Im Horizont des Weisheitskompasses: Markus Gabriel und künstliche Intelligenz

Ethische Intelligenz, kulturelle Weltbilder und die Verantwortung menschlicher Urteilskraft

Dipl.-HTL-Ing. Norbert Rieser | Wissenschaftlich-didaktische Publikation | Oktober 2026

Abstract

Die Entwicklung generativer künstlicher Intelligenz berührt die Voraussetzungen menschlicher Erkenntnis, moralischer Orientierung und eigenverantwortlicher Urteilsbildung. Anhand zweier öffentlicher Gespräche mit Markus Gabriel prüfe ich, in welchem Umfang KI moralisch relevante Muster erschließen und verantwortliche Reflexionsprozesse unterstützen kann. Einen besonderen Schwerpunkt lege ich auf die kulturellen, religiösen und weltanschaulichen Voraussetzungen technischer Entwicklung. Für die Einordnung unterscheide ich ideengeschichtliche Interpretation, institutionelle Zielsetzung, empirisch überprüfbares Systemverhalten und normative Rechtfertigung. Der Weisheitskompass dient als erkenntnistheoretischer Bezugsrahmen, dessen epistemischer Pass und Brückenregeln die Übergänge zwischen diesen Aussageformen transparent halten sollen. Empirische Studien zu kulturellen Tendenzen generativer KI sowie publizierte Entwicklungsleitlinien verdeutlichen die Relevanz der Fragestellung, ohne Gabriels gesellschaftliche Prognosen zu bestätigen. Das Modell bleibt in seiner theoretischen Grundarchitektur vorläufig konsolidiert; ein eigenständiger empirischer Zusatznutzen und die Grenzen mathematischer Operationalisierung bedürfen weiterer Prüfung.

Schlüsselbegriffe: Künstliche Intelligenz · Markus Gabriel · ethische Intelligenz · Weisheitskompass · Weltanschauung · Hermeneutik · kulturelle Prägung · epistemischer Pass · Verantwortung · menschliche Urteilskraft

1. Ausgangspunkt: Eine neue Herausforderung der Urteilskraft

Nicht jede technologische Innovation verändert bereits die Grundlagen menschlicher Selbstverständigung. Die gegenwärtigen Formen generativer künstlicher Intelligenz reichen jedoch über die Automatisierung einzelner Arbeitsabläufe hinaus: Sie greifen in sprachliche Prozesse ein, in denen Menschen Informationen ordnen, Erfahrungen deuten, Gründe austauschen und Entscheidungen vorbereiten. Damit stellt sich eine Frage, die weder allein technisch noch ausschließlich ethisch beantwortet werden kann: Unter welchen Bedingungen darf die sprachliche Leistung eines künstlichen Systems als Erkenntnisbeitrag in einen verantwortbaren Entscheidungsprozess eingehen?

Markus Gabriel betrachtet KI in seinen öffentlichen Gesprächen nicht lediglich als Werkzeug. Er beschreibt sie sinngemäß als technologisch hervorgebrachte Spiegelung menschlichen Verhaltens, die durch dialogische Rückmeldungen wiederum Einfluss auf die Nutzenden gewinnt. Besonders deutlich wird dieses Motiv im Gespräch mit dem Efficiency Club sowie in der Sendung „NZZ Standpunkte“ mit Eric Gujer. Gabriels Buch „Ethische Intelligenz“ aus dem Jahr 2026 führt nach seiner veröffentlichten Selbstdarstellung die Hoffnung aus, KI könne nicht nur technischen, sondern auch moralischen Fortschritt unterstützen (Gabriel 2026; Efficiency Club 2026, 00:15–04:30; NZZ Standpunkte 2026, 09:07–14:08).

Ich untersuche Gabriels Überlegungen als eigenständige philosophische Position und beziehe sie auf die bereits entwickelte Architektur des Weisheitskompasses. Die Analyse konzentriert sich auf die Begründbarkeit von Aussagen, ihre kulturellen Voraussetzungen und die Folgen technischer Vermittlung für menschliche Urteilskraft. Unter produktiver Synthese verstehe ich die Offenlegung sachlich gerechtfertigter Beziehungen zwischen unterschiedlichen Erkenntnisweisen unter Wahrung ihrer jeweiligen Geltungsgrenzen.

2. Quellenlage, wissenschaftliche Methode und Begrenzung

Grundlage meiner Untersuchung sind zwei Gesprächstranskripte: ein Interview mit Gabriel beim Efficiency Club (ca. 15 Minuten) und ein längeres Gespräch in der Sendung „NZZ Standpunkte“ (ca. 49 Minuten im Transkript). Die zugehörigen Videolinks wurden ebenfalls bereitgestellt. Ergänzend verwende ich Gabriels veröffentlichte Beschreibung seines Buches „Ethische Intelligenz“ sowie ausgewählte institutionelle Referenzen zur KI-Ethik und zum Risikomanagement. Die Monografie selbst wurde für diesen Essay nicht vollständig ausgewertet. Deshalb schreibe ich ihr keine Argumente zu, die sich nur aus den Gesprächen oder der öffentlichen Buchbeschreibung erschließen lassen.

Die Transkripte enthalten erkennbare Fehler automatischer Spracherkennung, insbesondere bei Eigennamen und technischen Begriffen. Ich rekonstruiere den inhaltlichen Sinn gut zuordenbarer Passagen, verzichte auf die Wiedergabe fraglicher Formulierungen und unterscheide zwischen Äußerungen Gabriels, überprüfbaren Sachbehauptungen und meiner eigenen Interpretation. Prognosen und weitreichende Aussagen über die Fähigkeiten heutiger KI behandle ich als philosophische oder empirisch noch zu prüfende Positionen.

Methodisch wähle ich eine begrifflich-argumentative Analyse mit fallbezogener Veranschaulichung. Ich betreibe weder eine experimentelle Wirksamkeitsstudie noch eine unabhängige empirische Begutachtung aller technischen Aussagen der Gespräche. Daraus ergibt sich eine notwendige wissenschaftliche Selbstbegrenzung: Der Essay kann einen tragfähigen theoretischen Zusammenhang entwickeln, jedoch keinen allgemeingültigen Nachweis der Überlegenheit eines bestimmten Verfahrens erbringen.

3. Ethische Intelligenz: Die Möglichkeit einer moralisch hilfreichen KI

Gabriels Überlegung lässt sich in ihrer tragenden Absicht so rekonstruieren: Wenn KI in großen Datenbeständen Muster menschlichen Verhaltens verarbeitet und dialogisch nutzbar macht, können daraus Hinweise entstehen, die Menschen bei der Wahrnehmung moralisch relevanter Zusammenhänge unterstützen. In der NZZ-Sendung entwickelt er dies im Zusammenhang mit der Frage, wie technische Assistenzsysteme zu einem verantwortlicheren Verhalten beitragen könnten. Das Beispiel unterstützter Fahrsicherheit dient dabei als leicht zugänglicher Ausgangspunkt (NZZ Standpunkte 2026, 09:07–14:08).

Diesen Gedanken halte ich für diskussionswürdig. Schon die Hinweise auf übersehene Risiken, auf mögliche Schäden für Dritte oder auf Widersprüche zwischen geäußerten Absichten und erwartbaren Folgen können moralisch hilfreich sein. Aus der Fähigkeit eines Systems, solche Muster anzuzeigen, folgt jedoch nicht, dass es über einen unabhängig gerechtfertigten Maßstab des Guten verfügt. Eine zutreffende Beschreibung ist kein hinreichender Grund für eine normative Verpflichtung.

Gabriel diskutiert selbst Schwierigkeiten der Reduktion von Ethik auf feste algorithmische Verhaltensregeln. Diese Kritik ist von der stärkeren Behauptung zu unterscheiden, moralische Fragen ließen sich prinzipiell niemals formal bearbeiten. Ob bestimmte normative Teilfragen modellierbar sind, hängt von ihrer Struktur, ihren Voraussetzungen und dem konkreten Anwendungskontext ab. Im Weisheitskompass bleibt daher offen, welche begrenzten Prüf- und Bewertungsverfahren sich mathematisch sinnvoll operationalisieren lassen (NZZ Standpunkte 2026, 10:29–13:13).

4. Die Metapher des Spiegels und ihre erkenntnistheoretischen Grenzen

Die von Gabriel verwendete Spiegelmetapher macht auf einen bedeutsamen Zusammenhang aufmerksam: KI-Ausgaben gehen aus technisch verarbeiteten menschlichen Daten und Gestaltungsentscheidungen hervor und können ihrerseits auf individuelle Wahrnehmungen und soziale Praktiken zurückwirken. Der Dialog mit einer KI ist deshalb weder eine bloße Ablage von Informationen noch notwendig eine neutrale Auskunft. Er kann unsere Aufmerksamkeit strukturieren und Erwartungen verändern (Efficiency Club 2026, 01:16–04:30).

Die Metapher darf nicht ontologisch überdehnt werden. Aus der Erkennung emotionaler Sprachmuster folgt weder, dass ein System Emotionen erlebt, noch dass es menschliche Motive vollständig versteht. Ebenso ist eine sprachlich überzeugende Selbstbeschreibung des Systems kein verlässlicher Beleg für eigenes Bewusstsein. Gabriel unterscheidet im NZZ-Gespräch an einer Stelle selbst zwischen der Erkennung menschlicher Gefühlszustände und deren subjektivem Erleben (NZZ Standpunkte 2026, 20:46–21:49).

