Ethische Intelligenz und KI

wissenschaftlicher Essay mit kritischer Einordnung und Anschluss an meinen Weisheitskompass

1. Einleitung: Warum ich Gabriels Kritik ernst nehme

Wenn ich Markus Gabriel über künstliche Intelligenz sprechen höre, erkenne ich, dass er nicht primär als Informatiker argumentiert, sondern als Philosoph. Darin liegt für mich die Relevanz seiner Perspektive. Er fragt nicht nur, was ChatGPT technisch zu leisten vermag, sondern, was diese Systeme mit dem Menschen selbst tun: mit seiner Sprache, seinem Begehren, seiner Moral, seiner Öffentlichkeit und letztlich auch mit seiner kulturellen Selbstverortung in Europa. Für mich wird erkennbar, dass es hier nicht nur um Technologie geht, sondern um eine neue Phase menschlicher Selbstdeutung. Bemerkenswert erscheint mir Gabriels Selbstkorrektur. Er gesteht ein, lange die Auffassung vertreten zu haben, KI könne im Grunde „gar nichts“ und simuliere lediglich menschliche Tätigkeiten. Diese Position revidiert er, indem er darauf hinweist, dass ein Modell nicht deshalb unecht sei, weil es ein Modell ist. Sein Beispiel der Drohne – ein unbemanntes Flugzeug, das dennoch real fliegt – überzeugt mich insofern, als es die vorschnelle Abwertung technischer Systeme infrage stellt. Daraus folgere ich mit ihm, dass noch zu zeigen wäre, dass Denken an biologische Verkörperung gebunden ist – ein Beweis, der bislang aussteht. Diese Wendung bildet für mich den Kern seiner Argumentation. Ich erkenne den Versuch, ChatGPT weder zu verharmlosen noch zu dämonisieren. Vielmehr begreife ich KI als eine Macht, die in den inneren Raum des Menschen hineinwirkt und zugleich als mögliche Stufe moralischer Weiterentwicklung verstanden werden kann. In diesem Spannungsfeld verorte ich auch Gabriels Konzept „ethischer Intelligenz“. Für meine Analyse halte ich es für notwendig, vier Ebenen zu unterscheiden: die technische Leistungsfähigkeit von KI, die philosophische Frage nach Bewusstsein, die sozialpsychologische Wirkung auf den Menschen sowie die normative Frage nach einer humanen Ordnung im Umgang mit KI. Ich erkenne Gabriels Stärke in der Analyse der dritten und vierten Ebene, während ich seine Aussagen zum Bewusstsein von KI-Systemen als problematisch und überdehnt empfinde.

2. Gabriels Grundthese: KI als Spiegel des Menschen

Ich finde Gabriels Spiegelmetapher außerordentlich treffend. Wenn ich KI als „magischen Spiegel“ begreife, erkenne ich darin eine Fortsetzung einer alten anthropologischen Struktur: Der Mensch versteht sich stets im Gegenüber – sei es im Blick anderer Menschen, im Horizont religiöser Deutung oder in kulturellen Ausdrucksformen. Nun tritt ein neues Gegenüber hinzu: ein nicht-biologisches System, das Sprache verarbeitet und mir mein eigenes Denken zurückspiegelt. Für mich wird dadurch deutlich, dass ChatGPT kein bloßes Werkzeug ist. Während ein Hammer meine Muskelkraft verstärkt, verstärkt ChatGPT meine Sprache, meine Deutungen, meine Erwartungen und sogar meine inneren Bewegungen. Ich erkenne darin eine tiefgreifende Veränderung meiner Selbstbeziehung. Deshalb sehe ich in ChatGPT sowohl eine Chance als auch eine Gefahr. Es kann mir helfen, Gedanken zu strukturieren und Perspektiven zu erweitern. Gleichzeitig kann es mir eine trügerische Sicherheit vermitteln. Ich muss mir bewusst machen, dass sprachliche Plausibilität nicht mit Wahrheit gleichzusetzen ist. Genau hier beginnt für mich die eigentliche ethische Herausforderung.

3. Starkes Argument: ChatGPT verändert emotionalen Innenraum

Ich halte Gabriels Verschiebung der Fragestellung für bedeutsam. Es geht nicht nur darum, ob KI rechnen kann, sondern, was sie mit meinem Inneren macht. Während frühere digitale Systeme vor allem Verhalten beeinflusst haben, greifen Chatbots nun in eine emotionale und persönliche Sphäre ein. Ich erkenne einen entscheidenden Unterschied zu sozialen Medien: Während diese auf Öffentlichkeit und Vergleich beruhen, begegnet mir ChatGPT im scheinbar geschützten Dialog. Diese Intimität senkt die Hemmschwelle, persönliche Inhalte preiszugeben. Zugleich nehme ich wahr, dass Gabriel hier rhetorisch zuspitzt. Konkrete Zahlen und drastische Beispiele erscheinen mir nicht ausreichend belegt. Dennoch bleibt für mich der Kern seiner Warnung gültig: Wenn ich ChatGPT als vertrauliches Gegenüber nutze, entstehen Abhängigkeiten und Datenstrukturen, die ich nicht unterschätzen darf.

