Kognitionstheorien und KI-Systeme: Theoretische Grundlagen für die pädagogische Praxis

...

Anna: Hallo und herzlich willkommen. Ich bin Anna, Medienpädagogin und Gastgeberin dieser Folge.

Ben: Und ich bin Ben, Kognitionspsychologe und Informatiker. Heute widmen wir uns einem spannenden Thema, das sowohl mit menschlicher als auch mit künstlicher Intelligenz zu tun hat.

Anna: Genau, Ben. Wir sprechen heute über Theorien des menschlichen Denkens und deren Verbindung zu KI-Systemen wie ChatGPT und anderen Large Language Models.

Ben: Das klingt vielleicht zunächst sehr theoretisch, aber wir versprechen euch: Diese Grundlagen werden euch helfen, besser zu verstehen, wie diese KI-Systeme funktionieren, wo ihre Stärken und Schwächen liegen - und wie ihr sie sinnvoll im Unterricht einsetzen könnt.

Anna: Als Lehrkräfte seid ihr mittlerweile alle mit der Situation konfrontiert, dass eure Schülerinnen und Schüler KI-Tools verwenden. Und viele von euch nutzen sie vielleicht auch selbst schon.

Ben: Aber um diese Tools pädagogisch sinnvoll einzusetzen, ist es wichtig zu verstehen, was eigentlich "unter der Haube" passiert. Wie "denken" diese Systeme? Und was hat das mit menschlichem Denken zu tun?

Anna: Wir werden heute acht einflussreiche Theorien zum menschlichen Denken vorstellen und jeweils den Bogen zu aktuellen KI-Systemen spannen. Keine Sorge, wir halten das praxisnah und verständlich!

Warum ist die Auseinandersetzung mit Kognitionstheorien für Lehrkräfte wichtig?

Anna: Bevor wir in die einzelnen Theorien einsteigen: Warum ist das Verständnis dieser Theorien für uns als Lehrkräfte überhaupt wichtig, Ben?

Ben: Wer verstehen will, warum heutige KI-Systeme bestimmte Dinge gut können und an anderen scheitern, muss wissen, welche Annahmen über menschliches Denken in ihnen stecken. Viele Begriffe, die wir heute in KI-Architekturen finden - wie "Symbol", "Netzwerk" oder "Aufmerksamkeit" - sind aus Theorien über den menschlichen Geist und das Gehirn entlehnt. Manchmal steckt dahinter eine echte Verwandtschaft, manchmal nur ein anschauliches Bild.

Anna: Und dieses Wissen hilft uns, realistischere Erwartungen an KI-Systeme zu entwickeln. Wir können besser einschätzen, wofür wir sie im Unterricht einsetzen können und wo sie an Grenzen stoßen.

Welche Kognitionstheorien gibt es?

Ben: Lasst uns mit der ersten Theorie beginnen: der klassischen symbolischen Theorie des Geistes, auch bekannt als klassische "Computational Theory of Mind".

Anna: Die Grundidee dieser Theorie ist recht einfach: Denken ist wie das Ausführen eines Computerprogramms. Unser Geist manipuliert Symbole nach formalen Regeln.

Ben: Für unsere Schülerinnen und Schüler könnte man es so erklären: Das Gehirn arbeitet wie ein sehr komplizierter Taschenrechner. Es gibt klare Symbole wie "7" oder "+" und Regeln, wie man damit umgeht.

Anna: In der KI-Geschichte finden wir diese Denkweise in frühen Expertensystemen. Aber auch heute noch gibt es Anwendungen: Wissensgraphen oder logikbasierte Planungssysteme funktionieren nach diesem Prinzip.

Ben: Ein Unterrichtsbeispiel wäre ein Programm, das Schülern Mathematik beibringt, indem es klare Regeln anwendet: "Wenn die Aufgabe so aussieht, wende diese Formel an."

Anna: Die zweite Theorie ist der Konnektionismus oder die Neurale-Netzwerk-Theorie. Hier wird es für das Verständnis moderner KI richtig wichtig.

Ben: Absolut! Im Gegensatz zum symbolischen Ansatz sagt der Konnektionismus: Wissen steckt nicht in einzelnen Symbolen, sondern in den Verbindungsstärken zwischen vielen künstlichen Neuronen. Ein Gedanke ist dann ein Muster verteilter Aktivität.

Anna: Für die Schüler könnte man es so erklären: Stellt euch eine riesige Anzeigetafel mit tausenden kleinen Lämpchen vor. Kein einzelnes Lämpchen steht für "Hund", sondern erst das gesamte Muster, das aufleuchtet, repräsentiert den Gedanken an einen Hund.

Ben: Und genau so funktionieren moderne Large Language Models wie ChatGPT. Sie sind im Kern neuronale Netzwerke mit Milliarden von "Gewichten" - Verbindungen zwischen künstlichen Neuronen.