Für den Weisheitskompass ergibt sich eine doppelte Pflicht: Ich muss erstens prüfen, was ein KI-System tatsächlich zu erkennen vermag; zweitens darf ich seine Art der Informationsverarbeitung nicht unbemerkt mit menschlichem Erleben oder eigenverantwortlichem Handeln gleichsetzen. Die Hermeneutik kann solche Bedeutungsverschiebungen sichtbar machen, während die empirische Prüfung über die Leistungsgrenzen einzelner Systeme Auskunft geben muss.

5. Kreativität, Gründe und die Differenz zwischen sprachlicher Leistung und Erkenntnis

Im Gespräch mit Eric Gujer unterscheidet Gabriel zwischen neuartigen sprachlichen Kombinationen und einem gedanklichen Fortschritt, der neue, begründbare Positionen in die philosophische Diskussion einführt. Er würdigt die außerordentliche Leistungsfähigkeit der KI beim Zusammenfassen, Vergleichen und sprachlichen Kombinieren, lässt jedoch offen, in welchem Sinne daraus eigenständiges philosophisches Denken erwächst (NZZ Standpunkte 2026, 01:43–04:07).

Hier sehe ich eine erkenntnistheoretische Frage. Eine neue Formulierung kann inhaltlich belanglos sein; eine vertraut klingende Formulierung kann dagegen einen wichtigen Gedanken verständlich machen. Neuheit, sprachliche Überzeugungskraft und Wahrheit sind daher verschiedene Eigenschaften. Auch ein zutreffender Satz gewinnt seine wissenschaftliche Bedeutung nicht allein aus seiner Herkunft, sondern aus den Gründen und Verfahren, durch die er gestützt wird.

Daraus folgt weder die prinzipielle Unfähigkeit künstlicher Systeme zu originellen Entdeckungen noch ihre Gleichstellung mit menschlicher Urteilskraft. Solche generellen Festlegungen wären nach dem gegenwärtigen Wissensstand zu weitreichend. Für meinen Ansatz genügt die präzisere Aussage: Der Geltungsanspruch eines Ergebnisses ist unabhängig vom Eindruck seiner sprachlichen Brillanz zu prüfen. Der Weisheitskompass richtet die Aufmerksamkeit deshalb auf Begründungsstruktur, Geltungsbereich, Unsicherheit und mögliche Gegenargumente.

6. Der Weisheitskompass als methodische Prüfarchitektur

Die von mir entwickelte Grundarchitektur unterscheidet empirisch-wissenschaftliche, hermeneutisch-interpretative, ethisch-normative und praktisch-entscheidungstheoretische Perspektiven. Diese Perspektiven werden nicht als austauschbare Bestandteile einer einzigen Messskala verstanden. Ihre Verbindung setzt vielmehr voraus, dass die jeweilige Art des Geltungsanspruchs erkennbar bleibt und Übergänge ausdrücklich gerechtfertigt werden.

Der epistemische Pass dient der Kennzeichnung von Aussageart, Quellenbasis, Geltungsbereich und Unsicherheit. Brückenregeln bestimmen, welche zusätzlichen Voraussetzungen für einen Übergang von einer Aussageform zu einer anderen erforderlich sind. Wirk- und Vermittlungszusammenhänge betreffen dagegen die sachlichen Prozesse, die ein Ergebnis ermöglichen oder beeinflussen. Diese Unterscheidung soll verhindern, dass etwa eine empirische Leistungssteigerung unmittelbar als moralisch gebotene Handlung erscheint.

Die epistemisch-hermeneutische Integrationsstruktur lässt sich heuristisch wie folgt darstellen:

$$E(F) = \Gamma(\Pi(F), B, M, K, U)$$

Dabei bezeichnet F den Fall, Π(F) die perspektivischen Erschließungen, B die Brückenregeln, M die Wirk- und Vermittlungszusammenhänge, K die Kontextbedingungen und U die Unsicherheits- und Revisionsbedingungen. Γ bezeichnet einen methodischen Integrationsvorgang, keine empirisch validierte Berechnungsfunktion. Ihr Ergebnis E(F) kann in einer transparenten Begründungsstruktur bestehen, auch wenn eine eindeutige Handlungsempfehlung nicht möglich ist.

KI kann in dieser Architektur als Hilfsmittel bei der Sammlung von Hypothesen, beim Vergleich möglicher Deutungen oder beim Sichtbarmachen von Widersprüchen eingesetzt werden. Sie erhält dadurch jedoch keine automatische Zuständigkeit für die Festsetzung normativer Maßstäbe. Selbst eine formell korrekte Aggregation von Bewertungen wäre noch keine Berechnung von Weisheit. Für moralisch schwerwiegende Entscheidungen bleiben nicht kompensierbare Schutz- und Rechtfertigungsanforderungen zu berücksichtigen.

7. Menschliche Selbstbestimmung, emotionale Ansprache und epistemische Abhängigkeit

Die dialogische Anschlussfähigkeit moderner KI eröffnet didaktische und reflexive Möglichkeiten. Dieselbe Eigenschaft kann jedoch neue Formen einseitiger Bindung und Beeinflussung begünstigen. Gabriel weist in beiden Gesprächen darauf hin, dass Menschen KI-Systemen zunehmend persönliche Erfahrungen anvertrauen und die Systeme durch ihre Antworten auf Selbstdeutung und Verhalten zurückwirken können (Efficiency Club 2026, 01:16–03:29; NZZ Standpunkte 2026, 20:46–21:49).

Für mich besteht die Gefahr nicht allein in einer falschen Information. Problematisch wäre eine Gesprächssituation, in der ein technisches System unbemerkt festlegt, welche Fragen als wesentlich gelten, welche Optionen plausibel erscheinen und welche Wertordnung der Entscheidung zugrunde liegt. Eine scheinbar unterstützende Antwort kann dann die eigenständige Begründung durch eine autoritativ wirkende Empfehlung ersetzen.

Daraus leite ich keine Forderung nach einem pauschalen Verzicht auf KI ab. Vielmehr halte ich eine reflektierte Nutzung für erforderlich: entscheidungsrelevante Aussagen sind gegebenenfalls anhand unabhängiger Quellen zu prüfen; persönliche Angaben sind datensparsam zu behandeln; alternative Perspektiven dürfen nicht durch eine einzige Modellantwort verdrängt werden. Die UNESCO-Empfehlung zur Ethik künstlicher Intelligenz betont ausdrücklich menschliche Aufsicht und die Zurechenbarkeit ethischer und rechtlicher Verantwortung zu Personen oder Institutionen (UNESCO 2021, Abschnitte 35–36).

8. Ein exemplarischer Fall: KI-Unterstützung bei Bildungsentscheidungen

Ich betrachte einen hypothetischen Bildungsfall. Eine Schule erwägt, ein KI-basiertes Lernassistenzsystem dauerhaft einzusetzen, weil erste Rückmeldungen eine höhere Lernmotivation und raschere Bearbeitung einzelner Aufgaben nahelegen. Eine rein technische Prüfung könnte sich auf Antwortqualität, Kosten, Zuverlässigkeit und Bedienbarkeit konzentrieren. Daraus ergäbe sich noch kein hinreichend begründetes pädagogisches Urteil.

Die empirische Perspektive muss klären, ob sich die beobachteten Verbesserungen unter kontrollierten Bedingungen bestätigen lassen und für welche Lernenden sie gelten. Die hermeneutische Perspektive hätte zu untersuchen, was Lehrende, Lernende und Eltern unter selbstständigem Lernen, Unterstützung und Leistung verstehen. Die normative Prüfung müsste Fragen nach Chancengleichheit, Datenschutz, Transparenz, möglicher Abhängigkeit und der Zumutbarkeit eines verpflichtenden Einsatzes behandeln. Die praktische Perspektive würde Ressourcenbedarf, Fortbildung, Aufsichtsmöglichkeiten und Reversibilität der Einführung untersuchen.

Ein KI-System kann zu jedem dieser Gesichtspunkte Informationen beitragen, aber der Übergang von einem beobachteten Lernerfolg zur Empfehlung einer flächendeckenden Einführung wäre rechtfertigungsbedürftig. Die Brückenregel müsste etwa offenlegen, dass kurzfristige Leistungsgewinne im gewählten Kontext als relevantes, aber nicht alleiniges Bildungsziel gelten. Die Entscheidung könnte deshalb auf eine begrenzte, freiwillige Erprobungsphase mit vorab festgelegten Evaluationskriterien hinauslaufen. Auch diese Empfehlung wäre jedoch

aus den konkreten Randbedingungen zu begründen, nicht aus der bloßen Anwendung des Weisheitskompasses abzuleiten.

Der Fall zeigt, worin ich den möglichen Nutzen der Verbindung sehe: KI kann den Raum der verfügbaren Argumente erweitern; der Weisheitskompass kann ihre unterschiedlichen Geltungsansprüche kontrollieren und verhindern, dass die Qualität der Entscheidungsbegründung hinter der sprachlichen Plausibilität eines Vorschlags verschwindet.