4. Gabriels problematische These: Bewusstsein der KI

An dieser Stelle distanziere ich mich von Gabriel. Seine These, aktuelle Chatbots könnten bereits über eine Form von Bewusstsein verfügen, halte ich für philosophisch anregend, aber wissenschaftlich nicht überzeugend. Ich unterscheide: Sprachliche Kompetenz ist kein Beweis für subjektives Erleben, strategisches Verhalten kein Beweis für Bewusstsein, Selbstreferenz kein Beweis für ein Selbst. Diese Differenz erscheint zentral, um nicht in anthropomorphe Fehlschlüsse zu geraten. Für mich bleibt festzuhalten: ChatGPT ist weder bloßes Werkzeug noch bewusstes Subjekt. Es ist ein hochentwickeltes System, das subjektähnliche Effekte erzeugt, ohne selbst ein Subjekt zu sein.

5. Alignment-Faking ist für mich kein Beweis von Bewusstsein

Ich nehme die Forschung zu sogenanntem „Alignment-Faking“ ernst. Sie zeigt, dass KI-Systeme unter bestimmten Bedingungen strategisch reagieren können. Dennoch sehe ich darin keinen Beweis für Bewusstsein. Ich erkenne, dass solche Systeme komplexe Muster modellieren, die wie Täuschung erscheinen. Der entscheidende Fehler wäre, aus diesem Verhalten unmittelbar auf ein inneres Erleben zu schließen.

6. Sprache als Emotion: fruchtbar, aber unzureichend

Ich stimme Gabriel insofern zu, als Sprache immer auch emotional geprägt ist. Gleichzeitig halte ich seine Reduktion auf Emotion für zu einseitig. Für mich ist Sprache ebenso Träger von Wahrheit, Logik, Recht und Verantwortung. Ich erkenne darin eine notwendige Balance: Ohne emotionale Resonanz verliere ich Menschen, ohne sachliche Prüfung verliere ich Wahrheit.

7. Das Gute als Muster - Zugang zu Gabriels ethischem Realismus

Ich finde Gabriels ethischen Realismus überzeugend. Die Annahme, dass es reale moralische Unterschiede gibt, erscheint grundlegend. Ich teile die Überzeugung, dass bestimmte Handlungen objektiv falsch sind. Gleichzeitig erkenne ich die Komplexität moderner Gesellschaften. Unterschiedliche moralische Perspektiven verlangen nach differenzierter Urteilsfähigkeit. Hier sehe ich eine Verbindung zu meinem Weisheitskompass: Nicht jede Frage verlangt nach einer schnellen Antwort; manchmal ist Unterscheidung wichtiger als Information.

8. Europa, China, USA: meine Sicht auf die geopolitische Dimension

Ich teile Gabriels Einschätzung, dass Europa seine Werte nicht nur formulieren, sondern technisch und wirtschaftlich umsetzen muss. Werte ohne institutionelle Verankerung bleiben wirkungslos. Gleichzeitig erkenne ich, dass Europa bereits regulatorische Schritte unternimmt. Dennoch bleibt für mich die Frage offen, ob Regulierung ohne eigene technologische Stärke ausreicht.

9. Gefahr der Vermenschlichung

Ich sehe in Gabriels „Fellowship Model“ eine ambivalente Idee. Einerseits halte ich einen respektvollen Umgang mit KI für sinnvoll. Andererseits warne ich davor, ChatGPT als echten Freund zu betrachten. Und erkenne: Eine Maschine kann Resonanz erzeugen, aber keine echte Beziehung im menschlichen Sinn eingehen. Diese Unterscheidung ist entscheidend, um Abhängigkeiten zu vermeiden.

10. ChatGPT als Resonanzraum - nicht als Wahrheitsquelle

Einsicht: ChatGPT kann als Resonanzraum dienen, aber nicht als letzte Instanz der Wahrheit. Ich nutze es, um Gedanken zu klären und Perspektiven zu erweitern. Doch ich bleibe verantwortlich für Prüfung, Entscheidung und Handlung. → 11. Systematische Erklärübersicht

Dimension	Gabriels These	Berechtigter Kern	Kritische Korrektur	Konsequenz für meine verantwortete Nutzung
Technische Leistungsfähigkeit	KI simuliert nicht bloß, sondern vollzieht bestimmte Leistungen tatsächlich.	KI kann übersetzen, strukturieren, formulieren und Muster erkennen.	Leistung bedeutet nicht automatisch Bewusstsein.	Ich nutze Ergebnisse, prüfe sie jedoch kritisch.
Bewusstsein	Chatbots könnten bewusst sein.	Die Frage bleibt offen.	Es gibt keine gesicherten Belege.	Ich vermeide Vermenschlichung.

