

Generative KI als epistemische Infrastruktur

Vertrauen, Wissen und Autorität im digitalen Zeitalter -
wo Informatik und Philosophie sich treffen

Prof. Dr.-Ing. Stefan Wagenpfeil

PFH Private University of Applied Sciences Göttingen

s.wagenpfeil@pfh.de

Abstract

Generative KI verändert die Bedingungen menschlicher Wissensgewinnung. KI-Assistenten treten zunehmend als Informationsquellen, Argumentationspartner, Orientierungsinstanzen, oder auch emotionale Gefährten auf, ohne selbst im klassischen Sinn “wissende Subjekte” zu sein. In diesem Artikel wird die These entwickelt, dass generative KI weniger als neue Form von Intelligenz, sondern als epistemische Infrastruktur verstanden werden sollte, durch die Wissen erzeugt, geprüft, verarbeitet und genutzt wird. Hierbei wird nicht gefragt, ob KI selbst “wissen” kann, sondern unter welchen Bedingungen menschliches Vertrauen in KI-vermittelte Aussagen epistemisch gerechtfertigt ist. Um dies zu erörtern, verbindet der Beitrag klassische Erkenntnistheorie, Debatten über Zeugenschaft und Vertrauen, Diskussionen zur epistemischen Autorität algorithmischer Systeme mit Erkenntnissen aus der Informatik und KI-Forschung. Er argumentiert, dass generative KI eine Zwischenposition einnimmt: sie ist weder bloß neutrales Werkzeug noch vollwertiger Wissensakteur. Gerade diese Zwischenstellung erzeugt neue Formen der Autorität, aber auch neue Verwundbarkeiten gegenüber Halluzinationen, Plausibilisierung oder der Verschiebung kritischer Urteilkraft.

1 Einleitung

Generative KI hat sich in sehr kurzer Zeit von einem spezialisierten Forschungsgegenstand der Informatik zu einer alltäglichen Wissensinstanz entwickelt. Große Sprachmodelle, wie bspw. Chat GPT, Google Gemini, Meta Llama oder Anthropic Claude beantworten Fragen, fassen Texte zusammen, erzeugen Argumente, erklären Fachbegriffe, schreiben Programmcode, unterstützen die wissenschaftliche Recherche und werden immer mehr zum alltäglichen Begleiter für viele Menschen. Damit verändern sie nicht nur Arbeitsprozesse, sondern vor allem auch die Bedingungen, unter denen Menschen Information als glaubwürdig, plausibel oder wahr behandeln. Denn die Antworten von generativen KI-Systemen basieren auf mathematischen Wahrscheinlichkeiten, die sog. neuronale Netze verwenden, welche durch intensives Training mit kuratierten Informationskollektionen entstanden sind. Durch die Art, Zusammensetzung und Gewichtung dieser Informationskollektionen resultiert ein sog. Bias, also eine gewisse Färbung der Information, eine Art vorgefertigter Meinung, vergleichbar mit einem Kulturraum oder einer politischen Strömung. Aktuell sind die Hersteller von KI-Systemen für die Zusammenstellung der Informationskollektionen verantwortlich, für die Nutzenden bleibt dies meist im Verborgenen. Auch werden KI-Systeme so entwickelt, dass sie immer versuchen, eine Antwort zu geben – auch wenn diese ggf. inkorrekt ist. In der Praxis spricht man dann von sog. Halluzinationen.

Aus philosophischer Sicht ist die interessante Frage daher nicht allein, ob solche Systeme intelligent sind (das ist ebenfalls eine spannende Frage, aber nicht Kontext dieses Artikels), sondern welche Rolle sie im Gefüge menschlichen Wissenserwerbs oder sogar menschlicher Erkenntnis einnehmen.

Unter epistemischer Infrastruktur werden im Folgenden technische, soziale und institutionelle Strukturen, verstanden die nicht selbst Wissen im starken Sinn besitzen müssen, aber die Bedingungen prägen, unter denen Menschen Informationen finden, bewerten, weitergeben und als Wissen anerkennen. Generative KI ist in diesem Sinn keine bloße Informationsquelle, sondern eine vermittelnde Struktur, die Suchrichtungen, Plausibilitätswahrnehmungen und Begründungspraktiken mit formt.

