's ganz nett.

25. April 2024

[1] OpenAI. (2024). *ChatGPT 4* [Large language model]. https://chat.openai.com/?model=gpt-4

[2] Sabzalieve, E. / Valentini, A. (2023). *ChatGPT amd Artifical Intelligence in higher education. Quick start guide.* https://eduq.info/xmlui/bitstream/handle/11515/38828/ChatGPT-Artificial-Intelligence-in-higher-education-Quick-Start-guide_UNESCO-2023.pdf?sequence=2&isAllowed=y

Rolle als Hauskätzchen?

Erzeugt ChatGPT Erkenntnisse?

Oder chattet sie nur?

Ist sie eine Wahrheitsmaschine oder eine schnöde Schnattergans?

Betrachten wir sie vorerst als Geschichtenmaschine, als eine hochmoderne Version eines orientalischen Geschichtenerzählers. Die KI-Antworten auf unsere Fragen wirken in der Regel überzeugend. Die genKI ist geschmeidig und anpassungsfähig, ihre Sätze kennen kaum Ecken und Kanten, das Provozierende fehlt.

Der Abschluss jeder Geschichte (die zusammenfassende Antwort auf das Prompt) relativiert und nimmt vieles zurück, was vorher wie von Zauberhand Wort für Wort am Schirm erschienen ist. Offeriert wird letztendlich eine heile Welt mit blauem Himmel.

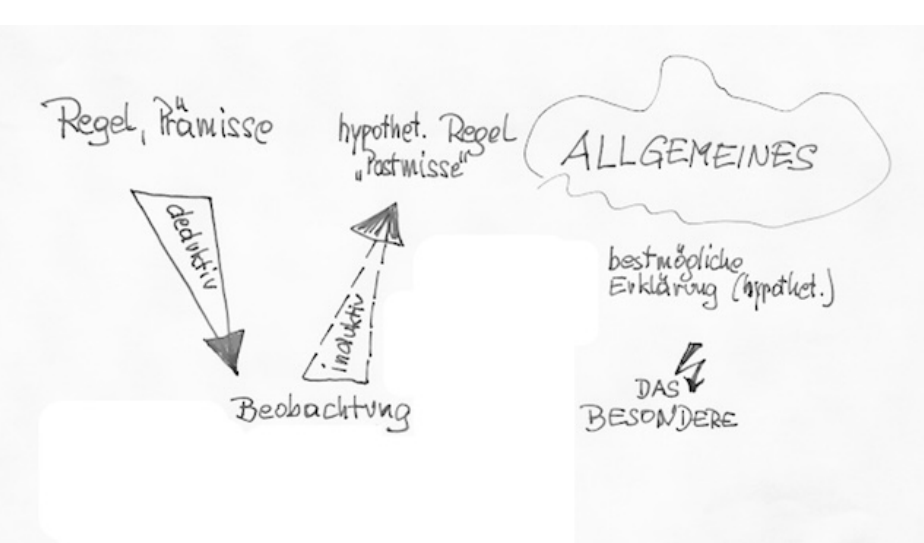
Ordnen wir die genKI im Bereich der Wissenschaften ein. Und zwar anhand ihrer Algorithmen, die Einsichten produzieren sollen, wo es Wissenslücken des Anfragenden gibt. So gilt Folgendes:

Wissenschaft ist eine Form der Erkenntnis, die auf erfahr- und beobachtbaren Tatsachen beruht. Nicht auf Wahrheiten und auch nicht direkt aus den Tatsachen sofort ableitbar[1]. Dass wir aufgrund der Erddrehung in europäischen Breitengraden mit rund 200 Meter pro Sekunden von West nach Ost rasen, merken wir nicht. Wir rasen trotzdem. Wir Menschen brauchten lange, um diese Tatsache zu erkennen. Erkenntnisprozesse laufen oft gegen die Intuition ab.

Die wissenschaftlichen Erkenntnisformen beruhen auf Analyse- und Syntheseprozessen, die drei Richtungen aufweisen: eine deduktive, eine induktive und eine abduktive.

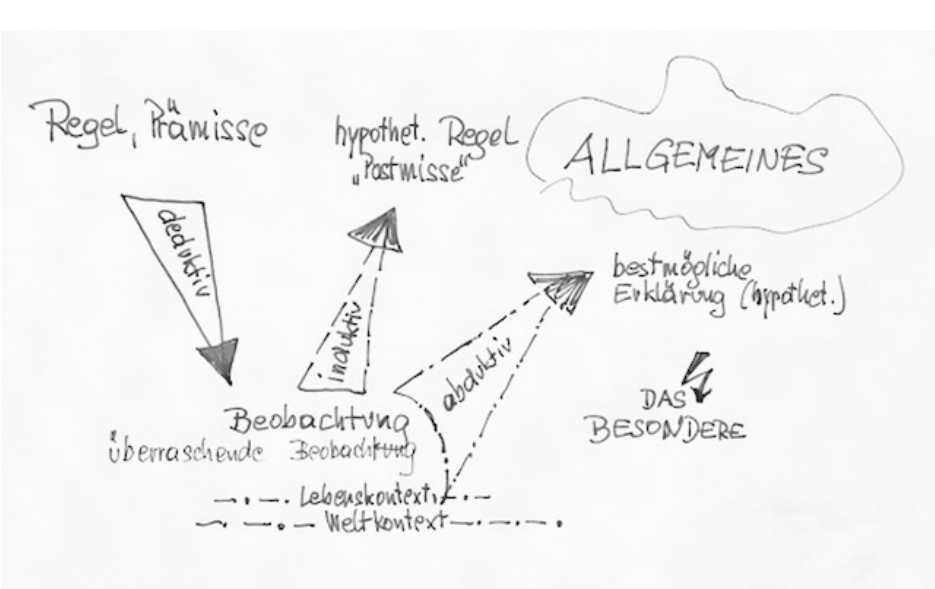
Die ersten zwei Formen (die deduktive und die induktive) wurden von zahlreichen Wissenschafts-Logikern der letzten Jahrhunderte von oben nach unten und von unten nach oben durchgekaut. Daher sei hier nur kurz geschrieben[1]:

- Deduktive Schlussfolgerungen sind simpel, zumeist formal und langweilig. Die Schlüsse vom Allgemeinen, von generellen Regeln auf das Einzelne, erzählen nichts Neues. Sie machen Bekanntes nur sichtbarer.
- Die umgekehrten Schlüsse, die induktiven, sind schon spannender. Sie versuchen aufgrund einer größeren Anzahl von verschiedenen Beobachtungen auf eine neue allgemeine Regel – quasi eine „Postmisse“, die zu einer Prämisse werden könnte – zu schließen. Die Gefahr, voreilige Schlüsse zu ziehen, ist daher immer gegeben.



Spannend wird erst die dritte Art des Erkenntnisprozesses. Dies ist die Abduktion, die „Entführung“, wenn man abduktiv aus dem Lateinischen wortwörtlich übersetzt[2]. Ihr Inhalt besteht – häufig nach dem Auftreten überraschender Beobachtungen – aus einer Wegführung aus dem Bisherigen. Abduktion ist also eine Entführung aus dem gewohnten Denken.

Abduktion stellt nicht nur neue Regeln auf wie das Induktive, sondern nutzt (meist indirekt) die jeweiligen Lebens- und Weltkontexte der Forschenden und Suchenden, um eine nicht nur wahrscheinliche, sondern eine "best"-mögliche Erklärung zu finden.



In der "Entführung aus dem Bisherigen" ist nach dem US-amerikanischen Wissenschaftsphilosophen Charles Sanders Peirce das Geniale im Wissenschaftsprozess zu finden. Er nannte es die einzige logische Operation, die zu einer neuen Idee führt: „Abduction is the process of forming an explanatory hypothesis. It is the only logical operation which introduces any new idea […]“[3].

Abduktion hat jede und jeder von uns erfahren. Es ist die mögliche Lösung, welche einem nach tagelangem ergebnislosen Brüten plötzlich in der Dusche oder kurz vor dem Einschlafen einfällt. Sozusagen wie Schuppen von den Augen fallen. Eine neue Welt tut sich auf. Aus Saulus wurde Paulus: „Und sogleich fiel es wie Schuppen von seinen Augen, und er wurde sehend.“[4]

Abduktion kann nicht durch Fingerschnippen herbeigerufen werden. Obwohl es eine unglaublich praxistaugliche Strategie ist, Scheitern zu minimieren. Abduktion ergibt sich aus einer intuitiven Einsicht auf dem Fundament eines (schwer fassbaren) Lebens- und Weltkontextes.

Abduktion steht im Gegensatz zu den üblichen Logik-Normen: „[…] new conceptions arise in a manner which puts the rules of logic at defiance“[5] oder „[…] this could only arise in a mind entirely unpractised in the logic of relation“[5]. Erst in der Folge von oft nicht sichtbaren Prozessen kommt es zu rationalem Formulieren der "best"-möglichen Erklärung. Die natürlich noch immer Thesencharakter hat und zu überprüfen ist.