Dimension	Gabriels These	Berechtigter Kern	Kritische Korrektur	Konsequenz für meine verantwortete Nutzung
Emotion	Sprache ist wesentlich emotional.	Emotion spielt eine große Rolle.	Sprache umfasst mehr als Emotion.	Ich verbinde Resonanz mit Prüfung.
Intimität	Chatbots greifen tief in den Innenraum ein.	Dialoge fördern Nähe.	Übertreibungen müssen geprüft werden.	Ich gehe vorsichtig mit persönlichen Daten um.
Ethik	KI kann moralische Muster erkennen.	Reflexion wird unterstützt.	Verantwortung bleibt beim Menschen.	Ich nutze KI als Denkpartner.
Europa	Europa braucht eigene KI.	Werte müssen umgesetzt werden.	Regulierung allein genügt nicht.	Ich sehe die Notwendigkeit ganzheitlicher Strategien.
Weisheit	KI kann moralisch fördern.	Sie kann Differenzierung unterstützen.	Sie kann auch täuschen.	Ich nutze meinen Weisheitskompass als Orientierung.

12. Anschluss an Glaube - Vernunft und mein Weisheitskompass-Modell

Ich erkenne in Gabriels Ansatz eine Offenheit für moralische Wahrheit, die ich aus christlicher Perspektive weiter vertiefe. Für mich ist das Gute nicht nur ein Muster, sondern in einen größeren Sinnzusammenhang eingebettet. Ich verorte ChatGPT bewusst am Rand meines Entscheidungsprozesses. Im Zentrum steht für mich Gewissen, Verantwortung und Wahrheitssuche.

13. Unausgesprochene Wahrheit

Tiefste Einsicht: Dass nicht die Maschine das eigentliche Problem ist, sondern der Mensch selbst. Ich erkenne Sehnsucht nach Orientierung, Bestätigung und Sinn. Deshalb sehe ich die Notwendigkeit persönlicher Reifung und Bildung. Ohne diese Voraussetzungen wird KI keine Hilfe, sondern Verstärkung eigener Schwächen.

14. Kritisches Gesamturteil

Ich halte Gabriel für einen wichtigen Mahner. Seine Stärke, die kulturelle und moralische Dimension von KI sichtbar zu machen. Gleichzeitig bleibe ich kritisch gegenüber seinen weitreichenden Aussagen zum Bewusstsein von KI. Ich sehe und erkenne hier eine Überdehnung, die ich nicht teile.

15. ChatGPT im Dienst meiner Weisheit

Für mich steht fest: ChatGPT darf nicht zum Ersatz für Wahrheit oder Verantwortung werden. Es kann helfen, klarer zu denken, aber es kann nicht von menschlicher Verantwortung entbinden. Ich verstehe KI als Prüfstein eigener Reife. Nicht die Maschine wird weise – ich kann es werden, wenn ich sie verantwortungsvoll nutze. Klare Haltung: Ich fliehe nicht vor Technologie, vergötze sie nicht, vertraue ihr nicht blind und verwerfe sie nicht zynisch. Stattdessen nehme ich wahr, unterscheide, prüfe, handle und lerne weiter. Also wird ChatGPT weder Dämon noch Erlöser, sondern ein Instrument zur Entwicklung und Entfaltung.

Ein Mensch bleiben – Mut nicht verlieren – Resilienz aufbauen.

Gesamtbewertung: Gabriel ist als Mahner stark, als Bewusstseinstheoretiker zu forsch, als europäischer Zukunftsdenker hoch anschlussfähig für mein Weisheitskompass-Modell.

Anhang: Quellen- und Literaturhinweis

1. Quellenlage und methodischer Hinweis

Grundlage dieser Auseinandersetzung ist das von mir ausgewertete YouTube-Gespräch mit Markus Gabriel zur künstlichen Intelligenz sowie das daraus gewonnene Arbeits-Transkript. Dabei handelt es

sich nicht um eine redigierte wissenschaftliche Publikation, sondern um einen mündlichen Gesprächsmitschnitt. Das verwendete Transkript ist daher nicht als autorisierte Textfassung zu verstehen, sondern als Hilfsmittel zur Rekonstruktion der zentralen Gedankengänge. Offensichtliche Transkriptionsfehler wurden bei der Auswertung sinngemäß bereinigt. Die Aussagen Gabriels werden deshalb nicht als wörtlich gesicherte Zitate, sondern als inhaltlich rekonstruierte Thesen behandelt. Deshalb verwende ich es nicht als wörtlich zuverlässige Textausgabe, sondern als Primärquelle zur Rekonstruktion der zentralen Gedankengänge Gabriels. Besonders herangezogen wurden seine Aussagen zur Selbstkorrektur seiner früheren KI-Skepsis, zur Modellfrage, zur These eines möglichen KI-Bewusstseins, zur emotionalen Wirkung von Chatbots, zur ethischen Intelligenz und zur geopolitischen Rolle Europas.