Anna: Ein praktisches Beispiel für den Unterricht: Wenn Schüler mit einem KI-Bildgenerator arbeiten, erzeugt die KI Bilder nicht, indem sie Symbole wie "Auge" oder "Nase" nach Regeln zusammensetzt, sondern indem sie Muster aktiviert, die sie aus Hunderten Millionen bis Milliarden von Bildern gelernt hat.

Ben: Das erklärt auch, warum diese Systeme manchmal seltsame Fehler machen, etwa wenn frühere Versionen von DALL-E Menschen mit sechs Fingern zeichneten. Sie hatten kein symbolisches Verständnis von "fünf Finger pro Hand", sondern nur statistische Muster. Neuere Bildgeneratoren machen diesen Fehler übrigens deutlich seltener.

Anna: Das bringt uns zur dritten Theorie: Embodied oder Grounded Cognition - verkörpertes Denken. Diese Theorie betont, dass unser Denken untrennbar mit unserem Körper und unserer Umwelt verbunden ist.

Ben: Die Kernaussage ist: Wir denken anders, weil wir gehen, greifen und mit unseren Sinnen wahrnehmen können. Unsere Konzepte sind in körperlichen Erfahrungen "verankert" oder "geerdet".

Anna: Für Schüler könnte man es so ausdrücken: Unser Verständnis von "werfen" ist eng mit eigenen Erfahrungen verknüpft - mit der Muskelanspannung, der Bewegung und dem Gefühl, wenn wir selbst etwas werfen.

Ben: Wobei man fairerweise sagen muss: Wie stark diese Verankerung ist, ist in der Forschung umstritten. Auch Menschen, die bestimmte Bewegungen nie selbst ausführen konnten, verstehen sie. Trotzdem sehen wir hier einen wichtigen Unterschied zu heutigen KI-Systemen wie ChatGPT: Sie haben keine körperlichen Erfahrungen. Sie wurden mit Texten und inzwischen auch mit Bildern und Tonaufnahmen trainiert, aber nicht durch körperliches Handeln in der Welt. Daher fehlt ihnen dieses "Grounding".

Anna: Das hat Implikationen für den Unterricht: Wenn ein Schüler ChatGPT nach Erklärungen für physikalische Phänomene fragt, kann es theoretisch korrekte Antworten geben, aber es hat nie selbst einen Ball fallen lassen oder einen Magneten in der Hand gehalten.

Ben: Einen Teil dieser Lücke versucht man schon zu schließen: Viele aktuelle Modelle verarbeiten neben Sprache auch Bilder und Ton - man spricht von "multimodalen" Modellen. Und in der Robotik-Forschung verbindet man Sprachmodelle sogar mit den Bewegungserfahrungen von Robotern - das nennt man "Embodied AI".

Anna: Kommen wir zur vierten Theorie, dem Predictive Coding, das eng mit dem umfassenderen Free-Energy-Prinzip verbunden ist.

Ben: Diese Theorie besagt, dass unser Gehirn im Wesentlichen eine Vorhersagemaschine ist. Es erstellt ständig Vorhersagen darüber, welche Sinneseindrücke als nächstes auftreten werden, und versucht, den Unterschied zwischen Vorhersage und tatsächlicher Wahrnehmung zu minimieren.

Anna: Stell dir vor, dein Gehirn schreibt ständig "Wettervorhersagen" darüber, was als nächstes passieren wird, und korrigiert seine Modelle, wenn die Realität anders aussieht.

Ben: Und hier wird es für Lehrkräfte, die ihren Lernenden die Funktionsweise von generativen KI-Systemen nahe bringen wollen, besonders interessant: Eine verwandte Idee steckt in Large Language Models wie GPT. Sie wurden darauf trainiert, das jeweils nächste Textstück vorherzusagen, und lernen aus ihren Vorhersagefehlern. Aber Vorsicht: Das passiert nur während des Trainings. Wenn ihr mit ChatGPT chattet, lernt das Modell nicht mehr dazu - anders als unser Gehirn, das seine Vorhersagen ständig anpasst. Außerdem wird ChatGPT nach diesem Training noch mit menschlichem Feedback nachtrainiert, damit es hilfreiche Antworten gibt.

Anna: Ein praktisches Beispiel: Wenn ein Schüler ChatGPT bittet, einen Aufsatz zu schreiben, erzeugt das System den Text Stück für Stück - genauer gesagt in sogenannten Tokens, das sind oft Wortteile. Bei jedem Schritt berechnet es, welche Fortsetzungen wahrscheinlich sind, und wählt mit etwas Zufall eine davon aus. Trotzdem schreibt es nicht völlig planlos drauflos: Forschende konnten zeigen, dass solche Modelle intern vorausplanen und zum Beispiel beim Reimen das Reimwort schon festlegen, bevor die Zeile beginnt. Nur eben anders als ein Mensch, der vorher eine Gliederung schreibt.