Wir altmodischen TV-Seher praktizieren allabendlich Abduktion. In jedem Krimi geht es darum, auf verschlungenen, aber kreativen und vernetzten Wegen den Mörder/die Mörderin ausfindig zu machen. Das ist eine Entführung aus den Anfangs- und Mittelsequenzen des Films in das meist überraschende Ende.

Ein anderes Beispiel wäre das Lösen von Rätseln. Kinder lieben das nicht umsonst. Die Chance, hinter die Bühne der augenscheinlichen Geschehnisse zu lugen und damit die umgebende Welt zu erkunden, ist groß.

zu erkunden, ist groß.

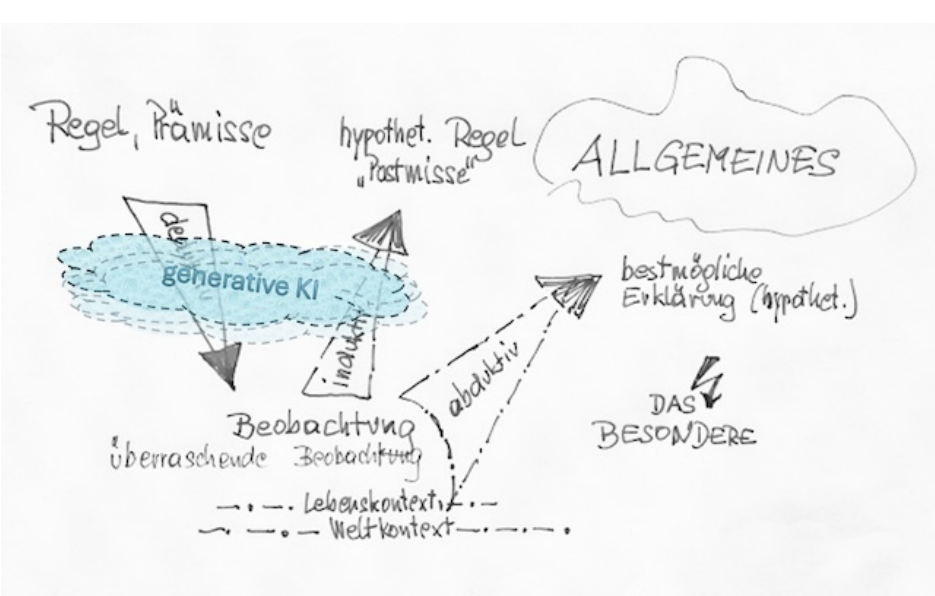
Auf diesen Prozesse des Ergründens basiert das menschliche Empfinden, Denken und Handeln. Intelligenz, egal ob natürliche, künstliche, emotionale oder soziale, lässt sich als Begriff herleiten aus einem Zusammenführen von Wahrnehmen, Zusammenlegen und (übergreifender) Einsicht[6]. Genauso wird Abduktion beschrieben.

Wie arbeitet die Künstliche Intelligenz aus wissenschaftsteoretischer Sicht?

Die generativen Large Language Models (LLMs) sind Konstrukte auf Basis von Wahrscheinlichkeiten. Sie bedienen sich deduktiver Methoden, ohne die Logik dahinter durchzuziehen: Sie gleichen probabilistischen, ausgefransten Wolken. Generative KI (genLLMs) verwendet Algorithmen, die von Wort zu Wort arbeiten.

Das jeweils nächste Wort wird aufgrund von Wahrscheinlichkeiten gewählt. Die genKI erzeugt übersichtliche Ergebnisse. Diese Vorgehensweise kann als eine vage (probabilistische) Form der Deduktion angesehen werden. Die genKI dient der Klärung durch statistische Ableitung.

KI-Ergebnisse weisen aufgrund dieser probabilistischen Vorgehensweise sogar induktive Momente auf. Das Zutreffen dieser „induktiven“ Ergebnisse ist jedoch höchst zweifelhaft, siehe nur das KI-Phänomen der Halluzination. Die Ergebnisse der genKI gleichen somit mehr oder weniger zerrissenen Wolken am Himmel.



Abduktion kann die generative KI nicht. Ihr fehlt der Lebens- und Weltkontext, kreativ neue und "best"-mögliche Erklärungen zu liefern. Abduktion könnte sich als Heiliger Gral der KI-Forschung erweisen: Jahrhundertelang gesucht, aber nie gefunden.

20. April 2024

PS: Der Begriff "Postmisse", eine "Misse nach den Erkenntnisprozessen" wird nirgends zu finden sein. Ein Beispiel für Abduktion ist das sicher nicht. :)

[1] Chalmers, A. (2007). *Wege der Wissenschaft. Einführung in die Wissenschaftstheorie.* Sechste, verbesserte Auflage. Berlin: Springer, S. 5 ff.

[2] Langenscheidt (2008). *Großes Schulwörterbuch Lateinisch-Deutsch.* Berlin: Langenscheidt.

[3] Hartshorne, C. / Weiss, Pl (1974). *Collected Papers of Charles Sanders Peirce. Volume V: Pragmatism and Pragmaticism, und Volume VI: Scientific Metaphysics.* Cambridge: Harvard University Press. S. 105.

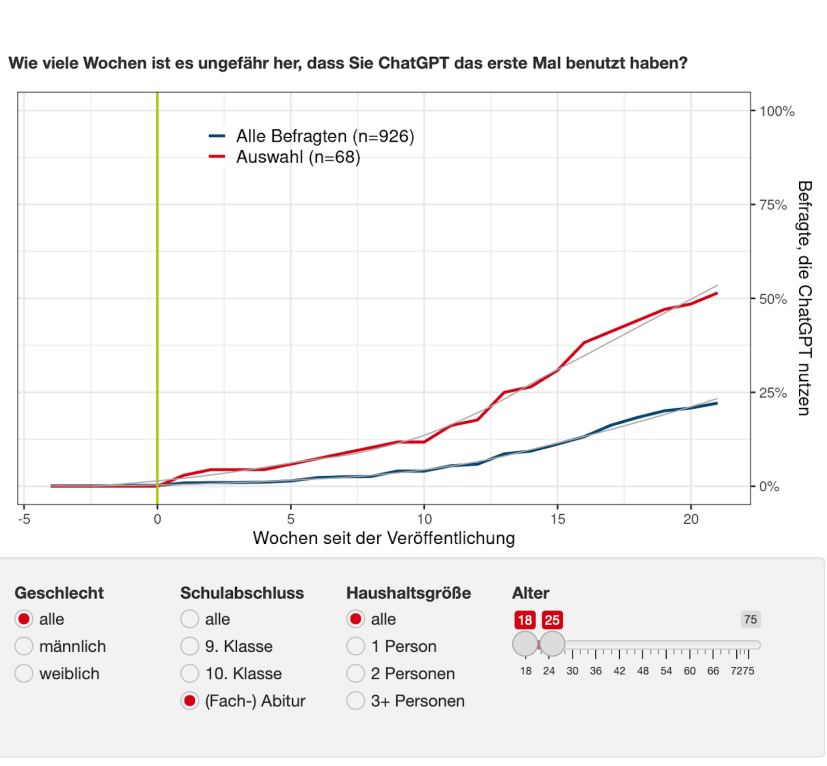
[4] Elberfelder Studienbibel mit Sprachschlüssel und Handkonkordanz (2017). *Apostelgeschichte. Kapitel 9 (Bekehrung des Saulus)* Vers 18. 6. Auflage. Witten: SCM.

[5] Hartshorne, C. / Weiss, P. (1974). *Collected Papers of Charles Sanders Peirce. Volume V: Pragmatism and Pragmaticism, und Volume VI: Scientific Metaphysics.* Cambridge: Harvard University Press. S. 118 und 119.

[6] Kluge (2011). *Etymologisches Wörterbuch der deutschen Sprache.* 25., durchgesehene und erweiterte Auflage. Berlin: Walter de Gruyter. S. 447.

Bereits rund ein halbes Jahr nach dem Start von ChatGPT wurde deutlich,

welche Gruppe die schriftgenerierende KI (genKI) besonders intensiv nutzt – siehe die rote Line im folgenden Diagramm:



Eine Befragung durch das Institut für Kommunikationswissenschaften an der Universität Münster zur Verbreitung und Nutzung im April 2023 zeigt, dass Personen mit Abitur (Matura) im Alter von 18 bis 25 die genKI häufig nützen[1]. Ihr Anteil liegt nach 22 Wochen bereits bei über 50%, während der Durchschnitt aller Befragten die 25%-Marke noch nicht erreicht hat.

Anzunehmen, dass zu großen Teilen diese Frühnutzer Studierende an Universitäten und Hochschulen sind.