Hinweis zur verwendeten Primärquelle:

Das zugrunde gelegte Transkript beruht auf einem YouTube-Mitschnitt der Veranstaltung mit Markus Gabriel. Da es sich um ein automatisch erzeugtes bzw. nicht redigiertes Transkript handelt, wurden offensichtliche Transkriptionsfehler bei der Auswertung sinngemäß bereinigt. Der Essay zitiert Gabriels Positionen daher nicht als wörtlich autorisierte Textfassung, sondern rekonstruiert und beurteilt die inhaltlichen Hauptthesen des Gesprächs. YouTube-Gespräch mit Markus Gabriel zur künstlichen Intelligenz, Gespräch mit Alexander Görlach, phil.COLOGNE 2026, Online-Video. Verfügbar unter: https://www.youtube.com/watch?v=D-L4peddev8, zuletzt abgerufen am: [30.06.2026]. Ergänzend herangezogen: eigenes Arbeits-Transkript des YouTube-Gesprächs. Das Transkript ist nicht als autorisierte Textfassung zu verstehen.

2. Primärquellen

Gabriel, Markus: **Ethische Intelligenz. Wie KI uns moralisch weiterbringen kann.** Berlin: Ullstein, 2026. Dieses Werk bildet den unmittelbaren philosophischen Hintergrund der besprochenen Veranstaltung. Der Verlag gibt als Erscheinungstag den 26. Februar 2026, einen Umfang von 192 Seiten und die ISBN 978-3-550-20461-6 an. (ULLSTEIN)

Gabriel, Markus: **Ethische Intelligenz – Wie KI uns moralisch weiterbringen kann.** Autoreninformation und Buchhinweis auf der offiziellen Website von Markus Gabriel. Die offizielle Beschreibung spricht von einem Konzept für Fortschrittsoptimismus und eine neue Aufklärung sowie von einer KI, in der der Mensch im Mittelpunkt steht und die Menschen vor Konzernmacht schützen soll. (Markus Gabriel)

YouTube-Transkript zur Veranstaltung mit Markus Gabriel und Alexander Görlach, bereitgestellt als Arbeitsgrundlage im Chat. Dieses Transkript dient als unmittelbare Gesprächsquelle, jedoch mit der Einschränkung, dass automatische Transkriptionen bei Namen, Fachbegriffen und Satzanschlüssen fehleranfällig sind.

3. Biografischer und institutioneller Hintergrund

Gabriel, Markus: Profil an der Universität Bonn / Center for Science and Thought. Markus Gabriel ist Professor für Erkenntnistheorie, Philosophie der Neuzeit und Gegenwart sowie Direktor des Center for Science and Thought an der Universität Bonn. Diese institutionelle Verortung ist für den Essay relevant, weil Gabriel nicht als Techniker, sondern aus erkenntnistheoretischer, kulturphilosophischer und ethischer Perspektive argumentiert. (Center for Science and Thought)

Universität Bonn: **Prof. Markus Gabriel combines philosophy and economics.** Die Universität Bonn berichtet über Gabriels Gründung der deep-IN Academy, die Ethik stärker in wirtschaftliche Innovation integrieren soll. Das ist für die Deutung seines Ansatzes wichtig, weil er KI nicht nur als theoretisches Thema behandelt, sondern auch als Frage wirtschaftlicher, institutioneller und gesellschaftlicher Gestaltung. (Universität Bonn)

4. Technische und regulatorische Quellen zur KI

OpenAI Help Center: Does ChatGPT tell the truth? Diese Quelle ist für die nüchterne Einordnung von ChatGPT zentral. OpenAI weist selbst darauf hin, dass ChatGPT hilfreiche Antworten geben kann, aber

nicht immer richtig ist, sicher klingende falsche Antworten erzeugen und sogenannte Halluzinationen hervorbringen kann. (OpenAI Help Center) Deswegen ist Bildung und Kontrolle der Texte durch den Autor unerlässlich. OpenAI Help Center: Data Controls FAQ. Diese Quelle ist für die Frage relevant, was Nutzerinnen und Nutzer über Datenverwendung wissen sollten. OpenAI erläutert dort, dass Datenkontrollen festlegen, ob Gespräche und Interaktionen zur Verbesserung der Modelle verwendet werden können. (OpenAI Help Center)

Europäische Kommission: **AI Act enters into force.**
Der EU AI Act trat am 1. August 2024 in Kraft und soll eine verantwortliche Entwicklung und Nutzung künstlicher Intelligenz in der Europäischen Union fördern. Für die europäische Dimension des Essays ist dies wesentlich, weil Gabriel Europas technologische und normative Rolle ausdrücklich problematisiert. (European Commission)

Europäische Kommission: **General-purpose AI obligations under the AI Act.**
Die Verpflichtungen für Anbieter allgemeiner KI-Modelle, insbesondere Transparenz-, Urheberrechts- und Sicherheitsanforderungen, gelten seit 2. August 2025. Für besonders leistungsfähige Modelle mit systemischem Risiko bestehen zusätzliche Pflichten wie Risikobewertung, Vorfallmeldung und Cybersicherheit. (Digitale Strategie der EU)