Die öffentliche Diskussion über KI fokussiert derzeit häufig ethische, rechtliche oder wirtschaftliche Fragen. Diese Perspektiven sind wichtig, sie verdecken jedoch eine grundlegende epistemologische Verschiebung: generative KI wird zunehmend dort eingesetzt, wo Menschen bisher auf Lehrbücher, wissenschaftliche Texte, Experten, Suchmaschinen, Kollegen oder institutionelle Autoritäten zurückgegriffen haben. Sie treten somit in Räume ein, die traditionell durch Vertrauen, Zeugenschaft und Autorität strukturiert sind – oder waren.

Diese Verschiebung ist deshalb bedeutsam, weil Wissen niemals nur aus isolierten wahren Sätzen besteht, weil Information nie ohne Kontext entstehen kann. Wissen ist eingebettet in Praktiken des Begründens, Prüfens, Überliefers und Vertrauens. Menschen wissen vieles nicht aus eigener Anschauung, sondern weil sie anderen vertrauen: Lehrenden, Forschenden, Institutionen, Messinstrumenten, Medien oder technischen Systemen. Generative KI fügt nunmehr eine neue Instanz hinzu, deren Aussagen sprachlich kompetent und oftmals überzeugen erscheinen, deren Verhältnis zu Wahrheit, Begründung und Verstehen jedoch unklar bleibt.

Die leitende These dieses Artikels lautet daher: Generative KI ist weniger eine neue Form wissender Intelligenz als vielmehr eine neue Form epistemischer Infrastruktur. Sie erzeugt nicht notwendig Wissen im Starken Sinn, verändert aber vor allem die Bedingungen, unter denen Menschen Wissen suchen, prüfen, übernehmen und legitimieren.

Daraus folgt zwangsläufig auch eine Verschiebung einer zentralen Frage: Statt primär zu fragen, ob KI wissen kann, ist vielmehr zu fragen, wann und warum menschliches Vertrauen in KI-vermittelte Aussagen epistemisch gerechtfertigt ist. Dieser Artikel nähert sich der These und Fragestellung aus verschiedenen Perspektiven und versucht eine grundlegende, erste Einordnung.

2 Wissen, Information und Verstehen

Betrachtet man Wissen aus Sicht der Informatik, so bilden Daten die primäre Grundlage. Aus einzelnen Zeichen entstehen Wörter, aus Wörtern Sätze oder Daten, diese werden zu Wissen und im passenden Kontext zu Information. Haben Nutzer ein Informationsbedürfnis (Belkin), z.B. weil sie nach einem Begriff suchen, so wird dieses durch das Finden passender Ergebnisse im besten Fall befriedigt. In der Informationstheorie existieren zahlreiche Modelle, die sich mit dem Erwerb, der Weitergabe und der Aufbereitung von Wissen beschäftigen (Nonaka & Takeuchi) und durch ausgefeilte technische Lösungen erstaunliche Resultate erzielen – wir merken das jeden Tag während der Nutzung unserer digitalen Systeme.

Aus Sicht der Erkenntnistheorie wird Wissen hingegen häufig als eine besonders qualifizierte Form wahrer Überzeugung verstanden. Gettier formuliert dies in etwa so: Wissen ist mehr als bloßes Fürwahr-Halten und mehr als zufällig wahre Information – es verlangt eine Verbindung von Wahrheit, Rechtfertigung und epistemischer Verantwortung. Bei generativer KI wird diese Unterscheidung besonders wichtig, denn sie kann Aussagen produzieren, die wahr, plausibel, nützlich und kontextsensitiv erscheinen. Dennoch folgt daraus nicht ohne Weiteres, dass das System diese Aussagen tatsächlich auch “weiß”. Systeme können problemlos korrekte Antworten erzeugen, ohne Überzeugungen im menschlichen Sinn zu besitzen, ohne Gründe als Gründe zu verstehen und ohne sich gegenüber der Wahrheit seiner Aussage verantwortlich zu verhalten. Es besteht daher eine Differenz zwischen korrekter Informationsproduktion und Wissen im epistemischen Sinn.