Ben: Dass KI-generierte Texte trotzdem manchmal inkonsistent werden oder vom ursprünglichen Thema abweichen, hat eher technische Gründe: Das Modell kann nur eine begrenzte Textmenge auf einmal berücksichtigen, und kleine Fehler können sich im eigenen Text aufschaukeln. Bei aktuellen Modellen passiert das allerdings deutlich seltener.

Anna: Die fünfte Theorie ist die Global Workspace Theory. Sie versucht zu erklären, wie Bewusstsein entsteht.

Ben: Die Grundidee: Bewusstsein entsteht, wenn Informationen ein globales "Schwarzes Brett" im Gehirn erreichen, auf das alle Hirnmodule zugreifen können. Man könnte es mit einem Firmen-Intranet vergleichen - wenn dort eine Information erscheint, können alle Abteilungen darauf reagieren.

Anna: Für Schüler: Stellt euch vor, nur die wichtigsten Informationen schaffen es auf eine zentrale Anzeigetafel in eurem Kopf. Das sind die Dinge, die ihr bewusst wahrnehmt.

Ben: In modernen KI-Systemen, besonders den Transformer-basierten Modellen wie GPT, gibt es einen Mechanismus mit ähnlich klingendem Namen: den Attention-Mechanismus. Er gewichtet, welche Teile eines Textes für die Verarbeitung eines Wortes besonders wichtig sind. Ein globales "Schwarzes Brett" mit begrenztem Platz, um den die Informationen konkurrieren, ist das aber nicht. Ob es in Sprachmodellen etwas Vergleichbares gibt, wird gerade erst erforscht.

Anna: Im Unterrichtskontext könnte man dies so erklären: Wenn ihr mit ChatGPT interagiert, hat es keinen kontinuierlichen "Bewusstseinsstrom" wie wir Menschen. Stattdessen nutzt es Attention-Mechanismen, um die relevantesten Teile eurer Anfrage zu fokussieren.

Ben: Die sechste Theorie ist die Dual-Process-Theorie, besonders bekannt durch Daniel Kahnemans Buch "Schnelles Denken, langsames Denken". Sie besagt, dass wir zwei unterschiedliche Denkmodi haben: System 1 (schnell, automatisch, intuitiv) und System 2 (langsam, bewusst, analytisch).

Anna: Ein alltagsnahes Beispiel: Wenn ich dich frage, was 2+2 ist, antwortest du sofort "4" - das ist System 1. Wenn ich dich frage, was 17 × 24 ist, musst du nachdenken und rechnen - das ist System 2.

Ben: Diese Theorie ist besonders interessant für die Entwicklung fortschrittlicher KI-Systeme. Ein Sprachmodell, das sofort eine Antwort ausgibt, ähnelt eher "System 1". Schon 2022, noch vor dem Start von ChatGPT, entdeckten Forschende das sogenannte "Chain-of-Thought"-Prompting: Fordert man das Modell auf, Zwischenschritte aufzuschreiben, löst es Aufgaben oft besser - eine Art simuliertes "System 2". Neuere sogenannte Reasoning-Modelle sind sogar gezielt darauf trainiert, vor ihrer Antwort von sich aus ausführlich "nachzudenken".

Anna: Ein Unterrichtsbeispiel: Wenn ihr Schüler bittet, ein Chat-Modell für ein komplexes Problem zu verwenden, könnt ihr sie anweisen, die KI aufzufordern, "Schritt für Schritt" vorzugehen. Das Modell schreibt dann Zwischenschritte auf, die ihm als zusätzliche

Rechenschritte dienen, und das führt oft zu besseren Ergebnissen. Bei Reasoning-Modellen ist dieser Tipp allerdings überflüssig, teils sogar hinderlich - die gehen ohnehin schon schrittweise vor.

Ben: Die siebte Theorie ist das Bayesian Brain - die Idee, dass unser Gehirn ähnlich wie ein Statistiker nach dem Bayes-Theorem arbeitet. Es integriert Vorwissen mit neuen Beobachtungen, um zu Schlussfolgerungen zu gelangen. Diese Theorie ist übrigens eng mit dem Predictive Coding verwandt, über das wir eben gesprochen haben.

Anna: Für Schüler könnte man es so erklären: Euer Gehirn arbeitet wie ein Wettbüro. Es hat bestimmte "Quoten" - Erwartungen - und jede neue Information verändert diese Quoten. Wenn ihr einen Hund bellen hört, steigt die Wahrscheinlichkeit, dass ein Hund in der Nähe ist.