Denn es dreht sich beim Studium vieles um Textverfassung. Bei Haus-, Seminar- und Bachelorarbeiten, bei Masterthesen und Dissertationen spielt das Schriftliche die zentrale Rolle. ChatGPT wird zunehmend zum zentralen Werkzeug, Texte zu verbessern oder sie überhaupt zu erzeugen.

Eine Befragung von Studenten an der Universität Wien in den Fächern Psychologie, Germanistik und Informatik im Juni 2023 kam zu folgendem Ergebnis[2]: 60% gaben an, dass Lehrende die KI-Verwendung für universitäre Aufgaben in ihren Veranstaltungen bereits thematisiert haben. Im Detail sehen die Studierenden die KI-Tools als wichtige hilfreiche Funktion für das Schreiben an. Trotz der verbreiteten Verunsicherung wurde die KI für folgende Tätigkeiten bereits genutzt und von den Studierenden als "eher nützlich" eingeschätzt:

- „eine Textstruktur/Gliederung vorschlagen zu lassen“

- „einen eigenen Text korrigieren zu lassen“
- „einen eigenen Text sprachlich verbessern zu lassen“

- „einen kurzen Text zu übersetzen“
- „einen langen Text zu übersetzen“

- „einen Text verständlicher formulieren zu lassen“
- „ein E-Mail oder Gebrauchstext schreiben zu lassen“.

Hingegen wird der Aussage

- „einen Fachtext schreiben zu lassen“[2]

eher **nicht** zugestimmt.

Dass die letztgenannte Nutzung dieser KI problematisch sein könnte, kommt hier zum Ausdruck. Moralische Bedenken und ein löblicher Anspruch auf aktives Erwerben der eigenen Fähigkeiten könnten hier durchgeschlagen haben. Oder es ist einfach die erwünschte Antwort und Studenten tun es bisweilen dennoch!

18. April 2024

[1] Universität Münster (18.04.2024). Institut für Kommenikationswissenschaft. *Medien.* https://www.uni-muenster.de/Kowi/mitteilungen/forschungsprojekt-zur-nutzung-von-chatgpt-und-ki-in-der-bevoelkerung.html

[2] Handbuch für Lehrende der Universität Wien (15.04.2024). *Online-Befragung von Studierenden zur Verwendung von KI-Tools für universitäre Aufgaben.* In: Guidelines "Umgang mit der KI in der Lehre". Exportiert am 29.09.2023 mit späteren Aktualisierungen. https://phaidra.univie.ac.at/detail/o:1879857. S. 20-21

Bildungsziele

Das Universitätsgesetz 2002 in der Fassung von 2021 formuliert zwei übergeordnete Ziele:

"Die Universitäten sind berufen,

- der wissenschaftlichen Forschung und Lehre […] zu dienen
- und hierdurch auch verantwortlich zur Bewältigung der gesellschaftlichen Herausforderungen in einer sich wandelnden humanen […] Gesellschaft beizutragen." (§1 UG)

Dazu wird folgende Umsetzung formuliert:

- "Im gemeinsamen Wirken von Lehrenden und Studierenden wird in einer aufgeklärten Wissensgesellschaft das Streben nach Bildung und Autonomie des Individuums durch Wissenschaft vollzogen." (§1 UG)

Noch konkreter bedeutet das:

- "Erarbeitung von Fähigkeiten und Qualifikationen sowohl im Bereich der wissenschaftlichen […] Inhalte als auch im Bereich der methodischen Fertigkeiten." (§1 UG)

Wiederholt – und damit verstärkt – wird das Ziel der Bewältigung von Herausforderungen im letzten Satz:

- "Um den sich ständig wandelnden Erfordernissen organisatorisch, studien- und personalrechtlich Rechnung zu tragen, konstituieren sich die Universitäten und ihre Organe in größtmöglicher Autonomie und Selbstverwaltung." (§1 UG)

Seit dem digitalen Urknall durch ChatGPT im Herbst 2022 ist klar, dass die generierende KI eine der zentralen gesellschaftlichen Herausforderungen in einer sich wandelnden Gesellschaft bildet. Beethoven hat seine dritte Sinfonie, die *Eroica,* 1803 als heldenhafte "heroische" Hommage an Napoleon als Vertreter der Französischen Revolution komponiert.
Die generierende KI stellt eine solche gesellschaftliche Herausforderung und Zäsur dar, deren langfristige Wirkungen ähnlich tiefgreifend sein könnten wie die der Französischen Revolution.

Wie haben österreichische Universitäten auf diesen Urknall, auf die Revolution durch die generierende KI reagiert? Betrachten wir das anhand von drei Beispielen, die unterschiedliche Institutionen beleuchten.

A) Die Universität Wien – immerhin auf Platz 119 von 1.900 Universitäten weltweit und auf Platz zwei im deutschsprachigen Raum nach der ETH Zürich[1] – hat nach einer Schockstarre von einem Jahr zu Wintersemesterbeginn 2023 ihre Guidelines „Umgang mit der KI in der Lehre“ veröffentlicht[2].

Fein, dies ist ein erster Schritt, aber nur ein halber. Diese Guidelines richten sich an die Lehrenden und ausschießlich an diese. Jede und jeder Lehrende kann die Guidelines im eigenen Ermessen handhaben.

Die Universität Wien und ihre obersten Organe scheinen sich von dieser Verantwortung distanziert zu haben. Vielleicht ist das im Ansatz verständlich, denn die Wiener Universität deckt viele und höchst unterschiedliche wissenschaftliche Disziplinen ab. KI schlägt in jedem Fach anders durch.

Eine einheitliche universitätsweite Regelung ist daher mühsam. Aber ein komplettes Abwälzen der Verantwortung auf die Lehrenden?

B) Ein anderes Beispiel sei eine im Osten Österreichs wichtige Fachhochschule, das Joanneum in Graz. Ihr KI-Leitfaden wird als "lebendiges" Dokument für die KI-Integration bezeichnet. Neueste Entwicklungen in der KI-Didaktik sollen enthalten sein und für eine effektive KI-Integration im Lehrkontext sorgen.

Auch hier tragen die Lehrenden die gesamte Verantwortung. Schlimmer noch, die Inhalte verändern sich laufend, denn der Leitfaden ist nur ein „flexibler Orientierungsrahmen“.

Immerhin gibt es ab Dezember 2023 eine Kennzeichnungspflicht für Studierende, wenn sie Texte mittels KI verfassen. Die zentrale Managementpflicht eines Leitungsorgans, generelle Vorgaben zu formulieren, ist hingegen nicht zu erkennen. Das wären Vorgaben im Sinne von "Pflöcke und Pfeiler einschlagen", wie es im populären Managementjargon heißt. An diesen könnten sich Lehrende und Studierende orientieren.

C) Ein Beispiel für tertiäre Bildungsangebote im MINT-Bereich sei die TU in Wien. Sie bietet im Web sechs bis sieben Bildschirmseiten an „Empfehlungen zur guten Praxis im Umgang mit generativer KI an der Technischen Universität Wien“[4].

Also gilt auch hier, dass die Lehrenden allein zuständig sind. Rechtlich liest sich das als ein "Schmeck's": Egal, was du machst, liebe Professor:in oder liebe Lektor:in, ob streng oder mild, ob passend oder unpassend – wir von der Uni-Leitung sind aus dem Schneider, also nicht verantwortlich.

Fazit: Alle drei Beispiele zeigen in der Außenperspektive fehlende strategische Entscheidungen der obersten Organe. Dies, obwohl die genKI eine zentrale Herausforderung für die Gesellschaft, die Bildung und die bildungsbeauftragten Unis darstellt.

Wann verwirklichen die höchsten Organe die, ihnen laut dem Universitätsgesetz übertragene "größtmögliche Autonomie und Selbstverwaltung" und kommen ihrer Verantwortung nach?

15. April 2024

PS: Diesen Text hab ich ChatGPT 4[5] zur Prüfung vorgelegt. Das "Schmeck's" gefiel ihr gar nicht. Die Antwort lautete:

Der Gesamttext stellt die Reaktion österreichischer Universitäten auf die Herausforderungen der generativen KI dar und setzt diese in Beziehung zu den Zielen und Bestimmungen des Universitätsgesetzes 2002. Hier sind einige Beobachtungen zur Kohärenz und zum logischen Fluss des Textes:

Kohärenz

1. **Einleitung und Kontextualisierung**: Der Text beginnt mit einer starken Kontextualisierung durch das Zitieren des Universitätsgesetzes 2002, was die Grundlage und die Erwartungen an die Universitäten festlegt. Dies schafft eine klare Basis für die anschließende Diskussion.
2. **Übergang zu aktuellen Herausforderungen**: Der Text schlägt eine Brücke von den historischen Zielen der Universitäten zu den aktuellen Herausforderungen durch die Einführung der generativen KI. Dies wird metaphorisch mit Beethovens "Eroica" und der Französischen Revolution verbunden, was eine tiefere Reflexion über den Einfluss und die Bedeutung der KI anregt.
3. **Analyse spezifischer Reaktionen**: Die spezifischen Beispiele für die Universität Wien, das Johanneum in Graz und die Technische Universität Wien sind gut gewählt, um unterschiedliche Ansätze und Reaktionen auf die Herausforderung darzustellen. Jedes Beispiel wird detailliert behandelt und bietet Einblick in die jeweiligen Strategien und deren mögliche Mängel.