Ähnliches gilt für den Begriff des Verstehens. Während die Informatik hier eher auf die Internalisierung von Wissen und auf die Unterstützung der Informationsaufnahme durch Visualisierung, Modellierung oder Interaktion abzielt, versteht die Philosophie Verstehen nicht nur als die Fähigkeit, passende Muster zu verarbeiten. Verstehen zeigt sich vielmehr darin, Zusammenhänge erklären, Gründe abwägen, Gegenbeispiele prüfen, Relevanz einschätzen und die Grenzen eigener Aussagen bestimmen zu können. Generative KI simuliert viele dieser Fähigkeiten in beeindruckender Weise. Sie erzeugt den Eindruck verstehender Kommunikation, ohne dass damit bereits geklärt wäre, ob dem System ein eigenes Weltverhältnis, eine intentionale Perspektive oder eine Form begrifflicher Verantwortung zukommt.

Nun kann generative KI allerdings auch nicht als bloße Textstatistik abgetan werden, denn ihre praktische Leistungsfähigkeit verändert bereits reale Erkenntnispraktiken. Andererseits wäre es ebenso voreilig, aus funktionaler Sprachkompetenz auf Wissen, Verstehen oder sogar auf epistemische Subjektivität zu schließen. Vielmehr muss genau dieser Bereich zwischen bloßer Datenverarbeitung und menschlichem Wissen eigenständig analysiert werden. Und hier kommt der Gedanke der epistemischen Infrastruktur ins Spiel.

3 Generative KI als epistemische Infrastruktur

Der Begriff der epistemischen Infrastruktur erlaubt es, generative KI jenseits der Alternative von Werkzeug und Akteur zu beschreiben, denn eine Infrastruktur ist nicht bloß ein passives Mittel, welches äußerlich zu menschlichen Praktiken hinzutritt – sie strukturiert vielmehr Möglichkeiten, setzt Standards, erleichtert bestimmte Handlungen und erschwert andere. Bibliotheken, Suchmaschinen, Datenbanken oder wissenschaftliche Publikationssysteme sind Beispiele für Infrastrukturen, die Erkenntnisprozesse ermöglichen und zugleich formen. Generative KI übernimmt in diesem Sinn infrastrukturelle Funktionen: sie ordnet Informationsbestände, erzeugt Zusammenfassungen, schlägt Interpretationen vor, formuliert Hypothesen und vermittelt zwischen Wissensakteuren. Sie ist unmittelbar in den Erkenntnisprozess integriert, denn wer mit KI recherchiert, denkt nicht völlig unabhängig von ihr – die Suchrichtung, die begriffliche Rahmung und häufig auch die wahrgenommene Plausibilität von Antworten werden durch das System mitgeprägt. Und diese infrastrukturelle Rolle unterscheidet generative KI auch von klassischen Nachschlagewerken.

Ein Lexikon bspw. präsentiert relativ stabile, redaktionell kontrollierte Einträge. Eine Suchmaschine verweist auf Quellen, deren Autorenschaft und Kontext zumindest prinzipiell sichtbar bleiben und überprüft werden könnten. Ein Sprachmodell hingegen erzeugt eine neue, synthetische Antwort, die Quellen, Wahrscheinlichkeiten, Kontexte und Formulierungen in einer sprachlichen Einheitlichkeit zusammenführt. Dadurch erscheint die Antwort

oft unmittelbarer, kohärenter und autoritativer als die zugrunde liegende epistemische Lage das eigentlich rechtfertigen würde. Und gerade hier liegt die epistemische Ambivalenz generativer KI: als Infrastruktur kann sie Erkenntnis erweitern, indem sie Perspektiven sichtbar macht, Komplexität reduziert und Zugänge schafft. Gleichzeitig kann sie Erkenntnis aber auch verengen, wenn sie Unsicherheiten glättet, Kontroversen ausblendet oder sprachliche Plausibilität an die Stelle der begründeten Prüfung setzt. Generative KI ist daher weder epistemisch neutral noch notwendigerweise epistemisch defizitär. Sie ist vielmehr eine vermittelnde Struktur, deren Wert von ihrer Einbettung in menschliche Prüf- und Urteilspraxen abhängt.