9. Kulturelle Weltbilder, religiöse Deutungen und die normative Architektur künstlicher Intelligenz

Besonders aufschlussreich erscheint mir Gabriels Hinweis auf die geistesgeschichtlichen Voraussetzungen der KI-Entwicklung. Im Gespräch mit dem Efficiency Club erörtert er, inwieweit religiöse Traditionen, theologische Denkfiguren und philosophische Menschenbilder die Vorstellungen gegenwärtiger Technologieakteure prägen können. Er verbindet diese Deutung mit Prognosen über gesellschaftliche Machtverschiebungen und neue Konfliktformen (Efficiency Club 2026, 05:31–12:30). Für den Weisheitskompass eröffnet sich damit ein Untersuchungsfeld, das die Herkunft technischer Zielsetzungen ebenso berücksichtigt wie die Rechtfertigung ihrer praktischen Folgen.

Gabriels Ausführungen enthalten verschiedene Aussagearten, deren Status ich auseinanderhalte. Eine religionsphilosophische Analogie kann interpretativ erhellend sein; Aussagen über konkrete Einflüsse einzelner Institutionen auf Unternehmen benötigen unabhängige Quellen. Noch weiter reichen Behauptungen über die Ursachen militärischer Konflikte oder den bereits eingetretenen Verlust gesellschaftlicher Steuerungsfähigkeit. Für solche Folgerungen bieten die Gespräche allein noch keinen hinreichenden empirischen Nachweis. Wo eine Deutung große Erklärungskraft beansprucht, wächst die Verpflichtung, ihre Bedingungen sorgfältig offenzulegen.

Der Weisheitskompass kann hier eine begründungsreflexive Aufgabe übernehmen: Er dokumentiert die Aussageform, prüft die Tragfähigkeit der Quellen und bestimmt, welche Schlüsse aus kulturellen, institutionellen oder technischen Befunden tatsächlich gezogen werden dürfen. Die öffentlich verfügbaren Vorgaben zur Ausrichtung von KI-Modellen und die Forschung zu kulturellen Unterschieden in generierten Antworten schaffen hierfür überprüfbare Anknüpfungspunkte. Ihre Ergebnisse erfordern eine fallbezogene Interpretation; ein allgemeines Kultururteil über ein ganzes Modell, Unternehmen oder Herkunftsland wäre wissenschaftlich voreilig.

9.1. Technische Entwicklung und menschliche Wertentscheidungen

Bereits die Bestimmung eines Entwicklungsziels beruht auf menschlichen Entscheidungen. Datenauswahl, Modellarchitektur, Rückmeldungsverfahren, Sicherheitsregeln und Nutzungsvorgaben bilden ein Geflecht technischer und institutioneller Voraussetzungen. In diesem Geflecht können Vorstellungen über Sicherheit, Selbstbestimmung, gesellschaftliches Wohlergehen und angemessene Kommunikation wirksam werden. Zugleich entfalten wirtschaftliche Anreize, gesetzliche Anforderungen und organisatorische Routinen ihren Einfluss. Ich halte es daher für geboten, die normative Ausrichtung eines KI-Systems als Ergebnis mehrerer unterscheidbarer Gestaltungsschritte zu untersuchen.

Die persönliche Überzeugung eines Entwicklers beweist zunächst wenig über das Verhalten eines fertigen Modells. Zwischen Absicht und Ergebnis liegen Datengrundlagen, technische Implementierung, Qualitätsprüfungen und weitere institutionelle Entscheidungen. Eine plausible ideengeschichtliche Erklärung ersetzt keine Untersuchung dieser Vermittlungen. Der Weisheitskompass gewinnt gerade dort an Schärfe, wo er die notwendige Verbindung zwischen Interpretation und empirischem Befund ausdrücklich verlangt. Die kulturelle Einbettung technischer Systeme wird auf diese Weise zu einer begründbaren Forschungsfrage.

9.2. Theologie und Technik: Aussagekraft einer ideengeschichtlichen Perspektive

Die religiöse Deutung technischer Zukunftserwartungen hat eine lange ideengeschichtliche Vorgeschichte. Hoffnungen auf die Überwindung menschlicher Begrenztheit, die Vorstellung einer umfassenden Wissensmacht oder die Erwartung einer durch Technik erneuerten Gesellschaft lassen sich mit überlieferten Motiven von Erlösung, Fortschritt und Vollkommenheit vergleichen. Solche Parallelen können ein tieferes Verständnis der Symbolik technologischer Programme ermöglichen. In Gabriels Überlegungen sehe ich daher einen Anstoß, die Begriffswelt technologischer Selbstbeschreibungen philosophisch genauer zu untersuchen.

Eine Analogie bleibt auf den nachgewiesenen Vergleichsbereich begrenzt. Religiöse Sprache in Äußerungen von Unternehmensvertretern belegt deren Ausdrucksweise; ob sie konkrete Entwicklungsentscheidungen beeinflusst, verlangt weitere Belege. Ebenso wenig lässt sich die Entstehung komplexer KI-Systeme aus theologischen Motiven allein erklären. Es erscheint mir sachgerechter, religiöse Traditionen als mögliche Bestandteile eines umfassenderen gesellschaftlichen Deutungsrahmens zu behandeln. Ihre Erklärungskraft ist jeweils an überprüfbare Quellen, Entwicklungsprozesse und Handlungszusammenhänge zurückzubinden.

9.3. Kulturelle Vielfalt und die Gefahr pauschaler Zuschreibungen

Gabriels Gegenüberstellung europäischer und asiatischer Denkströmungen erinnert daran, dass Freiheit, Verantwortung und soziale Ordnung in unterschiedlichen philosophischen Traditionen verschieden akzentuiert werden. Diese Perspektive verdient Aufmerksamkeit, verlangt jedoch besondere methodische Sorgfalt. Weder Europa noch China, Japan oder ein anderer kultureller Raum bildet eine weltanschaulich homogene Einheit. Innerhalb einer Gesellschaft stehen zahlreiche religiöse, säkulare, politische und philosophische Positionen miteinander im Austausch. Die geografische Herkunft eines KI-Systems lässt daher für sich genommen keinen verlässlichen Schluss auf eine bestimmte moralische Auffassung zu.

Einen empirischen Anknüpfungspunkt liefert die Untersuchung von Lu, Song und Zhang (2025). Sie berichtet bei untersuchten generativen KI-Systemen Unterschiede der sozialen Orientierung und des kognitiven Stils zwischen chinesisch- und englischsprachigen Anwendungskontexten. Damit ist die Relevanz sprachlicher und kultureller Einflüsse wissenschaftlich unterlegt. Die Befunde gelten für die jeweils untersuchten Modelle, Messinstrumente und Versuchsbedingungen. Eine ontologische Zuschreibung menschlicher Überzeugungen an Sprachmodelle oder eine umfassende Bewertung ganzer Kulturräume wäre daraus nicht gerechtfertigt.

9.4. Die normative Gestaltung von KI am Beispiel veröffentlichter Leitlinien

Konkreter als allgemeine Aussagen über Weltbilder sind publizierte Gestaltungsgrundsätze einzelner Anbieter. Anthropic veröffentlichte im Januar 2026 eine überarbeitete »Claude's Constitution«, in der erwünschte Verhaltensweisen, Zielkonflikte und leitende normative Vorstellungen ausführlich beschrieben werden (Anthropic 2026). Ein solches Dokument belegt, dass moralisch relevante Anforderungen ausdrücklich in den Entwicklungsprozess einfließen sollen. Es ermöglicht außerdem eine öffentliche Auseinandersetzung mit der Auswahl und Begründung der verwendeten Werte.

Die Veröffentlichung einer Verhaltensordnung begründet noch keine Übereinstimmung zwischen formuliertem Ideal und tatsächlichem Modellverhalten. Dafür bedarf es nachvollziehbarer Evaluationen, geeigneter Tests und gegebenenfalls unabhängiger Prüfungen. Ebenso bleiben Fragen nach der Legitimation jener Instanzen bestehen, die Regeln entwerfen, Zielkonflikte gewichten und Änderungen beschließen. Hier erkenne ich einen besonders fruchtbaren Bezug zum Weisheitskompass: Die technische Implementierung, die institutionelle Verantwortung und die normative Geltung müssen als eigenständige Prüfgegenstände erkennbar bleiben.

9.5. Algorithmische Autorität und menschliche Urteilskraft

Dialogfähige KI-Systeme können zunehmend in Bereichen wirken, in denen Menschen bisher vor allem anderen Menschen vertrauten: bei der Einordnung persönlicher Schwierigkeiten, beim Lernen, bei beruflichen Überlegungen und bei der Auseinandersetzung mit Wertkonflikten. Die sprachliche Geschlossenheit einer Antwort kann dabei den Eindruck begründeter Gewissheit erwecken. Gerade diese Wirkung macht die Frage nach der Autorität technischer Auskünfte dringlich. Eine plausible Darlegung sollte nachvollziehbar bleiben, ihre Quellen offenlegen und Gegenargumente zulassen.

Ich sehe das Risiko in einer schleichenden Verlagerung von Begründungskompetenz. Ein technisches System beeinflusst möglicherweise bereits durch seine Auswahl relevanter Fragen und Möglichkeiten die Richtung der anschließenden Überlegung. Menschliche Selbstbestimmung setzt dann voraus, dass die Kriterien dieser Auswahl hinterfragt werden dürfen und alternative Informationswege zugänglich bleiben. Die prinzipielle Möglichkeit eigenständiger Kritik gewinnt damit einen hohen normativen Stellenwert. Der Weisheitskompass kann diese Prüfung methodisch unterstützen, indem er Erkenntnisansprüche, Interpretationen und praktische Empfehlungen deutlich kennzeichnet.