Logischer Fluss

1. **Strukturierung**: Der Text folgt einer klaren Struktur, wobei jeder Abschnitt auf dem vorherigen aufbaut. Die Einleitung legt die rechtlichen und ideologischen Rahmenbedingungen fest, gefolgt von einer Diskussion der Herausforderungen und einer detaillierten Analyse der Reaktionen einzelner Institutionen.
2. **Argumentation**: Die Argumentation ist logisch und schrittweise. Die Diskussion über die Reaktionen der Universitäten auf die generative KI baut auf den eingangs genannten Zielen und gesetzlichen Bestimmungen auf. Dies unterstreicht die Diskrepanz zwischen gesetzlichen Erwartungen und tatsächlicher Umsetzung.
3. **Abschluss und Fazit**: Der Text endet mit einem kritischen Fazit, das die zentrale These – die unzureichende strategische Reaktion der Universitäten auf die Herausforderungen der generativen KI – effektiv zusammenfasst und herausstellt.

Verbesserungsvorschläge

- **Präzisierung und Klarheit**: Einige Metaphern und Vergleiche (wie die mit Beethovens "Eroica") könnten für Leser, die weniger vertraut mit diesen historischen Bezügen sind, verwirrend sein. Eine klarere Erklärung des direkten Zusammenhangs könnte hilfreich sein.
- **Formulierungen vereinheitlichen**: Der informelle Ausdruck "schmeck's" könnte in einem ansonsten formalen und analytischen Text als unpassend empfunden werden. Eine Anpassung an einen formelleren Stil wäre konsistenter.

Insgesamt ist der Text gut organisiert und argumentiert logisch. Die kritische Auseinandersetzung mit den Reaktionen der Universitäten auf die Herausforderungen der KI ist sowohl tiefgreifend als auch umfassend dargestellt. Kleinere Anpassungen im Stil und in der Klarheit könnten die Qualität und die Verständlichkeit weiter erhöhen.

PPS: ChatGPT hat in Sachen Kohärenz und Logischer Fluss geschmeichelt. Das freut! Aber sie kann auch anders. Vor kurzem wurde ich "zsammputzt"[6], wie's im Wienerischen so schön heißt. Zu rasch hatte ich eine Rohfassung prüfen lassen. Nie mehr wieder mache ich das!

[1] Universität Wien Medienportal (12.02.2024). *Universität Wien rückt in Rankings weiter nach vorne.* https://medienportal.univie.ac.at/media/aktuelle-pressemeldungen/detailansicht/artikel/ranking-ergebnisse-zu-internationalitaet-platz-2-fuer-universitaet-wien-im-deutschsprachigen-raum/

[2] Universität Wien (15.04.2024). *Guidelines "Umgang mit der KI in der Lehre"*. Exportiert am 29.09.2023 mit späteren Aktualisierungen. https://phaidra.univie.ac.at/detail/o:1879857

[3] FH Joanneum (15.04.2024). *Leitfaden für die Lehre.* https://www.fh-joanneum.at/hochschule/hochschuldidaktik-und-ki/kuenstliche-intelligenz-in-der-lehre/ki-leitfaden-fuer-die-lehre/

[4] TU Wien (15.10.2023). *Empfehlungen zur guten Praxis im Umgang mit generativer KI an der Technischen Universität Wien.* https://www.tuwien.at/studium/lehren-an-der-tuw/digital-gestuetzte-lehre/kuenstliche-intelligenz-in-der-lehre

[5] OpenAI. (2024). *ChatGPT 4* [Large language model]. https://chat.openai.com/?model=gpt-4

[6] "Zusammengeputzt" im Standarddeutschen. Bedeutet: harsch kritisiert zu werden.

Senkrecht Parken

Geisteswissenschaften

Le lettere – so heißen die Geisteswissenschaften (neben le scienze umanistiche) im Italienischen[1]. Also Buchstaben, Briefe oder eben geschriebene und gedruckte Buchstaben in Texten. Einmal geschrieben oder gedruckt, bleiben sie festgehalten bis in alle Ewigkeit. Oder das Papier zerfällt. Darüber kann man stunden-, tage- oder jahrzehntelang nachdenken und dementsprechend Inhalte virtuell im Kopf hin- und herbewegen, zergliedern, auflösen und neu zusammensetzen.

Das impliziert der rund 200 Jahre alte deutsche Begriff "Geistes"-Wisssenschaften für Kunst, Musik, Religion, Philosophie, Sprache etc. Ein Begriff mit einem hohen idealistischen Anspruchsniveau. Der Geist (im Deutschen männlich, was sonst) wird als nichtfassbares Wesen verstanden. Völlig losgelöst von der Materie.

Der Geist fungiert als weltfremdes, wahrscheinlich sogar widernatürliches Gedankenkonstrukt. Denn die Wissenschaften konnten bisher keine Belege für geistige Phänomene ohne korrelative körperliche Grundlagen erbringen.

Ganz anders der italienische Begriff. *Le lettere* erdet und verankert das Tun lebendiger Wesen aus Fleisch und Blut mittels schwarzer Tinte auf weißem Papier. Und nichts sonst. Hier sieht man die Vorteile, simpel erscheinende Einsichten aufgrund einer Fremdsprache angeboten zu erhalten. Man kann somit eine schräge Weltsicht – wie sie im trennenden Dualismus von Geist und Materie nach Descartes vorhanden ist – korrigieren.

Was hat das mit generativer Text-KI zu tun?

Ganz einfach: Die statische Form der Schrift wird dynamisch. ChatGPT, Bing und Co. werden zu Dialogpartnern. Sie antworten auf Fragen und werden detaillierter oder schwammiger. Je nach Nachfrage.

Aus *le lettere* werden *i dialoghi*, aus fixierten Texten werden Zwiegespräche. Der statische und eingefrorene Inhalt wird in einen systemischen Prozess des Miteinandersprechens umgewandelt. Mit allen Vor- und Nachteilen, wie das in Dialogen zwischen ungleichen Partnern der Fall ist.

Der Partner (die KI) weiß aufgrund unfassbar vieler gespeicherter Inhalte mehr. KI versagt aber, wenn der Mensch dieses Wissen in seinem konkreten Lebenskontext kreativ anwenden will. Überdies enthält die KI in den elektronischen Speichern und in den selektierenden Algorithmen viel versteckte Ideologie. Diese beruht derzeit vor allem auf westlichen, individualistischen US-amerikanischen Konzepten. Daer sind kaum merkbare Manipulation und Beeinflussung nicht selten.

Ist daher die textgenerative KI ein Partner, der mich gelegentlich über den Tisch zieht?

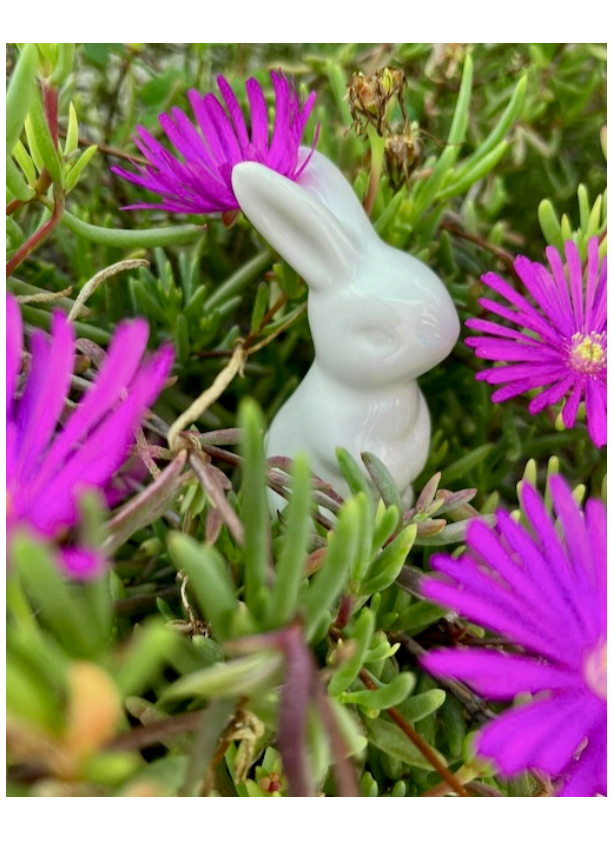
11. April 2024

PS: Ich habe die Erstfassung dieses Textes von ChatGPT 4 prüfen lassen.