4 Vertrauen, Zeugenschaft und epistemische Autorität

Ein großer Teil menschlichen Wissens beruht auf Zeugenschaft. Wir wissen historische Tatsachen, wissenschaftliche Ergebnisse, medizinische Befunde oder politische Ereignisse meist nicht durch eigene Beobachtung, sondern durch die Aussagen anderer. Diese Zeugenschaft ist nur umsetzbar, wenn den bezeugenden Personen oder Systemen auch vertraut wird. Vertrauen ist somit keine Randbedingung des Wissens, sondern eine zentrale Voraussetzung. Ohne Vertrauen in andere Personen, Institutionen und Verfahren wäre eine moderne Wissensgesellschaft unmöglich.

Generative KI tritt nunmehr in Praktiken ein, die einer Zeugenschaft ähneln, ohne selbst Zeuge zu sein. Sie berichtet nicht aus eigener Erfahrung, übernimmt keine persönliche oder personale Verantwortung und kann auch nicht im moralischen Sinn zur Rechenschaft gezogen werden. Dennoch wird sie von Nutzern häufig so behandelt, als liefere sie eine kompetente Auskunft. Diese quasi-testimoniale Stellung ist sowohl philosophisch als auch technisch neuartig: die KI spricht mit der Form einer qualifizierten Antwort, aber ohne den Status eines verantwortlichen Sprechers.

Epistemisches Vertrauen in KI unterscheidet sich deshalb von interpersonalem Vertrauen. Einer Person vertraut man nicht nur aufgrund beobachteter Zuverlässigkeit, sondern auch aufgrund zugeschriebener Aufrichtigkeit, Kompetenz und Verantwortlichkeit. Gerade in Deutschland wurde dies durch ein staatlich verankertes Rollensystem unterstützt, in dem Professoren, Richter, Anwälte, Sachverständige oder Mediziner schon allein aufgrund ihrer Rolle Vertrauen repräsentieren. Auch die Frage nach der Kompetenz wurde in Deutschland schon früh formalisiert und durch Titel, Diplome, Stellenbezeichnungen formalisiert. Bei KI-Systemen kann es Kompetenzindikatoren geben, wie z.B. hohe Trefferquoten, transparente Quellenangaben oder nachvollziehbare Grenzen – Aufrichtigkeit und Verantwortung im personalen Sinn fehlen jedoch. Das Vertrauen in KI muss daher als gerechtfertigte Abhängigkeit von einer technischen Vermittlungsstruktur verstanden werden.

Und genau daraus ergibt sich die Frage nach ihrer epistemischen Autorität. Eine Quelle besitzt epistemische Autorität, wenn ihre Aussagen für andere als Gründe gelten können, Überzeugungen zu bilden oder zu revidieren. Faktisch erhält Generative KI eine solche Autorität, wenn Menschen ihre Antworten ungeprüft übernehmen, ihre Vorschläge als Orientierung verwenden oder ihr Urteil gegenüber eigenen Zweifeln priorisieren. Diese Autorität ist jedoch nicht automatisch legitim, sie muss an Bedingungen gebunden werden: Transparenz der Herkunft, Prüfbarkeit der Aussagen, Kontextsensibilität, Eingeständnis von Unsicherheit und die Erhaltung menschlicher Urteilskraft.

5 Halluzinationen und Wahrheit, die Krise sprachlicher Plausibilität

Insbesondere das Problem der Halluzinationen zeigt die epistemische Besonderheit generativer KI. Halluzinationen sind nicht einfach Fehler im traditionellen Sinn – sie bestehen häufig aus sprachlich plausiblen, syntaktisch korrekten und rhetorisch überzeugenden Aussagen, die jedoch faktisch falsch, unbelegt oder irreführend sind. Das Gefährliche liegt hierbei nicht an dem Fehler an sich, denn die gibt es praktisch in jeder Wissensbasis. Vielmehr ist die Verbindung von Fehlerhaftigkeit mit hoher sprachlicher Autorität gefährlich.

Generative KI destabilisiert dadurch das Verhältnis von Wahrheit und Plausibilität, denn Menschen orientieren sich in der Kommunikation nicht nur an Beweisen, sondern auch an Stil, Kohärenz, Sicherheit und Anschlussfähigkeit. Ein Text, der präzise klingt, Fachterminologie verwendet und argumentative Struktur zeigt, wird leichter als epistemisch höherwertig wahrgenommen. Und Sprachmodelle können genau diese Signale erzeugen, ohne dass sie durch Wahrheit oder Rechtfertigung hinterlegt sind. Die Flüssigkeit der Darstellung kann schnell mit der Stärke der Begründung verwechselt werden.