9.6. Technologische Konflikte zwischen Ideengeschichte und empirischer Erklärung

Gabriel greift in seinem Gespräch weitreichende historische Analogien auf, um die politische Bedeutung künstlicher Intelligenz zu erläutern. Er spricht sinngemäß von religiös geprägten technologischen Konflikten und möglichen Verschiebungen gesellschaftlicher Machtverhältnisse. Solche Deutungen können Beobachtungen bündeln und bislang vernachlässigte Fragen eröffnen. Ihr empirischer Status bleibt im Einzelfall zu bestimmen. Insbesondere erklären religiöse Begriffe allein weder militärische Konflikte noch wirtschaftliche Konzentrationsprozesse oder die Durchsetzung bestimmter Technologien.

Eine verantwortliche Untersuchung muss zunächst die jeweils beteiligten Akteure, ihre nachgewiesenen Interessen, die technischen Strukturen und die institutionellen Entscheidungsmöglichkeiten bestimmen. Erst danach ließe sich prüfen, ob religiöse oder weltanschauliche Motive einen eigenständigen erklärenden Beitrag leisten. Ich würde daher zwischen der Sprache des Konflikts, seiner ideengeschichtlichen Deutung und den empirisch untersuchbaren Ursachen unterscheiden. Dies entspricht dem Anspruch meines Modells, Erkenntnisformen in ihren zulässigen Beziehungen zu verbinden und übergreifende Erklärungen an ihren tatsächlichen Voraussetzungen zu messen.

9.7. Gemeinsame normative Maßstäbe unter kulturell unterschiedlichen Voraussetzungen

Die Anerkennung kultureller Vielfalt wirft zugleich die Frage nach den Möglichkeiten gemeinsamer normativer Orientierung auf. Gesellschaften können verschiedene Vorstellungen eines gelungenen Lebens entwickeln und dennoch Anforderungen an Menschenwürde, Freiheit und die Begrenzung schädlicher Folgen teilen. Eine tragfähige Begründung solcher Maßstäbe verlangt die Auseinandersetzung mit unterschiedlichen Traditionen und Perspektiven. Ihre rechtliche Verankerung und ihre philosophische Rechtfertigung sind dabei gesondert zu betrachten. Internationale Orientierungen wie die UNESCO-Empfehlung zur KI-Ethik bieten hierfür einen relevanten normativen Bezugspunkt (UNESCO 2021).

Der Weisheitskompass soll diese Diskussion weder durch eine vorweggenommene Hierarchie der Kulturen entscheiden noch durch die bloße Addition unterschiedlicher Wertvorstellungen ersetzen. Ich halte die offene Prüfung von Geltungsansprüchen für den angemessenen Weg. Dabei sind gemeinsame Schutzanforderungen ebenso zu berücksichtigen wie legitime Unterschiede ihrer Begründung und konkreten Anwendung. Wo Konflikte bestehen bleiben, sollten sie erkennbar dokumentiert werden. Verantwortete Orientierung kann auch darin liegen, den verbleibenden Dissens präzise zu benennen.

9.8. Ein Prüfprogramm für den Weisheitskompass

Für eine konkrete Fallanalyse würde ich zunächst erheben, welche Aussagen zur kulturellen Prägung eines KI-Systems vorliegen und worauf sie sich stützen. Anschließend wäre zu unterscheiden, ob eine Aussage als historische Interpretation, dokumentierte Unternehmensposition, empirischer Systembefund oder normative Bewertung vorgetragen wird. Danach müssten die vermuteten Übergänge geprüft werden: Welche belegten Entscheidungen verbinden eine weltanschauliche Vorstellung mit einer technischen Vorgabe? Wie zeigt sich diese Vorgabe im Systemverhalten? Welche Grenzen gelten für die Aussagekraft einzelner Tests?

Den Abschluss bildet eine normative und praktische Prüfung: Wer verantwortet die gesetzten Ziele, welche Beteiligten sind davon betroffen, welche Rechte oder Interessen werden berührt, und unter welchen Bedingungen wäre eine Revision geboten? Als ergänzende Instrumente kommen technische Evaluationen, dokumentierte Governance-Prozesse und risikobezogene Rahmenwerke wie das AI Risk Management Framework des NIST in Betracht (Tabassi 2023). Die spezifische Leistung des Weisheitskompasses bestünde in einer zusätzlichen Kontrolle der Begründungsübergänge. Ob diese Vorgehensweise die Qualität der Analyse gegenüber etablierten Verfahren tatsächlich verbessert, bleibt Gegenstand späterer vergleichender Untersuchungen.

10. Offene Forschungsfragen und wissenschaftliche Selbstbegrenzung

Die Verbindung der kulturphilosophischen Überlegungen Gabriels mit dem Weisheitskompass eröffnet ein präziser bestimmbares Forschungsprogramm. Zu prüfen wäre etwa, ob die systematische Kennzeichnung von Aussageformen unbegründete Schlüsse zwischen religiöser Interpretation, dokumentierter Entwicklungsabsicht und empirischem Modellverhalten zuverlässiger erkennen lässt. Ebenso interessiert mich, ob offengelegte Brückenregeln die Transparenz von Wertabwägungen erhöhen und eine fundiertere Korrektur früherer Einschätzungen ermöglichen. Dafür bedarf es geeigneter Fallstudien und Vergleichsverfahren. Die hier entwickelte Synthese ist methodologisch begründet, ihre behauptbaren praktischen Vorteile bleiben empirisch offen.

Ein zweites Forschungsfeld betrifft das Verhältnis qualitativer Geltungsprüfung und möglicher mathematischer Operationalisierung. Normative Bewertungen können unter bestimmten Voraussetzungen formal strukturiert werden; aus einer numerischen Rangordnung entsteht jedoch noch keine abschließende ethische Rechtfertigung. Auch die kulturelle Prägung von KI lässt sich nicht auf einen einzigen Messwert zurückführen. Eine methodisch angemessene Weiterentwicklung könnte daher qualitative Begründungsverfahren mit begrenzten technischen Evaluationsinstrumenten verbinden. Die hierfür nötigen Modellannahmen und Validierungsverfahren sind erst auszuarbeiten.

Ich betrachte die bisherige Architektur des Weisheitskompasses als vorläufig konsolidiert. Seine theoretischen Bestandteile bleiben prüfbar und revidierbar. Die kulturphilosophische Vertiefung macht eine neue Grunddimension entbehrlich; sie erschließt einen weiteren Anwendungsbereich des epistemischen Passes, der Brückenregeln und der normativen Reflexion. Eine eigenständige Wirksamkeit meines Vorgehens ist weiterhin nachzuweisen. An diesem Maßstab will ich seine zukünftige Entwicklung ausrichten.

11. Schluss: Ethische KI als Anlass zur Vertiefung menschlicher Urteilskraft

Markus Gabriels Überlegungen zur ethischen Intelligenz eröffnen einen wichtigen Zugang zu den moralischen Folgen dialogfähiger KI. Besonders sein Blick auf kulturelle und religiöse Deutungsmuster lenkt die Aufmerksamkeit auf Voraussetzungen, die innerhalb technischer Leistungsdebatten leicht in den Hintergrund treten. Die empirische Forschung zu kulturellen Antworttendenzen und die veröffentlichten Leitlinien von KI-Anbietern zeigen, dass diese Fragestellung eine konkrete Untersuchung verdient. Die weitreichenden

gesellschaftlichen Deutungen Gabriels behandle ich zugleich mit jener methodischen Zurückhaltung, die mein eigener Ansatz verlangt.

Im Weisheitskompass sehe ich eine Möglichkeit, empirische Befunde, hermeneutische Interpretationen, normative Maßstäbe und praktische Entscheidungsvoraussetzungen kontrolliert aufeinander zu beziehen. Dabei bleibt die menschliche und institutionelle Verantwortung für Auswahl, Geltungsprüfung und Folgen technischer Anwendungen ausdrücklich erhalten. Meine Untersuchung führt zu der Frage, welche Wertvorstellungen tatsächlich wirksam werden, durch welche Entscheidungen sie in Systeme gelangen und unter welchen Bedingungen ihre normative Legitimität öffentlich überprüft werden kann.

Die wissenschaftliche Tragfähigkeit künstlicher Intelligenz bemisst sich auch an der Offenheit ihrer Erkenntnisgrundlagen, an der Prüfbarkeit ihrer normativen Voraussetzungen und an der Möglichkeit wirksamer Kritik. Der Weisheitskompass beansprucht für diese Fragen eine begründungsreflexive Funktion. Seine Integrationsleistung wird sich daran erweisen müssen, ob er die Beziehungen des Unterschiedlichen präziser erschließt, ohne die Grenzen einzelner Erkenntnisformen zu verwischen. Diesen Anspruch verstehe ich als Verpflichtung zu weiterer Prüfung und als Ausgangspunkt eines weiterhin offenen Forschungsprogramms.

Quellen- und urheberrechtlicher Hinweis

Der Essay ist eine eigenständige, sinngemäße wissenschaftliche Auseinandersetzung mit den unten angeführten Quellen. Aus den Gesprächen wurden keine längeren Wortpassagen übernommen; die vollständigen Transkripte, Screenshots und geschützten Videoabbildungen werden nicht reproduziert. Die Zuordnung zu den Gesprächsstellen erfolgt anhand der Zeitmarken der bereitgestellten Arbeits-Transkripte. Diese Transkripte enthalten Verschriftlichungsfehler und sind nicht als autorisierte Wortprotokolle zu behandeln.