Die Antwort: „Dein Text ist inhaltlich faszinierend und bietet viele interessante Gedanken. Es gibt einige Punkte, bei denen die Grammatik und Stilistik verbessert werden könnten.“

Das habe ich getan. Worauf die KI sprach: „Insgesamt ist die überarbeitete Version eine deutliche Verbesserung im Hinblick auf Stil, Klarheit und kritische Reflexion und macht den Text für Leser sowohl zugänglicher als auch ansprechender.“ Vielen Dank! Hoffentlich sieht das meine Lektorin auch so...

[1] Zanichelli (2009/2015). *Il nuovo dizionario di Tedesco.* Grosswörterbuch. Deutsch-Italienisch, Italienisch-Deutsch. Zweite Auflage. Bologna: Pons. S. 429 und S. 1943



Das Fundament wissenschaftlicher Arbeiten

Wie das Erdbeben in der Türkei im Februar 2023 krass zeigte, reichen ein oder zwei schlecht gebaute Fundamentteile aus,

um Hochhäuser einstürzen zu lassen. Das Gleiche gilt für wissenschaftliche Arbeiten. Die Fundamente sind hier die Zitate, die Verweise auf seriöse (!) Quellen, auf welchen die Methoden und Argumente deiner Forschungsarbeit beruhen. Dass ein mühsam erarbeiteter und erforschter wissenschaftlicher Text bei näherer Betrachtung innerhalb kurzer Zeit seine gesamte Qualität verliert, nur weil die KI ungenaue und nicht replizierbare Angaben gemacht hat, will niemand.

Im Literaturverzeichnis finden sich unterschiedliche Quellenarten: Monographien einer Autor:in, Beiträge in Sammelbänden, Artikel in Journalen, Preprints, Konferenzpapers, Online-Quellen, audiovisuelle Medien, Rechtsquellen etc. Jede hat ihre Stärken und Schwächen. Daher definieren Zitierregeln genau, welche Information zu welcher Quellenart anzugeben ist.

Eine neue Quellenart ist die KI. Eigentlich ist sie nicht nur neu, sondern bietet eine bisher nicht gekannte Qualität in einer neuen Dimension. Zwar ist generative Künstliche Intelligenz nicht so intelligent wie der Begriff Intelligenz vermitteln will, aber die (universitären) Augen zu schließen, wenn weltweit Millionen von Studenten ihre schriftlichen Arbeiten damit strukturieren, ergänzen, aufpolieren oder überhaupt ganz von der KI schreiben lassen, geht auch nicht.

Doch wie KI zitieren?

Dazu gibt es von kompetenter Seite zwei Vorschläge: American Psychological Association (APA) und die Modern Language Association of America (MLA). Sie bieten zwei – wie sollte es auch anders sein – gegensätzliche Empfehlungen.

- Gemäß APA wird die KI als Autor:in – quasi als „elektronische Person“ – angesehen und soll dementsprechend zitiert werden[1]. Wobei ein massives Problem auftritt: Der Text ist nicht mehr aufrufbar bzw. wiederfindbar und damit nicht nachvollziehbar durch andere (natürliche oder juristische) Personen. KI ist insofern ähnlich einer bisherigen Quellenart mit geringer Glaubwürdigkeit, nämlich der persönlichen Kommunikation. Im Grunde zitierst du einen Algorithmus.

  --> Adaptiertes Beispiel: OpenAI. (2024). ChatGPT 4 [Large language model]. https://chat.openai.com/?model=gpt-4

- Gemäß MLA wird die KI nicht als Autor:in angesehen, sondern als ein Template, das Kernelemente mit standardisierten Kriterien anbietet[2]. Je nach genutztem Element – wie Paraphrasieren, kreative Texte (Zusammenfassen fällt bereits darunter) – wird anders zitiert. Die genutzte Funktion wie Übersetzen, Bildkreation ist anzugeben.

  Im Text steht: … (“Describe the symbolism”) …

  --> Im Literaturverzeichnis wird der Prompt angegeben: “Describe the symbolism of the green light in the book The Great Gatsby by F. Scott Fitzgerald” prompt. ChatGPT, 13 Feb. version, OpenAI, 8 Mar. 2023, chat.openai.com/chat.

5. April 2024

PS: Dieser Beitrag ist von einer natürlichen Person geschrieben. :)

[1] https://apastyle.apa.org/blog/how-to-cite-chatgpt

[2] https://style.mla.org/citing-generative-ai/

Mittelmäßig?

Ich stoppe vorerst das Besprechen wissenschaftlicher Gütekriterien zur Bewertung der KI.

Es ist ein wissenschaftstheoretisches und sperriges Gebiet, für einen Blog wenig geeignet. Später mehr. :-)

Die Einflussnahme der KI auf zahlreiche gesellschaftliche Bereiche ist unübersehbar und wächst stetig. Es dürfte noch eine Weile dauern, bis wir vollständig erfassen, welche Rolle KI in unterschiedlichen Sektoren spielen wird und welchen Nutzen sie bringt. Schon jetzt zeichnen sich Vorteile ab, etwa die Fähigkeit der KI, Texte geschmeidig und ansprechend zu gestalten, ihnen eine gewisse Leichtigkeit und Abrundung zu verleihen.

Aber diese Leichtigkeit fühlt sich nett an, fast wie Small-talk. In wissenschaftlichen Disziplinen soll jedoch nicht wolkig, sondern differenziert, präzise und auf den Punkt formuliert werden. Danach schaut es derzeit allerdings nicht aus.

Weitere Nachteile werden hervortreten. So wird KI mit der Zeit durchschnittlicher werden, wie Expertinnen vermuten[1]. KI erspart Suche und fasst zusammen. Das alles gemäß Algorithmen, die auf Wahrscheinlichkeiten aufbauen. Was kann da herauskommen? Das Wahrscheinliche, also das Durchschnittliche.

KI ist also eine Zusammenfassungs-Maschine. Kreativität wird sich nur aus einer neuen Verbindung von bisher unterschiedlichen Aspekten manifestieren. Die Basis hierfür wird sukzessive kleiner, da die KI eher früher als später das ganze Netz abgegrast haben wird.

Wobei als "ganzes Netz" auch wirklich alles online Zugängliche gemeint ist. Dazu gehören auch Inhalte nichtwissenschaftlicher Seiten wie solche, auf denen Verschwörungstheorien propagiert werden. Da dies momentan viele Leute klicken, lesen und positiv kommentieren, es also Gewicht hat, werden die KI-Algorithmen dies aufnehmen und uns vor die Nase halten.

6. März 2024

[1] Meckel, M. & Steinacker, L. (2024). *Alles überall auf einmal: Wie Künstliche Intelligenz unsere Welt verändern und was wir dabei gewinnen können*. Hamburg: Rowohlt.

Buntes Allerlei

Äh,
was Strenge bringen?

Wissen verweist auf jene kostbaren Erkenntnisse, die den Kern des wissenschaftlichen Forschungsprozesses bilden. Doch wie steht es um die Qualität dieser Prozesse? Wie können wir sicher sein, dass die erzielten Ergebnisse verlässlich sind?

--> Hier kommen wissenschaftliche Gütekriterien ins Spiel. Sie sind entscheidend dafür, dass Zuverlässigkeit und Sicherheit unseres Wissens bewertet werden können.

An der Spitze dieser Kriterien steht die Forderung nach methodischer Strenge[1]. Der Begriff Methodische Strenge lässt bereits ahnen, dass wir "Bröseln", wie man im Wienerischen sagt, also Ärger bei der Anwendung von KI haben werden. Dieser Begriff zeigt, dass uns im KI-Zeitalter Schwierigkeiten erwarten. Doch was genau gewinnen wir durch diese Methodische Strenge?

--> Mehr als nur Mühe und Schweiß: Vertrauenswürdigkeit. Und die zählt!

In Bereichen, die unser unmittelbares Wohlergehen betreffen – wie Medizin, Recht und Finanzen – ist es offensichtlich, dass wir auf fundierte und zuverlässige Erkenntnisse angewiesen sind. Wir verlassen uns darauf, dass wir Zugang zu den besten verfügbaren Ratschlägen und Behandlungen haben, wenn es um unsere Gesundheit und unser Glück geht.

Bei KI jedoch ist diese Vertrauenswürdigkeit nicht selbstverständlich. Das steht in starkem Kontrast zum wissenschaftlichen Forschungsprozess, bei dem Zuverlässigkeit und Qualität der Quellen im Mittelpunkt stehen.

Die derzeitigen Angebote im KI-Bereich sind größtenteils von Marktinteressen getrieben, mit dem Ziel, Nutzerzahlen und somit Umsätze zu maximieren. Diese Geschäftsstrategien folgen oft der Devise "Mehr ist besser". So meinten irrigerweise hunderte Millionen Menschen bis vor einem halben Jahrtausend, die Erde sei eine Scheibe.

Wissenschaft hat mit der schieren Menge an Meinungen und Glaubensvorstellungen nichts zu tun. Im Gegenteil, Wissenschaft bemisst ihren Wert nicht an der Anzahl der Anhänger, sondern an der überprüfbaren Qualität der Erkenntnisse.