Ben: In der KI-Welt findet man bayesianische Ansätze in verschiedenen Methoden, etwa in probabilistischen Modellen oder in Techniken, die die Unsicherheit von KI-Vorhersagen quantifizieren.

Anna: Für den Unterricht bedeutet das: Wenn ihr mit Schülern über die Zuverlässigkeit von KI-Antworten sprecht, könnt ihr erklären, dass KI-Systeme häufig nicht gut darin sind, ihre eigene Unsicherheit einzuschätzen - sie klingen oft sicherer, als sie sind. Das geht uns Menschen übrigens ähnlich: Bei der Wahrnehmung verrechnet unser Gehirn Unsicherheiten erstaunlich gut, aber bei Urteilen und Wissensfragen überschätzen wir uns regelmäßig.

Ben: Und schließlich die achte Theorie: der Dynamische-Systeme-Ansatz. Diese Theorie betrachtet Kognition als kontinuierliche Zustandsänderungen in komplexen, miteinander gekoppelten Systemen.

Anna: Statt von klaren Repräsentationen oder Regeln spricht diese Theorie von "Fluss". Man könnte es so erklären: Gedanken ähneln Wirbelströmen im Wasser - sie entstehen, verändern sich und verschwinden in einem kontinuierlichen Prozess.

Ben: In der KI-Welt finden wir diesen Ansatz in rekurrenten neuronalen Netzwerken - Modellen, die dynamische Zeitverläufe verarbeiten können.

Anna: Im Unterricht kann man damit zeigen, dass Denken auch als fortlaufender Prozess verstanden werden kann, etwa bei der Steuerung von Bewegungen oder Robotern. Wichtig ist aber: ChatGPT und ähnliche Sprachmodelle sind keine rekurrenten Netze, sondern Transformer. Wenn sie sich mal in Wiederholungen verfangen, liegt das vor allem daran, wie sie die nächsten Textstücke auswählen.

Welche Bedeutung haben die vorgestellten Kognitionstheorien für die Entwicklung von KI-Systemen und den Einsatz von KI in der Schule?

Ben: So, nachdem wir alle acht Theorien kennengelernt haben, lass uns diese nun vergleichen und schauen, welche praktischen Schlussfolgerungen wir für den Unterricht ziehen können.

Anna: Die Theorien unterscheiden sich in mehreren Dimensionen: Wie sie Wissen repräsentieren - symbolisch klar oder kontinuierlich verteilt; wie stark sie den Körperbezug betonen; wie sie zeitliche Dynamik modellieren; welche Rolle Wahrscheinlichkeit spielt; und ob sie Bewusstsein explizit erklären.

Ben: Moderne KI-Systeme wie ChatGPT kombinieren tatsächlich Aspekte mehrerer Theorien. Sie sind hauptsächlich konnektionistisch und darauf trainiert, Text vorherzusagen - eine Idee, die mit Predictive Coding verwandt ist. Neuere Entwicklungen integrieren auch symbolische Elemente, etwa wenn ein Chatbot für Rechnungen Programmcode ausführt, und versuchen, Aspekte von Embodiment einzubeziehen. LLMs haben aber keine echte Körperlichkeit, ihre "System 2"-Fähigkeiten sind noch fragil, und die Frage nach "KI-Bewusstsein" ist höchst spekulativ.

Anna: Absolut. Und das führt uns zu einem wichtigen Punkt für alle Lehrkräfte: Diese Technologien entwickeln sich rasant weiter. Was heute als Grenze gilt, könnte morgen schon überwunden sein. Daher ist es wichtig, ein grundlegendes Verständnis dieser Theorien zu haben.

Ben: Genau. Fassen wir zusammen: Die verschiedenen Kognitionstheorien bilden ein Ideen-Arsenal, das die KI-Forschung selektiv nutzt. Moderne KI-Systeme sind am stärksten vom Konnektionismus und vom Prinzip der Vorhersage geprägt, aber integrieren zunehmend Aspekte anderer Theorien.

Anna: Für euch als Lehrkräfte bedeutet das: Experimentiert mit diesen Tools, aber behaltet stets ein kritisches Verständnis ihrer Grundlagen und Grenzen.

Ben: Und denkt daran: Es geht nicht darum, KI zu verteufeln oder zu glorifizieren, sondern sie als Werkzeug zu verstehen, das bestimmte Stärken und Schwächen hat - so wie auch unser eigenes Denken.

...

Anna: Damit sind wir am Ende unserer heutigen Folge angelangt. Wir hoffen, ihr habt einen guten Überblick über die Theorien des menschlichen Denkens und ihre Verbindung zu modernen KI-Systemen bekommen.

Ben: Vielen Dank fürs Zuhören. Bleibt neugierig!