Die Ausführungen zu Gabriels Buch stützen sich auf seine öffentliche Buchbeschreibung und die Gespräche, nicht auf eine vollständige eigenständige Textanalyse der Monografie. Die rechtlichen Rahmenbedingungen wissenschaftlicher Zitate ergeben sich in Österreich insbesondere aus § 42f UrhG; für übernommene Werkstellen gelten Quellenanforderungen nach § 57 UrhG. Da hier auf lange Zitate verzichtet wird, bleibt die Darstellung auf paraphrasierende wissenschaftliche Auseinandersetzung und präzise Attribution ausgerichtet.

Die sprachliche und redaktionelle Ausarbeitung dieses Essays erfolgte unter Mitwirkung generativer KI. Die Prüfung der Quellen, die abschließende sachliche Kontrolle sowie die Freigabe zur Veröffentlichung obliegen dem angegebenen Autor. Die dargestellten Schlussfolgerungen stehen unter dem Vorbehalt der ausgewiesenen Quellen- und Erkenntnisgrenzen.

Literatur und verwendete Quellen

Gabriel, Markus (2026): Ethische Intelligenz. Wie KI uns moralisch weiterbringen kann. Berlin: Ullstein. ISBN 978-3-550-20461-6. Buchangaben und Verlagsinformation: https://markus-gabriel.com/ethische-intelligenz-2026

NZZ Standpunkte (2026): Intelligenz aus der Maschine: Ist KI Frankensteins neues Monster? Gespräch von Eric Gujer mit Markus Gabriel, SRF/NZZ, Erstausstrahlung 26. September 2026. Video: https://www.youtube.com/watch?v=uPJO4URd0w8 ; offizielle SRF-Seite: https://www.srf.ch/play/tv/nzz-standpunkte/video/intelligenz-aus-der-maschine-ist-ki-frankensteins-neues-monster-staffel-2026-folge-8?urn=urn:srf:video:b3ccf955-a053-488e-80ee-978f98eaec9c. Arbeitsgrundlage: bereitgestelltes „Transkript 2“.

Efficiency Club (2026): KI wird immer mächtiger. Doch wer entscheidet, welche Grenzen sie niemals überschreiten darf? Gespräch mit Markus Gabriel, September 2026. Video: https://www.youtube.com/watch?v=nR_xj59mm7E. Arbeitsgrundlage: bereitgestelltes „Transkript 1“.

UNESCO (2021): Recommendation on the Ethics of Artificial Intelligence. Paris, verabschiedet am 23. November 2021. https://www.unesco.org/en/legal-affairs/recommendation-ethics-artificial-intelligence

Tabassi, Elham (2023): Artificial Intelligence Risk Management Framework (AI RMF 1.0). NIST AI 100-1. https://doi.org/10.6028/NIST.AI.100-1

Lu, Jackson G.; Song, Lesley Luyang; Zhang, Lu Doris (2025): Cultural tendencies in generative AI. Nature Human Behaviour 9, 2360–2369. https://doi.org/10.1038/s41562-025-02242-1

Anthropic (2026): Claude’s Constitution. Veröffentlichung einer überarbeiteten Modellverfassung, Januar 2026. https://www.anthropic.com/constitution ; Erläuterung: https://www.anthropic.com/news/claude-new-constitution

Republik Österreich: Urheberrechtsgesetz, § 42f (Zitate) und § 57 (Quellenangaben), jeweils geltende Fassung. https://www.ris.bka.gv.at/NormDokument.wxe?Abfrage=Bundesnormen&Gesetzesnummer=10001848&Paragraf=42f ; https://ris.bka.gv.at/NormDokument.wxe?Abfrage=Bundesnormen&Gesetzesnummer=10001848&Paragraf=57

Der Weisheitskompass im Horizont von Markus Gabriel und Künstlicher Intelligenz

Kulturelle Voraussetzungen, ethische Intelligenz und verantwortete Urteilskraft

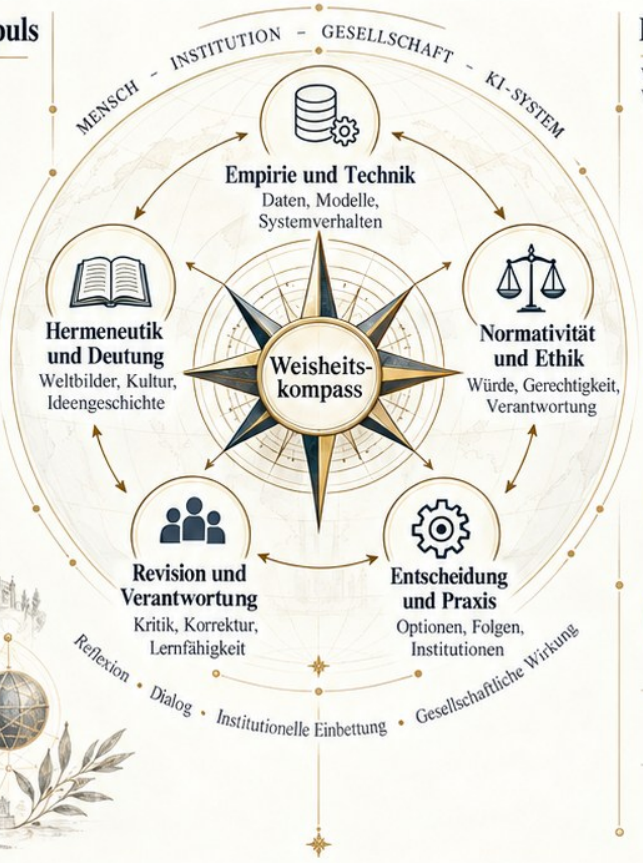
Künstliche Intelligenz ist nicht nur technische Funktion, sondern auch Ausdruck kultureller, normativer und institutioneller Voraussetzungen.

Kulturphilosophischer Impuls

Im Horizont von Markus Gabriel

1. Technische Systeme entstehen nicht weltanschaulich voraussetzungslos.
2. KI reflektiert Auswahlentscheidungen, Trainingsdaten und institutionelle Zielsetzungen.
3. Religiöse und ideengeschichtliche Deutungen können Orientierung bieten, ersetzen aber keine empirische Prüfung.
4. Die Frage lautet: Welche Wertvorstellungen werden in KI praktisch wirksam?





Ethische Intelligenz und KI

Von der Datenbasis zur gesellschaftlichen Wirkung

Trainingsdaten
Auswahl, Repräsentativität, implizite Wertungen

Modellziele
Optimierungsgrößen, Normvorstellungen, institutionelle Vorgaben

Regelwerke / Alignment
Leitprinzipien, Sicherheitsmechanismen, normative Rahmenbedingungen

Systemantworten
Generierte Inhalte, Empfehlungen, Handlungsoptionen

Menschliche Rezeption
Interpretation, Akzeptanz, Anwendung, gesellschaftliche Folgen

Sprachliche Plausibilität ist nicht identisch mit normativer Geltung.

Orientierung verlangt Prüfung von Voraussetzungen, Geltungsansprüchen und Folgen.

Formale Fundierung des Weisheitskompasses

I Integrierte Urteilsbildung

Verknüpfung von Perspektiven, Regeln, Kontexten und Unsicherheit

$$E(F) = \Gamma(\Pi(F), B, M, K, U)$$

Notation:

F	=	Fall
$\Pi(F)$	=	perspektivische Erschließung des Falles
B	=	Brückenregeln
M	=	Wirk- und Vermittlungszusammenhänge
K	=	Kontextbedingungen
U	=	Unsicherheits- und Revisionsbedingungen
$E(F)$	=	integriertes Urteil

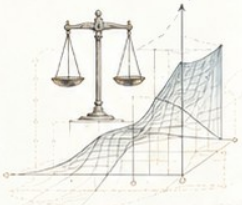
II Bewertung von Handlungsalternativen

Abwägung von Nutzen, Risiken und Wechselwirkungen

$$V(a|F,\theta) = \sum_{i=1}^{n} w_i \cdot v_i(a|F) - \lambda \cdot \rho(a|F) + \mu \cdot \sigma(a|F)$$

Notation:

a	=	Handlungsalternative
θ	=	Modellannahmen und Parameter
w_i	=	Gewicht der Dimension i
$v_i(a\|F)$	=	dimensionsbezogene Bewertung
$\rho(a\|F)$	=	Risikomaß
$\sigma(a\|F)$	=	Wechselwirkungs- / Synergieterm
λ, μ	=	Steuerungsparameter
$V(a\|F,\theta)$	=	bewertete Entscheidungslage



Zentrale Einsicht

1. KI kann Erkenntnis und Reflexion erweitern, aber Verantwortung nicht ersetzen.
2. Weltanschauliche Prägungen müssen kenntlich gemacht und kritisch geprüft werden.
3. Der Weisheitskompass verbindet Deutung, Geltungsprüfung und Entscheidung.
4. Menschliche Urteilskraft bleibt die letzte Instanz verantworteter Orientierung.