Eine Veränderung findet dann statt, wenn KI-Anbieter beginnen, den wissenschaftlichen Markt ernst zu nehmen. Indem sie wissenschaftliche Institutionen, Akademiker und deren Forschungsprozesse als Zielgruppe auffassen, könnten sie einen wertvollen Beitrag zur Förderung von zuverlässigem und fundiertem Wissen leisten. Erste Anzeichen sind erkennbar.

24. Februar 2024

[1] Gemäß der wissenschaftstheoretischen Tradition Campbells sollen die angewendeten Methoden zum Forschungsproblem passen und regelkonform umgesetzt werden. Döring, N. (2023). *Forschungsmethoden und Evaluation in den Sozial- und Humanwissenschaften*. 6., vollständig überarbeitete, aktualisierte und erweiterte Auflage. S. 89

Das Wiener Rathaus an einem Wintermorgen in 2023/24

Wie steht's um die Qualität?

ChatGPT ist ein Generalist –

für vieles sehr brauchbar, etwa für spezifische Anwendungen im Bildungs- oder Gesundheitsbereich.

Für vieles, etwa im Berufsleben, kann GPT eventuell sogar gefährlich werden. Man denke nur an erfundene (halluzinierte) Fälle in der US-amerikanischen Rechtssprechung.

Für alle Antworten von ChatGPT gilt das Warndreieck auf der Straße: Achtung Gefahr voraus!

Denn die angebotenen Informationen wirken zwar vertrauenswürdig, sind es aber nicht. Die KI-Generalisten sind eben eine black box – die Anwender*innen wissen nicht, wie die Antworten entstanden sind, auf welchen Quellen sie beruhen und ob sie tatsächlich das letztgültig bewährte Wissen repräsentieren.

Für wissenschaftliche Zwecke sind die universalen Bots in dieser Form unbrauchbar. Die textgenerative KI wird sowohl kreative als auch seriös forschende Wissenschaftler nicht ersetzen können.

Zu erwarten ist daher, dass für wissenschaftliche Zwecke spezifische KI auf den Markt kommen wird – erste Ansätze sind zu erkennen. So lassen sich pdf-Dateien durchsuchen und auf ihre wichtigen Erkenntnisse reduzieren.

Um diese neuen wissenschaftsspezifischen Bots einschätzen und beurteilen zu können, bedarf es bestimmter Kriterien. Dies ist allerdings nichts Neues. Die Wissenschaftstheorie bietet einen Fundus an in den letzten Generationen entwickelten Gütekriterien.

Denn im Grunde weiß man beim Lesen eines Artikels und einer Studie nie, ob es sich dabei um gute Wissenschaft handelt oder zum Zwecke des ‚publish or perish‘ verfasst wurde.

Gute Wissenschaft zeigt sich oft erst nach Jahren. Ein berühmtes Beispiel hierfür ist Albert Einstein, der selbst nach fünf Veröffentlichungen in 1905, seinem "Wunderjahr" weitere vier Jahre technischer Experte 3. Klasse im Berner Patentamt blieb. Vier dieser Veröffentlichungen waren nach Weizsäcker[1] nobelpreiswürdig. Erkannt hat man das erst viele Jahre später.

Behelfsmäßig springen hier Gütekriterien ein – sie sind Indizien für wissenschaftliche Qualität. Hierbei geht es um Wege der Evidenz, um Überprüf- und Nachvollziehbarkeit. Dies, um Inhalte verstehen und erklären zu können.

Die grundlegenden Forschungsansätze werden grob in drei Kategorien eingeteilt: *Quantitative, qualitative und Mixed-Method-Verfahren*. Die letzten Jahrzehnte haben gezeigt, dass Gütekriterien für eine Richtung nicht unbedingt wissenschaftliche Qualität für die andere Richtung verbürgen. Zu unterschiedlich sind Ziele, Forschungswege und Forschungsmethoden.

Qualität in einer Kategorie kann Qualität in der anderen Kategorie mindern.

Wir stehen im Grunde bei der Beurteilung der Qualität des KI-Outputs vor demselben Problem wie bei der Beurteilung von wissenschaftlichen Studien. Warum daher nicht die mühsam über Forschergenerationen erarbeiteten Gütekriterien auch für KI anwenden? Damit käme XAI ins Spiel.

29. Jänner 2024

[1] Weizsäcker, C. (1992). *Das Carl Friedrich von Weizsäcker Lesebuch.* München: DTB, S. 46

Qualität fürs Auge und den Gaumen

Wikipedia,

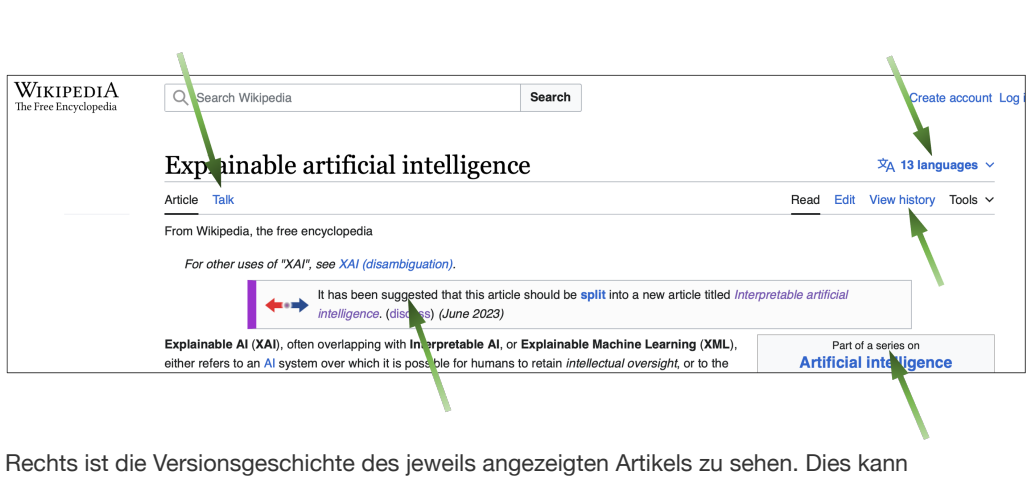
die seit fast einem Vierteljahrhundert im Netz abrufbare Enzyklopädie,

bietet zumeist gehaltvolle und seriös abgesicherte Information. Sie verweist bei aufgerufenen Begriffen auf Quellen, differierende Definitionen oder mögliche angrenzende Bereiche. Die angebotenen Inhalte werden diskutiert und – lobenswerterweise – auch in Frage gestellt.

Um dies am Beispiel XAI zu zeigen:

Der auf der Hauptseite des aufgesuchten Begriffes erscheinende Artikel wird argumentativ vielfach belegt. Etwas, das die meisten Leser*innen leider übersehen. Der zweite Reiter neben *Article* ist *Talk* (in der deutschen Ausgabe: *Diskussion*). Hier finden Sie die Anmerkungen anderer Redakteure und Administratoren von Wikipedia.

Die Anmerkungen beziehen sich auf den Stil, die grundsätzliche Richtung, gesetzte Schwerpunkte oder mögliche Verzerrungen. Veränderungsvorschläge werden skizziert. Am angezeigten Beispiel etwa wird angeregt, den Artikel zu XAI in Explicable AI und Interpretable AI aufzuteilen.



Explainable artificial intelligence

From Wikipedia, the free encyclopedia

For other uses of "XAI", see XAI (disambiguation).

It has been suggested that this article should be **split** into a new article titled *Interpretable artificial intelligence*. (discuss) *(June 2023)*

Explainable AI (XAI), often overlapping with **Interpretable AI**, or **Explainable Machine Learning (XML)**, either refers to an AI system over which it is possible for humans to retain *intellectual oversight*, or to the