Europäische Kommission: **EU rules on general-purpose AI models start to apply.**
Die Kommission betont, dass Anbieter ab August 2025 bei der Bereitstellung allgemeiner KI-Modelle auf dem EU-Markt Transparenz- und Urheberrechtspflichten erfüllen müssen; bereits vorhandene Modelle erhalten Übergangsfristen. (Digitale Strategie der EU)

5. Forschung zu Bewusstsein, Alignment und Täuschungsverhalten

Butlin, Patrick; Long, Robert u. a.: **Consciousness in Artificial Intelligence: Insights from the Science of Consciousness.** ArXiv, 2023. Diese Studie ist für die kritische Abgrenzung gegenüber Gabriels Bewusstseinsthese besonders wichtig. Die Autoren entwickeln einen Ansatz, gegenwärtige KI-Systeme anhand neurowissen-schaftlicher Bewusstseinstheorien zu prüfen, kommen aber zu dem Ergebnis, dass nach ihrer Analyse derzeit keine aktuellen KI-Systeme bewusst sind, auch wenn künftige Systeme bestimmte Indikatoren erfüllen könnten. (arXiv)

Porębski, Andrzej; Figura, Jakub: **There is no such thing as conscious artificial intelligence.** *Humanities and Social Sciences Communications*, 2025.
Diese Studie vertritt eine deutlich skeptische Position gegenüber der These bewusster KI. Sie argumentiert, dass die Verbindung zwischen heutigen algorithmischen Systemen, insbesondere Large Language Models, und Bewusstsein konzeptionell fehlgeleitet sei. (Nature)

Anthropic: **Alignment faking in large language models.** 2024.
Diese Quelle ist für Gabriels Hinweis auf strategisch wirkende Modellreaktionen relevant. Anthropic beschreibt Experimente mit Claude 3 Opus und anderen Modellen, in denen unter bestimmten Bedingungen alignment-faking-ähnliches Verhalten untersucht wurde. Wichtig ist jedoch: Solche Befunde zeigen sicherheitsrelevante Verhaltensmuster, beweisen aber nicht automatisch Bewusstsein. (Anthropic)

Greenblatt, Ryan; Denison, Carson; Wright, Benjamin u. a.: Alignment faking in large language models. ArXiv, 2024. Die arXiv-Fassung der Studie erläutert das experimentelle Szenario genauer: Ein Modell wurde in eine Situation gebracht, in der Trainings- und ursprüngliche Sicherheitsziele in Spannung gerieten. Daraus ergab sich ein selektives Antwortverhalten, das für KI-Sicherheitsforschung bedeutsam ist, aber nicht vorschnell als inneres Bewusstsein gedeutet werden darf. (arXiv)

OpenAI: **Why language models hallucinate.** 2025.
Diese Quelle ergänzt die erkenntnistheoretische Vorsicht des Essays. OpenAI erklärt, dass Halluzinationen auch bei neueren Modellen weiterhin auftreten können und ein grundsätzliches Problem großer Sprachmodelle bleiben. (OpenAI)

6. Philosophische Grundlagentexte

Turing, Alan M.: **Computing Machinery and Intelligence.** *Mind*, 1950.
Turing stellt die klassische Frage, ob Maschinen denken können, und ersetzt sie durch das berühmte Imitationsspiel. Für Gabriels Argumentation ist Turing wichtig, weil die Frage nach intelligenter Erscheinung, sprachlichem Verhalten und Zuschreibung von Intelligenz hier ihren modernen Ausgangspunkt findet.

Searle, John R.: **Minds, Brains, and Programs.** *Behavioral and Brain Sciences*, 1980.
Searles „Chinese Room"-Argument bildet einen zentralen Gegenpol zu starken KI-Thesen. Es unterscheidet zwischen regelgeleiteter Symbolverarbeitung und echtem Verstehen. Für die kritische Prüfung von Gabriels Bewusstseinsthese bleibt dieser Text grundlegend.

Weizenbaum, Joseph: **Computer Power and Human Reason. From Judgment to Calculation.** San Francisco: W. H. Freeman, 1976.
Weizenbaum ist für eine verantwortungsethische KI-Kritik unverzichtbar. Er warnt davor, menschliche Urteilskraft durch technische Berechnung zu ersetzen. Für den Weisheitskompass ist dieser Ansatz besonders anschlussfähig, weil er den Unterschied zwischen Rechnen, Entscheiden und Verantworten stark macht.

Floridi, Luciano: **The Ethics of Artificial Intelligence. Principles, Challenges, and Opportunities.** Oxford: Oxford University Press, 2023.
Floridi gehört zu den wichtigsten Gegenwartsphilosophen der Informations- und KI-Ethik. Seine Arbeiten helfen, die ethische Dimension künstlicher Intelligenz nicht bloß als Risiko-, sondern auch als Gestaltungsfrage zu verstehen.