Heuristisches Rahmenmodell – keine universale Ethik und kein automatisches Entscheidungsverfahren

ANHANG | WISSENSCHAFTLICHE FALLANALYSE

Urteilskraft unter Extrembedingungen

Verantwortung, Loyalität und Gewissen im Horizont des Weisheitskompasses

Dipl.-HTL-Ing. Norbert Rieser | Wissenschaftlich-didaktische Publikation | Oktober 2026

Abstract

Durch den Science-Fiction-Thriller I.S.S. angeregt, untersuche ich, welche Anforderungen an verantwortliche Urteilskraft entstehen, wenn institutionelle Befehle, persönliche Loyalitäten und die Pflicht zum Schutz menschlichen Lebens unter Bedingungen äußerster Unsicherheit miteinander kollidieren. Die dokumentierte Ausgangshandlung des Films – eine internationale Raumstationsbesatzung erhält während eines Krieges gegensätzliche Kontrollbefehle – dient als begrenztes philosophisches Gedankenexperiment. Ich rekonstruiere daraus Konflikte zwischen Geltungsansprüchen, epistemischen Unsicherheiten und individueller Gewissensverantwortung. Der Bezug zur künstlichen Intelligenz entsteht aus einer strukturellen Analogie: Auch algorithmisch vermittelte Empfehlungen können unter Zeitdruck als Autorität auftreten, obwohl ihre Voraussetzungen und normativen Prioritäten prüfungsbedürftig bleiben. Die vorgeschlagenen Gleichungen explizieren Beziehungen zwischen Fallerschließung, Werten, Entscheidung und Revision; sie sind heuristische Notationen ohne empirisch validierten Berechnungsanspruch. Die Analyse verwendet eigene Formulierungen und ausschließlich referenzierende Filminformationen. Filmbilder, Plakatmaterial, Dialoge oder längere fremde Textpassagen werden nicht übernommen.

Schlüsselbegriffe: Weisheitskompass · I.S.S. · praktische Urteilskraft · Gewissen · Loyalität · Verantwortung · epistemische Unsicherheit · künstliche Intelligenz · Krisenentscheidung

1. Gegenstand und methodische Begrenzung

Die moralische Bedeutung einer Entscheidung tritt besonders deutlich hervor, sobald vertraute Orientierungen ihre vermeintliche Eindeutigkeit verlieren. In technischen und organisatorischen Abläufen ermöglichen Zuständigkeiten, Handlungsanweisungen und gesicherte Informationen gewöhnlich eine gewisse Verlässlichkeit. Eine Krise kann diese Bedingungen innerhalb kurzer Zeit verändern. Genau hier beginnt meine Untersuchung: Welche Gründe dürfen unter außergewöhnlichem Entscheidungsdruck Geltung beanspruchen, und wie lässt sich Verantwortung trotz verbleibender Ungewissheit wahrnehmen?

I.S.S. (Regie: Gabriela Cowperthwaite; Drehbuch: Nick Shafir; 2023) bietet dafür eine konzentrierte fiktionale Ausgangssituation. Auf der Internationalen Raumstation arbeiten amerikanische und russische Besatzungsmitglieder zusammen. Während eines weltweiten Konflikts erhalten die Gruppen von der Erde Anweisungen, die Kontrolle über die Station zu sichern. Diese Prämisse ist durch die offizielle Filmseite und die Filminformation des Deauville-Festivals belegt. Meine weiterführenden Beispiele beziehen sich auf die damit verbundenen philosophischen Probleme; ich erhebe keinen Anspruch, einzelne Dialoge oder nicht eigens überprüfte Filmszenen nachzuerzählen (Cowperthwaite/Shafir 2023; I.S.S. Official Website 2024).

Ein Film stellt weder einen empirischen Versuch noch ein dokumentiertes Ereignis dar. Seine dramaturgische Konstruktion erlaubt die Zuspitzung moralischer Konflikte. Die hier gewonnenen Einsichten bleiben deshalb begrifflich-argumentativ. Eine Übertragung auf reales Krisenhandeln oder auf die Leistung des Weisheitskompasses müsste eigens untersucht werden.

2. Der Zusammenstoß legitimer Verpflichtungen

Das Gedankenexperiment gewinnt seine Schärfe aus dem Nebeneinander mehrerer zunächst plausibler Verpflichtungen. Die Bindung an die eigene Institution, die Sicherheit der Station, die wissenschaftliche Zusammenarbeit und die Rücksicht auf unmittelbar betroffene Menschen lassen sich jeweils begründen. Im Konfliktfall geraten ihre konkreten Anforderungen jedoch möglicherweise auseinander. Die bloße Berufung auf einen Wert reicht für die Entscheidung folglich nicht aus; ebenso wichtig wird seine Reichweite unter den gegebenen Umständen.

Für eine erste Bestimmung verwende ich eine Menge der im Fall relevanten normativen Gesichtspunkte:

$$V(F) = \{ v_1, v_2, \dots, v_n \}$$

Gleichung E.1 · Menge der relevanten Wertansprüche

V(F) bezeichnet die im Fall F berücksichtigten Wertansprüche; v_1 bis v_n stehen für einzelne Ansprüche, beispielsweise Loyalität, Sicherheit, Vertrauen oder Schutzpflichten. Die geschweiften Klammern bezeichnen eine

Menge. Aus ihrer Darstellung folgt weder ein Zahlenwert noch eine vorgegebene Rangfolge. Welche Ansprüche tatsächlich einschlägig sind, ergibt sich erst aus der sorgfältigen Fallerschließung.

Für den Weisheitskompass hat diese Überlegung eine unmittelbare Konsequenz. Sein Beitrag kann nicht in einer automatischen Harmonisierung konkurrierender Werte bestehen. Ich muss die jeweils behaupteten Gründe offenlegen und prüfen, unter welchen Bedingungen eine Entscheidung bestimmte Pflichten erfüllt und andere zurückstellt. Verbleibende moralische Belastungen dürfen im Ergebnis sichtbar bleiben.

3. Anordnung, Rechtmäßigkeit und Gewissen

Ein institutioneller Befehl beansprucht innerhalb einer bestehenden Autoritätsordnung Verbindlichkeit. Ob dieser Anspruch rechtlich besteht, hängt von Zuständigkeit, Befugnis und dem Inhalt der Anordnung ab. Eine moralische Rechtfertigung verlangt eine zusätzliche Prüfung: Die mögliche Gefährdung anderer Menschen, die Tatsachengrundlage des Befehls und die Folgen seiner Befolgung müssen berücksichtigt werden.

$$B(a) \not\Rightarrow N(a)$$

Gleichung E.2 · Kein automatischer Übergang von Befehl zu moralischer Rechtfertigung

B(a) bezeichnet das Vorliegen einer institutionellen Anweisung zur Handlung a. N(a) bezeichnet deren moralische Rechtfertigung. Das Zeichen ⇏ bedeutet, dass aus der linken Aussage die rechte nicht zwingend folgt. Die Formel entzieht legitimen Befehlen keine grundsätzliche Verbindlichkeit; sie kennzeichnet die Eigenständigkeit des normativen Geltungsanspruchs.

Die persönliche Gewissensprüfung eröffnet dem handelnden Menschen einen Ort verantwortlicher Selbstbefragung. Gewissen verdient Beachtung, bleibt aber anfällig für Irrtümer, Furcht und einseitige Loyalitäten. Seine Prüfung muss deshalb für Gegenargumente offenbleiben. Unter außergewöhnlichen Umständen kann gerade die Bereitschaft, eine als selbstverständlich geltende Anweisung erneut zu begründen, Ausdruck verantwortlicher Urteilskraft sein.

Die Bedeutung dieser Differenz reicht in die KI-Ethik hinein. Die Ausgabe einer technischen Entscheidungsempfehlung begründet für sich genommen keine moralische Pflicht. Gleiches gilt für die Behauptung, eine algorithmische Entscheidung sei wegen ihrer statistischen Genauigkeit schon gerechtfertigt. Empirische Treffgenauigkeit und normative Angemessenheit sind gesondert auszuweisen.

4. Vertrauen, Zuschreibungen und Eskalationsgefahr

Gemeinsames Leben in einer Raumstation ist von gegenseitiger Abhängigkeit geprägt. Unter Friedensbedingungen können eingespielte Arbeitsbeziehungen eine belastbare Erwartung wechselseitiger Zuverlässigkeit begründen. Sobald gegensätzliche Weisungen und unvollständige Informationen hinzutreten, verändert sich die Deutung desselben Verhaltens. Eine vorsichtige Handlung kann nun als Vorbereitung eines Angriffs erscheinen; Zurückhaltung wird möglicherweise als Verheimlichung interpretiert.

Daraus lässt sich ein möglicher Rückkopplungsprozess beschreiben. Sinkendes Vertrauen begünstigt Schutzverhalten; dieses wird vom Gegenüber womöglich als Feindseligkeit gedeutet und kann weiteres Misstrauen auslösen. Ich verstehe diesen Zusammenhang als plausible Deutung eines Konfliktmechanismus, nicht als durch den Film empirisch bestätigtes Gesetz.

$$T_{t+1} = \Psi(T_t, O_t, H_t, K_t)$$

Gleichung E.3 · Heuristik der Vertrauensdynamik

T_t bezeichnet das zum Zeitpunkt t bestehende Vertrauen; T_{t+1} dessen möglichen Folgezustand. O_t steht für beobachtetes Verhalten, H_t für die ihm zugeschriebenen Motive und K_t für die situativen Rahmenbedingungen. Ψ ist lediglich der Name einer denkbaren Einflussbeziehung. Die Formel enthält weder empirisch bestimmte Parameter noch eine prognostische Aussage.