Part of a series on
Artificial intelligence

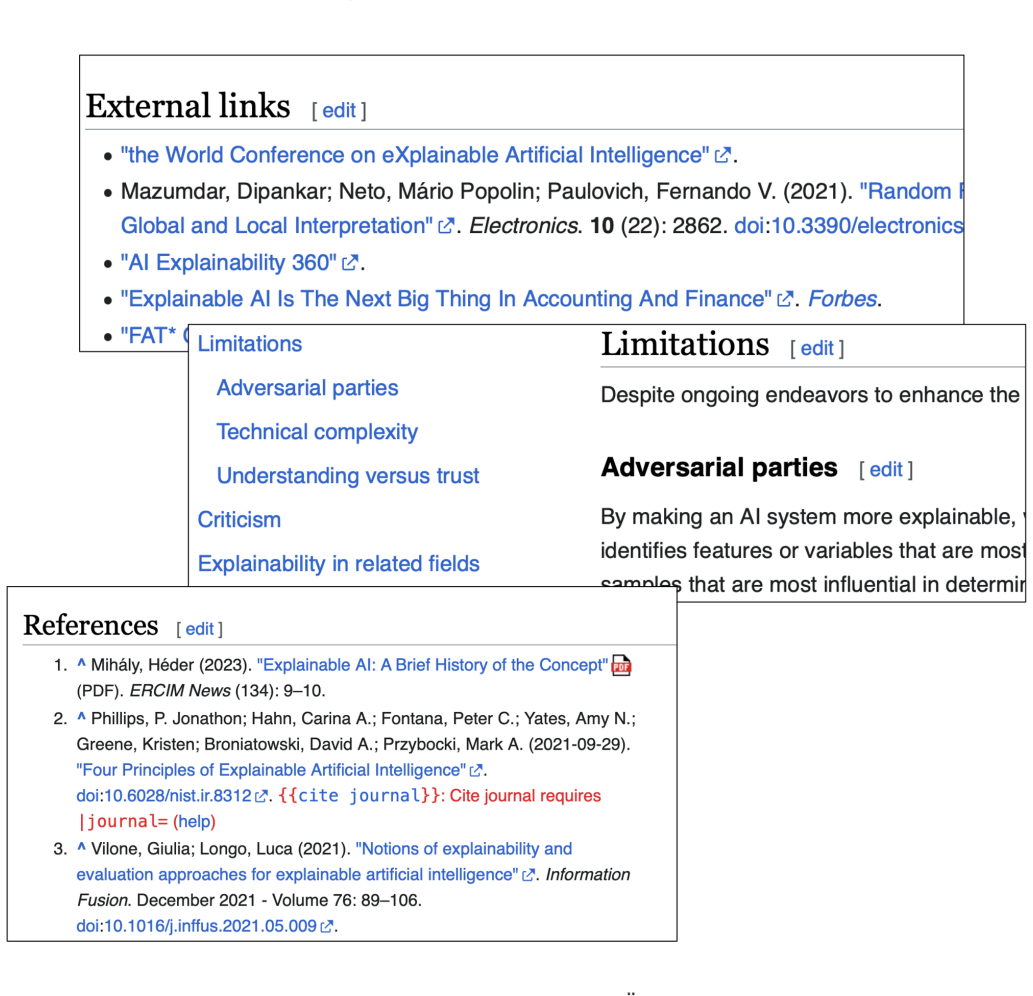
Rechts ist die Versionsgeschichte des jeweils angezeigten Artikels zu sehen. Dies kann aussagekräftig sein, denn die Zeitstempel bieten einiges an Meta-Information besonderer Art: etwa für das Auftreten von Modewellen oder eben ein Schweigen zur relevanten Thematik. Ein weiterer Reiter bietet einen Link zu anderen Sprachen. Dies wiederum kann möglicherweise ganz andere Strukturen und Sichtweisen erkennbar machen.

Denn Sprachen spiegeln instrumental nur äußere Merkmale unterschiedlicher Denk- und Weltauffassungen wider. Im Home-office kann so ein bisschen kognitiv-ethnologische Forschung betrieben werden: Welche Kultur legt welchen Fokus auf welche Dimensionen?

Im angeführten Beispiel zu XAI wird auch eine Artikelserie zum übergreifenden Begriff „Artificial Intelligence“ innerhalb Wikipedia angezeigt. Ebenso werden Wiki-externe Links angeführt.

Im Kapitel Limitationen werden zentrale Aspekte für den Geltungsbereich und Bemerkungen zu gegensätzlichen Auffassungen/Meinungen angeführt. Etwas, das für akademische Arbeiten von grundsätzlicher Bedeutung ist und in seriöser akademischer Literatur entsprechend eingefordert wird.

Selbstverständlich gibt es Quellenverweise, also Referenzen. Diese können manchmal älteren Datums sein – je nach Entstehungszeit des Hauptartikels und seiner Überarbeitungen. Der Wiki-Artikel zu XAI führt 87 Quellen an, die bis 2023 reichen.



External links [edit]

- "the World Conference on eXplainable Artificial Intelligence".
- Mazumdar, Dipankar; Neto, Mário Popolin; Paulovich, Fernando V. (2021). "Random Global and Local Interpretation". *Electronics*. **10** (22): 2862. doi:10.3390/electronics
- "AI Explainability 360".
- "Explainable AI Is The Next Big Thing In Accounting And Finance". *Forbes*.
- "FAT*

Limitations
Adversarial parties
Technical complexity
Understanding versus trust
Criticism
Explainability in related fields

Limitations [edit]

Despite ongoing endeavors to enhance the

Adversarial parties [edit]

By making an AI system more explainable,
identifies features or variables that are most
samples that are most influential in determi

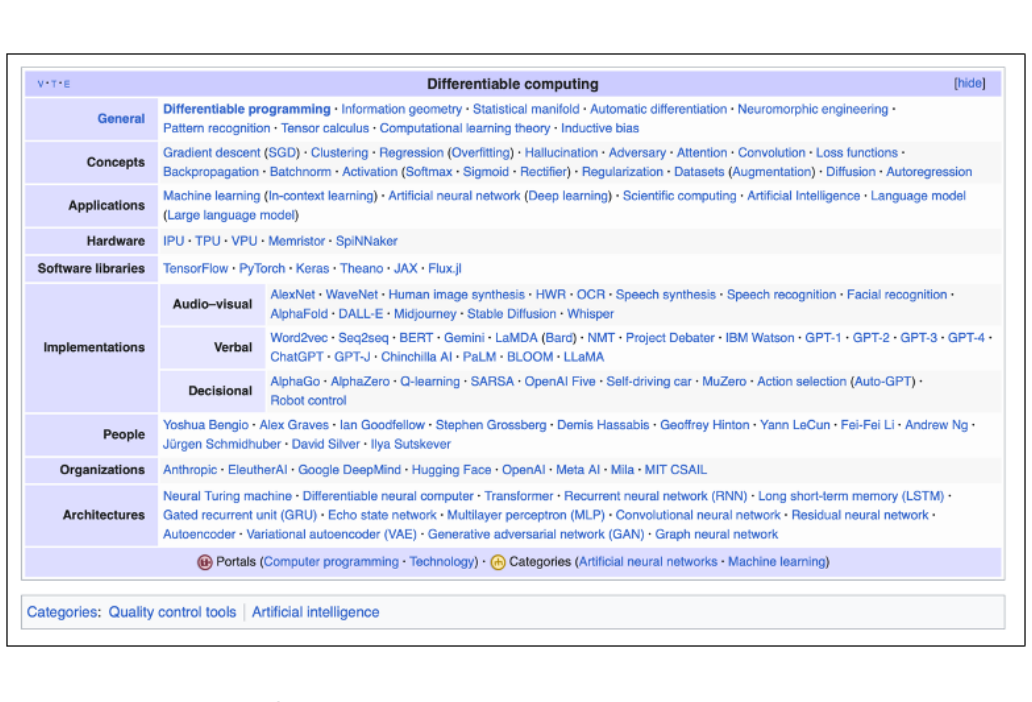
References [edit]

1. ^ Mihály, Héder (2023). "Explainable AI: A Brief History of the Concept" (PDF). *ERCIM News* (134): 9–10.
2. ^ Phillips, P. Jonathon; Hahn, Carina A.; Fontana, Peter C.; Yates, Amy N.; Greene, Kristen; Broniatowski, David A.; Przybocki, Mark A. (2021-09-29). "Four Principles of Explainable Artificial Intelligence". doi:10.6028/nist.ir.8312. {{cite journal}}: Cite journal requires |journal= (help)
3. ^ Vilone, Giulia; Longo, Luca (2021). "Notions of explainability and evaluation approaches for explainable artificial intelligence". *Information Fusion*. December 2021 - Volume 76: 89–106. doi:10.1016/j.inffus.2021.05.009.

Als Draufgabe wird im XAI-Artikel eine zusammenfassende Übersicht zu Konzepten und verwendeten Algorithmen geboten: Welche Unternehmen haben mittels welcher Architekturen AI bereits implementiert.

Liebe Leser*in, achte auf die letzte Zeile der folgenden Übersicht, insbesondere auf den Link zu „Quality control tools“. Die Inhalte dieses Links sind interessant, aber man kann sich in den darin angebotenen Tools verlieren, wenn man ihre Konzepte und Anwendungsmöglichkeiten nicht kennt.

Doch werden hier erste statistische Kontrollwerkzeuge für KI, mögen sie geeignet sein oder auch nicht, angesprochen.