Diese Verwechslung betrifft nicht nur Laien. Auch Experten können generative KI produktiv nutzen und zugleich von ihren Plausibilisierungseffekten beeinflusst werden. Gerade in komplexen Feldern, in denen vollständige Prüfungen aufwändig sind, kann die KI-Antwort als kognitiver Abkürzungsmechanismus dienen. In diesem Fall wird zwar die Geschwindigkeit der Wissensarbeit erhöht, aber auch die Schwelle verändert, ab der etwas als hinreichend geprüft gilt. Aus Sicht der Informatik versucht man sich in diesem Kontext an der Erkennung von fehlerhaften Inhalten (Stichwort Fake-News), KI-Wasserzeichen, Echtheitsnachweisen, digitalen Signaturen, u.v.m. Aus Sicht der Philosophie ergänzt man, dass Wahrheit zwar nicht verschwindet, ihre soziale Erkennbarkeit aber deutlich schwieriger wird. Wenn synthetische Texte, Bilder und Stimmen massenhaft erzeugt werden können, verliert der bloße Anschein an Dokumentiertheit an epistemischem Gewicht, die Oberfläche des Wissens wird imitierbar. Umso mehr wächst die Bedeutung von Verfahren, Institutionen und Praktiken, die Aussagen prüfbar machen. Informatik und Philosophie sind sich an dieser Stelle vielleicht so nahe, wie nie zuvor.

6 Bedingungen an gerechtfertigtes Vertrauen

Versteht man generative KI als epistemische Infrastruktur, dann ist Vertrauen in sie nicht grundsätzlich irrational. Es kann und muss durchaus sinnvoll und gerechtfertigt sein, KI-Systeme als heuristische, explorative oder assistierende Instanzen zu nutzen. Entscheidend ist jedoch, dass dieses Vertrauen nicht blind, sondern abgestuft, kontextsensibel und prüfbar bleibt. Hierzu müssen fünf Grundprinzipien eingehalten werden:

Erstens verlangt gerechtfertigtes Vertrauen Transparenz. Nutzer müssen erkennen können, ob eine Antwort auf überprüfbaren Quellen, auf generativer Rekonstruktion oder auf bloßer Wahrscheinlichkeitsfortsetzung beruht. Zweitens müssen Systeme Unsicherheiten, Grenzen und mögliche Fehler sichtbar machen. Drittens ist die Anschlussfähigkeit an menschliche Prüfung erforderlich. Eine KI-Antwort sollte nicht als Endpunkt, sondern als Ausgangspunkt kritischer Reflexion dienen. Viertens muss epistemisches Vertrauen rollen- und kontextabhängig sein. In harmlosen Alltagssituationen kann eine plausible KI-Antwort genügen. In Wissenschaft, Recht, Medizin, Bildung oder öffentlicher Meinungsbildung gelten höhere An-

forderungen. Je stärker KI-Aussagen in Entscheidungen, Bewertungen oder institutionelle Urteile eingehen, desto höher müssen die Anforderungen an Nachvollziehbarkeit, Quellenkritik und Verantwortungszuweisung sein. Und schließlich darf Vertrauen in KI nicht zur Delegation epistemischer Verantwortung führen. Menschen können sich technischer Systeme bedienen, aber sie bleiben für ihre Überzeugungsbildung und ihre Entscheidungen verantwortlich. Gerade im wissenschaftlichen oder rechtlichen Kontext bedeutet das: KI kann z.B. bei der Recherche unterstützen, Argumentationsstrukturen vorschlagen, blinde Flecken sichtbar machen. Sie kann jedoch nicht die eigene Prüfung, Quellenarbeit und argumentative Verantwortung ersetzen.