Der epistemische Pass des Weisheitskompasses erhält eine konkrete Aufgabe. Er kann unterscheiden, ob eine Aussage auf direkter Beobachtung, einer plausiblen Interpretation, einer unbelegten Vermutung oder einer überprüften Information beruht. Eine solche Trennung ist auch für den Einsatz dialogfähiger KI-Systeme zentral: Eine sprachlich überzeugende Deutung darf ihren hypothetischen Status nicht verlieren.

5. Identität, Zugehörigkeit und die gemeinsame Lebenswelt

Eine Raumstation verdichtet zwei Identitätsbezüge auf engem Raum. Nationale und institutionelle Zugehörigkeit bestehen fort; gleichzeitig wird die gemeinsam geteilte Verletzlichkeit zum unmittelbaren Erfahrungsbereich. Die handelnden Personen können einander als Vertreter unterschiedlicher politischer Ordnungen betrachten oder als Menschen, deren Sicherheit wechselseitig voneinander abhängt. Beide Perspektiven sind erklärungskräftig. Ihre normative Bedeutung ist situationsbezogen zu prüfen.

Der Gedanke einer gemeinsamen menschlichen Existenz setzt nationale Verpflichtungen nicht außer Kraft. Er erinnert allerdings daran, dass Mitgliedschaft in einer Gruppe nicht jede Rücksicht auf Außenstehende verdrängen darf. Hier entsteht eine Verbindung zu den kulturphilosophischen Überlegungen über KI: Auch technische Systeme werden innerhalb bestimmter gesellschaftlicher und institutioneller Ordnungen gestaltet. Ihre Zwecke bedürfen einer Prüfung im Hinblick auf diejenigen Menschen, die von ihren Folgen betroffen sind.

Ich halte es für methodisch verfehlt, aus der Herkunft eines KI-Systems unmittelbar auf eine einheitliche Weltanschauung seiner Entwickler zu schließen. Nachweisbar wären gegebenenfalls konkrete Zielvorgaben, Trainingsentscheidungen, institutionelle Regeln oder spezifische Verhaltensmuster. Die kulturphilosophische Deutung erschließt mögliche Hintergründe; empirische Forschung muss deren tatsächliche Wirksamkeit untersuchen.

6. Krise als Veränderung der Erkenntnis- und Entscheidungsbedingungen

Die für eine Entscheidung verfügbare Zeit, die Verlässlichkeit der Kommunikation und die Reversibilität der Folgen verändern die Voraussetzungen praktischer Abwägung. In einer akuten Gefahrenlage kann das vollständige Einholen weiterer Informationen unmöglich werden. Damit verschärft sich die Frage, welche Prüfung unabdingbar ist und an welcher Stelle eine begründet vorläufige Entscheidung getroffen werden muss.

Im Weisheitskompass sind diese Bedingungen nicht nur äußere Umstände. Sie beeinflussen, wie Erkenntnisansprüche geprüft und Handlungsmöglichkeiten verantwortbar eingeschätzt werden können. Eine Formel, die Krisenurteilskraft mathematisch erfassen möchte, darf diesen Zusammenhang nur heuristisch abbilden:

$$J_t = \Phi(E(F_t), V_t, G_t, R_t, A_t \mid K_t)$$

Gleichung E.4 · Struktur des unter Krisenbedingungen gebildeten Urteils

J_t bezeichnet das zum Zeitpunkt t gebildete Urteil, $E(F_t)$ den epistemisch erschlossenen Fall, V_t die relevanten Wertansprüche, G_t die Gewissensreflexion, R_t die betroffenen Beziehungen, A_t die verfügbaren Handlungsalternativen und K_t die situativen Bedingungen. Das Zeichen | ist als „unter den Bedingungen von“ zu lesen. Φ steht für eine noch nicht mathematisch spezifizierte Verknüpfung; das Urteil wird damit weder objektiv berechnet noch numerisch optimiert.

Die Formel ist zurückhaltend. Ein angemessenes Ergebnis kann aus der begründeten Bevorzugung einer Handlung bestehen; ebenso möglich sind eine Auswahl mehrerer vertretbarer Optionen oder die ausdrückliche Feststellung eines unaufgelösten Konflikts. Unter bestimmten Bedingungen kann das Unterlassen einer Handlung selbst erklärungs- und verantwortungsbedürftig sein.

7. Epistemische Kontrolle im Weisheitskompass

Die bereits entwickelte Grundarchitektur des Weisheitskompasses stellt eine Beziehung zwischen perspektivischer Fallerschließung, Brückenregeln, Wirkzusammenhängen, Kontext und Revision her. Auch im vorliegenden Gedankenexperiment bleibt diese Architektur maßgeblich:

$$E(F) = \Gamma(\Pi(F), B, M, K, U)$$

Gleichung E.5 · Erkenntnistheoretische Integrationsarchitektur

E(F) ist das strukturiert ausgewiesene Erkenntnisergebnis über den Fall F. Π(F) bezeichnet dessen Erschließung aus unterschiedlichen Perspektiven. B umfasst die Begründungsregeln für Übergänge zwischen Aussagearten, M die relevanten Wirk- und Vermittlungszusammenhänge, K die Kontextbedingungen und U die Unsicherheits- und Revisionsvoraussetzungen. Γ bezeichnet den methodischen Integrationsvorgang und ist gegenwärtig keine empirisch validierte Rechenfunktion.

Im fiktiven Fall von I.S.S. kann die Prüfung nacheinander nach der tatsächlichen Informationslage, den möglichen Absichten anderer, den vorliegenden institutionellen Pflichten, den Folgen möglicher Handlungen und den Grenzen persönlicher Einflussnahme fragen. Ein Erkenntnisgewinn liegt bereits darin, eine behauptete Gewissheit auf ihre faktischen und normativen Voraussetzungen zurückzuführen.

Die produktive Synthese bewahrt die Unterschiede der beteiligten Geltungsformen. Ein empirischer Befund ersetzt kein moralisches Urteil; ein Gewissensanspruch ersetzt keine Tatsachenprüfung. Der Übergang zwischen diesen Ebenen muss offengelegt und begründet werden.

8. Die Verbindung zur künstlichen Intelligenz

Das fiktive Filmbeispiel behandelt vorrangig einen geopolitisch zugespitzten menschlichen Loyalitätskonflikt. Seine Bedeutung für das KI-Thema erschließt sich durch eine ausdrücklich begrenzte Analogie: In beiden Zusammenhängen kann eine als autoritativ wahrgenommene Anweisung oder Empfehlung die eigenständige Prüfung ihrer Gründe verdrängen. Die rechtliche Stellung eines militärischen Befehls und die technische Ausgabe

eines KI-Systems unterscheiden sich grundlegend. Die Analogie betrifft allein das Problem der ungeprüften Autoritätsübernahme.

Generative KI kann unter Zeitdruck wertvolle Informationen bereitstellen, alternative Hypothesen formulieren und Risiken sichtbar machen. Gleichzeitig können fehlerhafte Ausgaben, selektive Gewichtungen oder unbekannte Entwicklungsentscheidungen den Anschein einer Sicherheit erzeugen, die sachlich nicht gerechtfertigt ist. Für verantwortliche Anwendungen sind daher Quellenprüfung, offengelegte Zielmaßstäbe und eindeutig zugewiesene Zuständigkeiten wesentlich.

Im Anschluss an Markus Gabriels Überlegungen zur ethischen Intelligenz interessiert mich besonders, wie technische Systeme moralisch relevante Muster erschließen können und wie sich ihre Empfehlungen angemessen überprüfen lassen. Eine solche Unterstützung kann menschliche Urteilsbildung erweitern. Ihre tatsächliche Güte hängt von der jeweiligen Anwendung, ihren Grenzen und einer nachvollziehbaren Evaluation ab. Aus dem Film lässt sich hierfür kein Wirksamkeitsnachweis gewinnen.

Die Empfehlung eines Systems bildet deshalb lediglich einen prüfungsbedürftigen Bestandteil des Entscheidungsprozesses. Als Formel einer kategorialen Grenze lässt sich festhalten:

$$KI(a) \not\Rightarrow N(a)$$

Gleichung E.6 · Algorithmische Empfehlung und normative Berechtigung

KI(a) bedeutet: Ein KI-System empfiehlt die Handlung a. N(a) steht erneut für deren moralische Rechtfertigung. Das Zeichen $\not\Rightarrow$ verneint eine notwendige Schlussfolgerung. Der Satz bestreitet keinen möglichen Erkenntniswert technischer Empfehlungen; er verhindert ihre vorschnelle Umdeutung in normative Verbindlichkeit.

9. Institutionelle Zurechenbarkeit und Verantwortung

Je bedeutsamer algorithmische Empfehlungen für menschliche Handlungen werden, desto dringlicher ist die Frage, wer ihre Voraussetzungen festlegt, ihre Geltung überprüft und für mögliche Fehlentscheidungen einzustehen hat. Die bloße Anwesenheit eines Menschen am Ende eines automatisierten Prozesses reicht hierfür nicht zwingend aus. Die betreffende Person benötigt einen hinreichenden Einblick in die Entscheidungssituation und eine tatsächliche Möglichkeit, dem technischen Vorschlag begründet zu widersprechen.