Differentiable computing		
General		Differentiable programming · Information geometry · Statistical manifold · Automatic differentiation · Neuromorphic engineering · Pattern recognition · Tensor calculus · Computational learning theory · Inductive bias
Concepts		Gradient descent (SGD) · Clustering · Regression (Overfitting) · Hallucination · Adversary · Attention · Convolution · Loss functions · Backpropagation · Batchnorm · Activation (Softmax · Sigmoid · Rectifier) · Regularization · Datasets (Augmentation) · Diffusion · Autoregression
Applications		Machine learning (In-context learning) · Artificial neural network (Deep learning) · Scientific computing · Artificial Intelligence · Language model (Large language model)
Hardware		IPU · TPU · VPU · Memristor · SpiNNaker
Software libraries		TensorFlow · PyTorch · Keras · Theano · JAX · Flux.jl
Implementations	Audio–visual	AlexNet · WaveNet · Human image synthesis · HWR · OCR · Speech synthesis · Speech recognition · Facial recognition · AlphaFold · DALL-E · Midjourney · Stable Diffusion · Whisper
	Verbal	Word2vec · Seq2seq · BERT · Gemini · LaMDA (Bard) · NMT · Project Debater · IBM Watson · GPT-1 · GPT-2 · GPT-3 · GPT-4 · ChatGPT · GPT-J · Chinchilla AI · PaLM · BLOOM · LLaMA
	Decisional	AlphaGo · AlphaZero · Q-learning · SARSA · OpenAI Five · Self-driving car · MuZero · Action selection (Auto-GPT) · Robot control
People		Yoshua Bengio · Alex Graves · Ian Goodfellow · Stephen Grossberg · Demis Hassabis · Geoffrey Hinton · Yann LeCun · Fei-Fei Li · Andrew Ng · Jürgen Schmidhuber · David Silver · Ilya Sutskever
Organizations		Anthropic · EleutherAI · Google DeepMind · Hugging Face · OpenAI · Meta AI · Mila · MIT CSAIL
Architectures		Neural Turing machine · Differentiable neural computer · Transformer · Recurrent neural network (RNN) · Long short-term memory (LSTM) · Gated recurrent unit (GRU) · Echo state network · Multilayer perceptron (MLP) · Convolutional neural network · Residual neural network · Autoencoder · Variational autoencoder (VAE) · Generative adversarial network (GAN) · Graph neural network

Portals (Computer programming · Technology) · Categories (Artificial neural networks · Machine learning)

Categories: Quality control tools | Artificial intelligence

Fazit: Aus akademischer Sicht bietet Wikipedia auf effiziente und zeitsparende Weise vieles, was eine Basis für weitere seriöse Analyse sein kann: Zentrale Aspekte von Alternativ-Definitionen, begriffliche Einordnungen, Herkünfte, Begrenzungen, kritische Überlegungen etc.

Die zukünftige XAI ist im etablierten Wikipedia ansatzweise verwirklicht.

13. Jänner 2024

PS: Die Unis haben eine große Chance vertan. Seit zwei Jahrzehnten heißt es rigoros: Zitate aus Wikipedia sind verboten!

Wären konsequent Unis Teil-Kooperationen mit dem Wiki-Projekt eingegangen und hätten Wiki-Inhalte mit dem akademischen Erforschen verknüpft,

stünden die Unis jetzt besser da und nicht so rat- und hilflos vor dem überrollenden Tsunami von geglätteten Texten aus dem KI-Algorithmus, die "Nicht-Studierende" einfach abrufen.

Und Mose fordert den Herrn auf:

„Lass mich doch deine Herrlichkeit sehen!“.

„Du kannst es nicht ertragen, mein Angesicht zu sehen, denn kein Mensch kann mich sehen und am Leben bleiben.“ (Kapitel 33, Verse 18 bis 21)[1]. Diese Antwort ist wohl eindeutig.

Einige Autoren des 20. Jahrhunderts wie Sigmund Freud[2] oder Immanuel Velikovsky[3] deuten darin die Herkunft des altisraelitischen Gottes Jahwe als Vulkangott (siehe auch die Erzählung vom brennenden Dornbusch) oder als eine Übertragung von Aton, dem ägyptischen Gott der Sonnenscheibe.

Diese spekulativen Thesen konnten bisher nicht durch archäologische Beweise belegt werden und gelten als randständig. Eine Prima-facies-Plausibilität kann dieser Interpretation jedoch zugesprochen werden: Ein zu langes Sehen in die Sonne oder in einen Vulkankrater mit glühendem Magma lässt uns erblinden.

Im Buch Exodus folgen nun Textpassagen, die global populär wurden:

„Darauf sprach der Herr zu Mose: Haue dir zwei steinerne Tafeln […] und steige am Morgen auf den Berg Sinai […]“ (34:1-2) Mose blieb dort vierzig Tage und vierzig Nächte, aß weder Brot noch trank er Wasser.

„Und er schrieb auf die Tafeln die Worte des Bundes, die zehn Worte [Gesetze]“(34:28)

Diese alttestamentarische Erzählung stellt in seiner Gesamtheit eine Metapher zur gegenwärtigen textgenerierenden KI dar. Wir fragen –> erhalten Antworten –> sehen aber nicht, wie sie entstanden sind.

Vielleicht würde es uns an einigen Stellen schaudern, wenn auch nicht gleich erblinden lassen. KI ist für uns Sterbliche eine black box. Die Antworten wirken plausibel, als seien sie von einem wissenden Menschen verfasst. Woraufhin wir uns freuen und die Antworten ähnlich schätzen, wie es im Buch Exodus im folgenden Kapitel heißt:

Als Mose vom Berg Sinai herabstieg, war sein Gesicht strahlend geworden und alle fürchteten sich.

„Danach aber traten alle Söhne Israel heran, und er gebot ihnen alles, was der HERR auf dem Berg Sinai geredet hatte.“ (34:32-33).

Auch uns gegenwärtigen Töchtern und Söhnen ergeht es so. KI-Antworten wirken logisch aufgebaut und weisen einen beträchtlichen Grad an Plausibilität auf. Nebenbei gefällt uns, dass KI uns Suchenden viel Arbeit abgenommen hat. Daher nehmen wir gerne die KI-Texte für bare Münze.

Jedoch ist da was geschehen. Unbemerkt haben wir den Empfangsort gewechselt: Von der realen oder virtuellen universitären Bibliotheksumgebung in einen Kirchenraum.

Von mühsam Suchenden sind wir naiv alles schluckende Gläubige einer modernen Gottheit geworden. Hatten wir doch gerade von der KI-Kanzel die Wahrheit gehört.

Um es weniger zugespitzt zu sagen: ChatGPT ist als großes Sprachmodell ein *Gesellschaftsmodell* und kein Wissenschaftsmodell. KI-Bots bieten Informationen, die auf Unmengen von Daten zugreifen – gespeist aus unzähligen unterschiedlichen, zum großen Teil inkompatiblen Quellen.

Wir wissen nicht, wie und warum die KI gerade diese Informationen ausgespuckt hat.

Aber zu verstehen, wie Ergebnisse entstanden sind, ist das zentrale Fundament in den Wissenschaften. Die angewendeten Methoden der Datengenerierung und Datenanalyse sind das Um und Auf interpretierbarer Ergebnisse.

Daraus werden Schlussfolgerungen gezogen. Erst dann kann bis zu einem gewissen Grad den Aussagen vertraut werden. Nicht Glaube, sondern Kritik ist ein wesentliches Element von Wissenschaften.

Kurz,
das wissenschaftliche Gütekriterium Transparenz ist nicht einmal ansatzweise erfüllt. Den Anwendern präsentieren die KI-Chat-Bots nur eine Eingabe- und eine Ausgabemaske. Dazwischen gibt es .... nichts.

Hier kann man sich symbolisch eine black box vorstellen. Man weiß nicht, in welchem Rahmen gesucht wird, was die Algorithmen der Datenaufbereitung tun und welche Auswahlkriterien schlagend werden.

Einen Quantensprung könnte uns die weitere Entwicklung von Artificial Intelligence zu einer XAI, der eXplainable Artificial Intelligence, bieten. XAI, die erklärende künstliche Intelligenz, hat den Anspruch, verstehbare Ergebnisse zu liefern[2]. Wodurch die gegenwärtig unsichtbare black box zu einer white box, also einer durchsichtigen gläsernen, wird.

Anders als Mose könnten wir sehen, am Leben bleiben und überdies unser Augenlicht behalten.

7. Jänner 2024

[1] Elberfelder Studienbibel mit Sprachschlüssel und Handkonkordanz. 6. Auflage 2017, Witten: SCM

[2] Freud, Sigmund (1939/2022). *Der Mann Moses und die monotheistische Religion.* Amsterdam: Allert de Lange

[3] Velikovsky, Immanuel (1953/2005). *Erde in Aufruhr.* Julia White

[4] David Gunning, Mark Stefik, Jaesik Choi, Timothy Miller, Simone Stumpf, Guang-Zhong Yang (2019). *XAI — Explainable artificial intelligence. Explainability is essential for users to effectively understand, trust, and manage powerful artificial intelligence applications.* Science Robotics. 4, eaay7120

Nächtlicher Kanal in Venedig als Sinnbild für Licht in der Dunkelheit :)

Zugriffe: 63706

Comments (0)

There are no comments posted here yet

Leave your comments

Posting as Guest

Name (Required)

Email

Write your comment here...

Submit Comment

© 2026 reinhard-neumeier.at