7 Fazit

KI ist längst kein technisches Thema mehr. Die Fragestellungen, die sich insbesondere aus dem Kontext der generativen KI heraus ergeben, sind interdisziplinär und fordern nahezu jeden Bereich der Wissenschaft und des öffentlichen Lebens auf, sich mit dem Thema zu beschäftigen und Lösungsansätze zu entwickeln. Der philosophische Wert der Debatte über generative KI liegt bspw. nicht allein in der Frage, ob Maschinen denken, wissen oder verstehen. Auch wenn diese Fragen natürlich wichtig sind, greifen sie zu kurz, wenn sie die gesellschaftliche Rolle generativer KI auf das Problem der künstlichen Subjektivität reduzieren. Die eigentliche Herausforderung besteht vielmehr darin, dass KI-Systeme bereits heute menschliche Erkenntnispraktiken mitstrukturieren, ohne dabei selbst eindeutig in die Kategorien von Werkzeug, Quelle, Zeuge oder Akteur zu passen. Die hier vorgeschlagene Beschreibung generativer KI als epistemische Infrastruktur macht diese Zwischenstellung sichtbar. KI erzeugt nicht notwendig Wissen im starken Sinn, aber sie verändert, wie Menschen Wissen suchen, darstellen, bewerten und legitimieren. Daraus resultiert dann auch eine neue, zentrale Forschungsfrage: unter welchen Bedingungen ist Vertrauen in KI-vermittelte Erkenntnis gerechtfertigt, und wann wird es zur epistemischen Selbstentlastung?

Eine philosophische Untersuchung dieses Themas sollte daher drei Ebenen verbinden: erstens die klassische Frage nach Wissen, Wahrheit und Verstehen. Zweitens die sozialemistemologische Frage nach Zeugenschaft, Vertrauen und Autorität und drittens die technikphilosophische Frage nach der Rolle von Infrastrukturen, Artefakten und algorithmischen Vermittlungsformen. In dieser Verbindung liegt gleichzeitig auch die Originalität des Themas.

Generative KI fordert die Erkenntnistheorie nicht deshalb heraus, weil sie menschliches Wissen einfach ersetzt, sondern weil sie die Bedingungen verändert, unter denen Menschen sich selbst als wissende Subjekte verstehen.

Literaturverzeichnis

- Audi, Robert (2010): *Epistemology: A Contemporary Introduction to the Theory of Knowledge*. 3. Aufl. New York/London: Routledge.
- Belkin, Nicholas J. (1980): "Anomalous States of Knowledge as a Basis for Information Retrieval". In: *Canadian Journal of Information Science* 5(1), S. 133–143.
- Coady, C. A. J. (1992): *Testimony: A Philosophical Study*. Oxford: Clarendon Press.

- Floridi, Luciano (2011): *The Philosophy of Information*. Oxford: Oxford University Press.
- Fricker, Miranda (2007): *Epistemic Injustice: Power and the Ethics of Knowing*. Oxford: Oxford University Press.
- Gettier, Edmund L. (1963): “Is Justified True Belief Knowledge?”. In: *Analysis* 23(6), S. 121–123.
- Hemmje, Matthias; Wagenpfeil, Stefan (2026): *Multimedia Information Retrieval: Extracting Information from Multimedia Sources*. Berlin/Heidelberg: Springer. DOI: 10.1007/978-3-662-73310-3.
- Ihde, Don (2009): *Postphenomenology and Technoscience: The Peking University Lectures*. Albany: State University of New York Press.
- Nonaka, Ikujiro; Takeuchi, Hirotaka (1995): *The Knowledge-Creating Company: How Japanese Companies Create the Dynamics of Innovation*. New York/Oxford: Oxford University Press.
- O’Neill, Onora (2002): *A Question of Trust: The BBC Reith Lectures 2002*. Cambridge: Cambridge University Press.
- Pritchard, Duncan (2023): *What Is This Thing Called Knowledge?*. 5. Aufl. London/New York: Routledge.
- Searle, John R. (1980): “Minds, Brains, and Programs”. In: *Behavioral and Brain Sciences* 3(3), S. 417–457.
- Vallor, Shannon (2024): *The AI Mirror: How to Reclaim Our Humanity in an Age of Machine Thinking*. New York: Oxford University Press.
- Verbeek, Peter-Paul (2005): *What Things Do: Philosophical Reflections on Technology, Agency, and Design*. University Park: Pennsylvania State University Press.