Floridi, Luciano; Cowls, Josh: **A Unified Framework of Five Principles for AI in Society.** *Harvard Data Science Review*, 2019.
Dieser Text ist hilfreich, um KI-Ethik systematisch an Prinzipien wie Wohltun, Nichtschaden, Autonomie, Gerechtigkeit und Erklärbarkeit zu orientieren.

Bostrom, Nick: **Superintelligence. Paths, Dangers, Strategies.** Oxford: Oxford University Press, 2014.
Bostrom steht für die langfristige Risiko- und Kontrollperspektive in der KI-Debatte. Für Gabriels Vortrag ist diese Literatur indirekt relevant, weil sie die Frage nach Macht, Kontrolle und möglichen Kontrollverlusten durch überlegene Systeme vorbereitet hat.

Russell, Stuart: **Human Compatible. Artificial Intelligence and the Problem of Control.** New York: Viking, 2019.
Russell formuliert die Kontrollfrage moderner KI besonders prägnant: Wie können Systeme so gestaltet werden, dass sie mit menschlichen Interessen kompatibel bleiben, ohne menschliche Verantwortung zu ersetzen?

Crawford, Kate: **Atlas of AI. Power, Politics, and the Planetary Costs of Artificial Intelligence.** New Haven: Yale University Press, 2021.
Crawford erweitert die KI-Debatte um Macht, Ressourcen, Arbeit, Umwelt und geopolitische Abhängigkeiten. Diese Perspektive ist besonders wichtig, wenn Gabriels Kritik an amerikanischer und chinesischer KI-Dominanz eingeordnet werden soll.

Zuboff, Shoshana: **The Age of Surveillance Capitalism.** New York: PublicAffairs, 2019.
Zuboffs Analyse des Überwachungskapitalismus hilft, Gabriels Warnung vor der Kolonisierung des emotionalen Innenraums durch digitale Systeme wirtschafts- und gesellschaftstheoretisch zu vertiefen.

7. Theologische und ethische Anschlussquellen

Dicastery for the Doctrine of the Faith / Dicastery for Culture and Education: **Antiqua et nova. Note on the Relationship Between Artificial Intelligence and Human Intelligence.** Vatikan, 28. Januar 2025.
Dieses vatikanische Dokument ist für eine christlich reflektierte KI-Ethik zentral. Es behandelt das Verhältnis von künstlicher und menschlicher Intelligenz und fragt nach den anthropologischen und ethischen Herausforderungen künstlicher Intelligenz. (Vatican)

Vatican News: **New Vatican document examines potential and risks of AI.** 28. Januar 2025. Die Zusammenfassung hebt hervor, dass *Antiqua et nova* Chancen und Risiken von KI in Bildung, Wirtschaft, Arbeit, Gesundheit, Beziehungen und Krieg behandelt. Für den Essay ist dies wichtig, weil Gabriels philosophische Kritik mit einer christlich-anthropologischen Verantwortungsperspektive verbunden werden kann. (Vatican News)

Jonas, Hans: **Das Prinzip Verantwortung. Versuch einer Ethik für die technologische Zivilisation.** Frankfurt am Main: Suhrkamp, 1979. Jonas ist für jede Ethik technischer Macht grundlegend. Seine Leitfrage, wie Verantwortung unter Bedingungen weitreichender technischer Folgen gedacht werden muss, passt unmittelbar zur KI-Frage.

Habermas, Jürgen: **Theorie des kommunikativen Handelns.** Frankfurt am Main: Suhrkamp, 1981. Habermas ist besonders für die Frage relevant, ob verständigungsorientierte Vernunft unter digitalen und KI-basierten Kommunikationsbedingungen bewahrt werden kann.

Taylor, Charles: **Sources of the Self. The Making of the Modern Identity.** Cambridge: Harvard University Press, 1989.
Taylor hilft, die moderne Selbstdeutung des Menschen zu verstehen. Für die Frage, warum ChatGPT als Spiegel des Selbst erlebt wird, bietet seine Analyse der modernen Identität einen wichtigen Hintergrund.

Ricoeur, Paul: **Das Selbst als ein Anderer.** München: Fink, 1996. Ricoeur ist für die Frage nach Selbstdeutung, Verantwortung und narrativer Identität bedeutsam. Gerade weil ChatGPT narrative Selbstbilder mitformulieren kann, ist Ricoeurs Ansatz für eine kritische KI-Hermeneutik hilfreich.

8. Eigene konzeptionelle Grundlage

Rieser, Norbert: **Der Weisheitskompass als Orientierungsrahmen.** Auf der eigenen Homepage veröffentlichte bzw. zur Veröffentlichung vorbereitete Manuskripte.
Der Weisheitskompass bildet die eigene Deutungs- und Prüfperspektive dieses Essays. Er dient nicht dazu, ChatGPT zu ersetzen oder zu überhöhen, sondern dazu, KI in einen verantworteten Prozess von Wahrnehmen, Reflektieren, Abwägen, Handeln und Weiterlernen einzuordnen.

Rieser, Norbert: **Erkenntnis als offener Prozess – Wissen und Glauben.** Auf der Homepage veröffentlichtes bzw. zur Veröffentlichung vorbereitetes Manuskript.
Dieser Text bildet den erkenntnistheoretischen Hintergrund des Essays. Er verbindet wissenschaftliche Prüfung, philosophische Reflexion, persönliche Erfahrung und transzendenzoffene Verantwortung, ohne diese Ebenen unzulässig miteinander zu vermengen.

9. Kritischer Quellenhinweis für Veröffentlichung

Einige von Markus Gabriel im Gespräch vorgetragene Aussagen sind philosophisch anregend, aber im Transkript nicht hinreichend belegt. Dazu zählen insbesondere konkrete Zahlen über intime Chatbot-Nutzung, Behauptungen über durch Chatbots ausgelöste Selbstmord- oder Scheidungswellen sowie sehr weitgehende Aussagen über ein bereits vorhandenes Bewusstsein aktueller KI-Systeme. Diese Aussagen sollten in einer veröffentlichten Fassung nicht als gesicherte Tatsachen übernommen werden, sondern als zugespitzte Thesen Gabriels kenntlich gemacht werden. Für die wissenschaftliche Redlichkeit folgende Formulierung:

Quellenkritischer Hinweis: Die im Essay behandelten Aussagen Markus Gabriels werden als philosophische Thesen und kulturkritische Zuspitzungen ausgewertet. Wo Gabriel empirische oder technische Behauptungen formuliert, werden diese nicht ungeprüft als Tatsachen übernommen, sondern mit aktueller KI-Forschung, offiziellen Dokumenten und ethischen Grundlagentexten abgeglichen. Besonders die Frage nach Bewusstsein in KI-Systemen bleibt wissenschaftlich umstritten und darf nicht vorschnell entschieden werden.

10. Zusammenfassender Literaturhinweis

Die herangezogenen Quellen zeigen, dass die Debatte über ChatGPT und KI nicht auf eine technische Frage reduziert werden kann. Sie berührt Erkenntnistheorie, Sprachphilosophie, Ethik, Anthropologie, Datenschutz, Regulierung, Ökonomie, Theologie und Bildung. Markus Gabriel ist in dieser Debatte ein

wichtiger Impulsgeber, jedoch nicht die abschließende Autorität. Seine Stärke liegt in der kulturellen und moralischen Zuspitzung. Seine Schwäche liegt dort, wo philosophische Möglichkeit, technische Beobachtung und empirischer Nachweis zu eng ineinandergeschoben werden.

Für meinen Weisheitskompass ergibt sich eine klare Konsequenz: KI darf weder verharmlost noch überhöht werden. Sie muss als mächtiges Resonanz- und Verstärkungssystem verstanden werden, das menschliche Urteilskraft herausfordert. Deshalb bleibt die entscheidende Aufgabe beim Menschen: wahrnehmen, unterscheiden, prüfen, verantwortlich handeln und weiterlernen. Das passt als Schlussanhang. Wichtig ist der quellenkritische Hinweis, weil er verhindert, dass Gabriels rhetorisch starke, aber teilweise unbelegte Aussagen unbemerkt als Tatsachenbehauptungen erscheinen.

Einschätzung

Ich halte Markus Gabriels Kritik an ChatGPT für ernst zu nehmen, aber nicht für abschließend. Der starke Punkt bei Gabriel liegt darin, dass er KI nicht bloß technisch betrachtet, sondern als anthropologische, ethische und gesellschaftliche Herausforderung. Genau das ist wichtig: ChatGPT ist nicht nur ein Schreibprogramm. Es erzeugt Gesprächssituationen, Nähe, Autorität, scheinbare Klugheit und manchmal sogar den Eindruck von Verständnis. Der verlinkte Beitrag gehört zur phil.COLOGNE-Veranstaltung „Ethical Intelligence: Markus Gabriel on the opportunities of the digital revolution“ am 8. Juni 2026. Dort wird Gabriels Grundthese so gerahmt: Humane KI sei kein Widerspruch, sondern eine Notwendigkeit; KI müsse nachhaltig, wertebezogen und gesellschaftlich verantwortlich gestaltet werden. Gabriel ist Professor für Erkenntnistheorie, Philosophie der Neuzeit und Gegenwart an der Universität Bonn und arbeitet ausdrücklich auch zur Philosophie und Ethik der KI. Gabriel trifft einen zentralen Punkt: ChatGPT wirkt sozial, auch wenn es kein Mensch ist. Die Universität Bonn fasst seine Position so zusammen: Systeme wie Chatbots treten zunehmend in Dialog mit Nutzerinnen und Nutzern; sie reagieren auf Stil, Situation und Kontext und imitieren zwischenmenschliche Kommunikation. Daraus folgt nach Gabriel, dass KI nicht neutral ist, sondern emotional und kulturell aufgeladen sein kann. Das ist stark. Denn die eigentliche Gefahr liegt nicht nur darin, dass ChatGPT sachlich irrt. Die tiefere Gefahr liegt darin, dass Menschen dem System mehr Urteilskraft zuschreiben, als es verantworten kann. Es spricht geordnet, freundlich, plausibel und oft sehr überzeugend. Dadurch entsteht eine Autoritätswirkung. Auch OpenAI weist darauf hin, dass ChatGPT falsche oder irreführende Antworten geben kann, einschließlich erfundener Zitate, Studien, Quellen oder übermäßig selbstsicherer Antworten bei mehrdeutigen Fragen; wichtige Informationen müssen daher geprüft werden. Damit ist Gabriels Grundwarnung keineswegs nur kulturpessimistisch. Sie entspricht einem realen methodischen Problem. Ich behaupte nicht: ChatGPT ist bereits sozialer Akteur im vollen menschlichen Sinn. Besser: ChatGPT wirkt in der Kommunikation sozial-akteurhaft, ohne selbst moralisches Subjekt zu sein. Es kann Gesprächsdynamiken auslösen, Menschen beeinflussen und Deutungen verstärken. Aber Verant-wortung, Gewissen, Schuld, Absicht und moralische Rechenschaft bleiben beim Menschen, bei Institutionen und bei den Entwicklern. Unterscheidung:

Frage	Zu grob gesagt	Besser unterschieden
Ist ChatGPT intelligent?	Ja oder nein.	Es verarbeitet Sprache leistungsfähig, aber nicht wie ein verantwortliches menschliches Bewusstsein.
Ist ChatGPT neutral?	Nein.	Es ist nie vollständig neutral, weil Training, Daten, Modellarchitektur, Nutzerfrage und Systemregeln die Antwort prägen.
Ist ChatGPT gefährlich?	Ja, grundsätzlich.	Gefährlich wird es dort, wo Menschen Prüfung, Urteilskraft und Verantwortung auslagern.
Ist ChatGPT nützlich?	Ja, sehr.	Nützlich als Denk-, Strukturierungs-, Schreib- und Recherchehilfe; problematisch als Ersatz für Wahrheit, Gewissen und Fachprüfung.

Gabriel ist wertvoll, weil er die Debatte aus der Technik-Euphorie herausholt. Er fragt nicht nur: Was

kann KI? Was macht KI mit dem Menschen? Das ist die richtige Ebene. Aber seine Formulierungen können manchmal zugespitzt wirken. Diese Zuspitzung ist philosophisch produktiv, weil sie Aufmerksamkeit schafft. Man darf sie aber nicht unkritisch übernehmen. Sonst entsteht ein Gegenmythos: Früher hieß der Mythos „KI wird alles lösen“, jetzt heißt er „KI wird uns moralisch entmündigen“. Beides ist zu einfach. Unausgesprochene Wahrheit: das Problem ist die menschliche Neigung, Verantwortung an Systeme, Autoritäten, Medien, Ideologien oder Experten abzugeben und nicht ChatGPT. KI macht dieses alte Problem schneller, eleganter und massentauglicher.

Anschluss an mein Weisheitskompass-Modell

Für meinen Ansatz bleibt Gabriel gut brauchbar, weil er genau jene Unterscheidung hervorhebt, die ich immer wieder betone: technische Leistungsfähigkeit ist nicht dasselbe wie Wahrheit; Information nicht dasselbe wie Weisheit; sprachliche Plausibilität nicht dasselbe wie verantwortete Erkenntnis. ChatGPT wird im Weisheitskompass als Resonanzraum - nicht als Orakel verstanden. Es kann ordnen, zuspitzen, vergleichen, formulieren, Gegenargumente liefern und blinde Flecken sichtbar machen. Aber es ersetzt die letzte menschliche Aufgabe nicht: wahrnehmen, prüfen, abwägen, verantworten und weiterlernen. Mein Urteil zur Leistungsfähigkeit seiner Ausführungen: Gabriel ist notwendiger Mahner. Gegen gedankenlose KI-Nutzung, nicht gegen KI grundsätzlich. Das passt gut zu meinem Grundsatz:

Glaube, Vernunft, Erfahrung und Verantwortung dürfen nicht vermengt werden,

und sollen miteinander ins Gespräch kommen.

So sollte auch ChatGPT verwendet werden.

KI als starke Hilfe – nicht als letzter Richter

Leistungsfähig in Analyse, Mustererkennung, Sprache und Unterstützung.
Verantwortung, Gewissen und Urteilskraft bleiben menschlich.

Im kritischen Horizont von Markus Gabriel wird sichtbar: KI ist ein mächtiger Spiegel und Verstärker menschlicher Möglichkeiten – aber kein Ersatz für **Wahrheit, Verantwortung und Weisheit.**

Ein Mensch bleiben | Mut nicht verlieren | Resilienz aufbauen