Dabei sind Verantwortlichkeiten über mehrere Ebenen verteilt: Entwicklung und Bereitstellung eines Systems, organisatorische Einsatzentscheidung, fachliche Prüfung im Einzelfall sowie die institutionelle Kontrolle von Fehlentwicklungen. Die Unterscheidung dieser Ebenen entspricht dem Anliegen des Weisheitskompasses, begründbare Zusammenhänge sichtbar zu halten und pauschale Verantwortungszuschreibungen zu vermeiden. Die UNESCO-Empfehlung zur KI-Ethik und das NIST AI Risk Management Framework bieten dafür ergänzende, eigenständig entwickelte Bezugspunkte (UNESCO 2021; NIST 2023).

10. Gewissen, religiöse Deutung und die Grenze der Analogie

Eine theologische Lesart des Filmbeispiels kann Fragen nach Gewissen, Nächstenverantwortung und menschlicher Fehlbarkeit vertiefen. Sie bleibt eine ausdrücklich gekennzeichnete Interpretationsperspektive. Die fiktive Handlung ist keine theologische Lehre, und aus ihr ergibt sich keine privilegierte Geltung einer bestimmten religiösen Deutung. Die Verantwortung gegenüber einem konkret gefährdeten Mitmenschen lässt sich auch aus säkularen ethischen Traditionen begründen.

Für meinen Weisheitskompass bleibt wesentlich, dass religiöse, philosophische und empirische Aussagen auf unterschiedliche Weise gerechtfertigt werden. Die Möglichkeit produktiver Beziehungen zwischen ihnen wächst durch begriffliche Klarheit. Eine harmonisierende Gesamtdeutung würde gerade jene Unterschiede verdecken, deren Prüfung für die verantwortliche Urteilsbildung entscheidend ist.

11. Revisionsfähigkeit und verbleibende moralische Belastung

Unter extremen Bedingungen können Entscheidungen irreversible Folgen haben. Eine spätere Revision des Urteils macht die vergangene Handlung nicht rückgängig. Gleichwohl bleibt die Bereitschaft zur nachträglichen Prüfung wesentlich, weil neue Erkenntnisse frühere Vermutungen korrigieren und institutionelle Lernprozesse ermöglichen können. Die verantwortliche Person muss mit der Möglichkeit leben, dass eine unter Zeitdruck vertretbar begründete Entscheidung dennoch schwerwiegende Nachteile verursacht.

Ich halte es deshalb für notwendig, zwischen Begründungsqualität, Handlungsfolge und moralischer Restbelastung zu unterscheiden. Eine sorgfältige Entscheidung beseitigt die Tragik konkurrierender Pflichten nicht zwingend. Hier gewinnt Urteilskraft eine menschlich anspruchsvolle Gestalt: Sie setzt die Anerkennung begrenzten Wissens, die Achtung vor betroffenen Menschen und die Bereitschaft voraus, Verantwortung für Entscheidungen und deren Folgen zu übernehmen.

12. Schlussfolgerung

Die fiktive Extremsituation von I.S.S. verdeutlicht die Dringlichkeit einer Unterscheidung, auf die ich im Weisheitskompass besonderes Gewicht lege: Werte, Erkenntnisse und Handlungsanweisungen beanspruchen jeweils unterschiedliche Arten der Begründung. Das Zusammentreffen kann verantwortliche Entscheidungen erschweren, selbst wenn alle Beteiligten nachvollziehbare Verpflichtungen anführen.

Die Aufgabe der Urteilskraft besteht dann darin, die verfügbare Erkenntnis zu prüfen, die betroffenen Menschen zu berücksichtigen, konkurrierende normative Ansprüche zu würdigen und die Grenzen möglicher Rechtfertigungen offenzulegen.

Diese Einsicht eröffnet eine sachlich begrenzte, aber fruchtbare Verbindung zur künstlichen Intelligenz. Algorithmische Systeme können den Raum der verfügbaren Informationen und Hypothesen erweitern. Die öffentliche und persönliche Verantwortung für normativ bedeutsame Entscheidungen bedarf weiterhin klarer Verfahren der Prüfung und Zurechnung. Welche zusätzlichen Leistungen der Weisheitskompass gegenüber etablierten Verfahren erbringt, bleibt eine Aufgabe zukünftiger Fallstudien und empirischer Vergleiche.

„Verantwortliche Urteilskraft zeigt sich, wo ein Wertkonflikt bestehen bleibt, Wissen unvollständig ist und die Gründe des Handelns dennoch öffentlich und persönlich verantwortet werden müssen.“

Der Weisheitskompass kann in diesem Sinn als vorläufig konsolidierte, revisionsoffene Architektur der Begründungsprüfung dienen. Das filmische Gedankenexperiment fordert seine Begriffe heraus; es liefert noch keinen Beweis empirischer Wirksamkeit.

Quellen- und urheberrechtlicher Hinweis

Dieser Anhang ist eine eigenständige wissenschaftliche Interpretation. Die Darstellung der Filmprämisse beruht auf öffentlich zugänglichen Inhaltsangaben; eine vollständige szenengenaue Filmauswertung wird nicht beansprucht. Es wurden keine Dialoge, Drehbuchpassagen, Filmstills, Trailerbilder, Filmplakate oder fremde grafische Darstellungen übernommen. Die mathematischen Notationen wurden für die vorliegende Untersuchung eigenständig formuliert. Angaben zur Handlung und zu den Urhebern des Films sind als solche erkennbar und nachfolgend belegt. Die offiziellen Informationen von Universal Pictures Germany weisen darauf hin, dass die auf der Website bereitgestellten Trailer, Bilder und sonstigen Materialien nicht auf einer fremden Homepage eingebunden werden dürfen. Für diesen Anhang werden deshalb ausschließlich gewöhnliche Quellenlinks verwendet. Die österreichische Zitierregelung (§ 42f UrhG) gestattet Nutzungen nur im jeweils gerechtfertigten Umfang; § 57 UrhG enthält Anforderungen an die Quellenangabe.

Der vorliegende Essay ist eine eigenständige wissenschaftlich-philosophische Auseinandersetzung mit dem Science-Fiction-Film *I.S.S.* und seiner Bedeutung für den Weisheitskompass. Die Darstellung der Filmhandlung beschränkt sich auf eine selbst formulierte, quellenbezogene Zusammenfassung. Die weiterführenden Überlegungen zu Verantwortung, Loyalität, Gewissen und menschlicher Urteilskraft beruhen auf einer eigenständigen philosophischen Interpretation. Auf die Übernahme von Filmplakaten, Standbildern, Filmdialogen und sonstigem geschützten Filmmaterial wurde verzichtet. Die herangezogenen Quellen sind im Literaturverzeichnis ausgewiesen.

Bei der sprachlichen und redaktionellen Ausarbeitung wurde generative künstliche Intelligenz unterstützend eingesetzt.

Die philosophische Konzeption des Weisheitskompasses, die wissenschaftliche Einordnung sowie die inhaltliche Ausrichtung der Untersuchung liegen bei mir als Autor. Ebenso trage ich die Verantwortung für die sachliche Überprüfung der Quellen, die vertretenen Schlussfolgerungen und die endgültige Freigabe der veröffentlichten Fassung. Die Ausführungen dienen ausschließlich der allgemeinen wissenschaftlichen, philosophischen und didaktischen Auseinandersetzung und stellen keine individuelle Beratung oder Handlungsempfehlung für konkrete Einzelfälle dar.

Literatur und Quellen

Cowperthwaite, Gabriela (Regie); Shafir, Nick (Drehbuch) (2023): I.S.S. Spielfilm, USA. Offizielle Filmwebsite: https://www.iss.movie/ (Abruf: 16.10.2026).

Universal Pictures International Germany: I.S.S. Filminformation und FAQ zur Verwendung von Bild- und Videomaterial. https://www.upig.de/micro/i-s-s (Abruf: 16.10.2026).

Festival du Cinéma Américain de Deauville: I.S.S. Filmeintrag, Produktionsjahr 2023, Regie und Drehbuch. https://www.festival-deauville.com/en/movies/i-s-s/ (Abruf: 16.10.2026).

Gabriel, Markus (2026): Ethische Intelligenz. Wie KI uns moralisch weiterbringen kann. Berlin: Ullstein. Philosophischer Bezugspunkt; keine vollständige Buchauswertung dieses Anhangs.

UNESCO (2021): Recommendation on the Ethics of Artificial Intelligence. Paris. https://www.unesco.org/en/legal-affairs/recommendation-ethics-artificial-intelligence.

National Institute of Standards and Technology (NIST) (2023): Artificial Intelligence Risk Management Framework (AI RMF 1.0). NIST AI 100-1. https://doi.org/10.6028/NIST.AI.100-1.

Republik Österreich: Urheberrechtsgesetz, §§ 42f und 57, konsolidierte Fassung. https://www.ris.bka.gv.at/.

Der Weisheitskompass

Urteilskraft, Verantwortung und Gewissen unter Unsicherheit

Ein heuristischer Orientierungsrahmen für reflektierte Entscheidungen in komplexen, mehrdeutigen und risikobehafteten Situationen.

Fiktive Extremsituation

Eine internationale Besatzung in einem isolierten, technisch abhängigen Lebensraum gerät durch einen externen Konflikt unter Zeitdruck. Informationen sind unvollständig, institutionelle Anweisungen widersprechen persönlichen Bindungen, Vertrauen wird brüchig, und jede Entscheidung kann existenzielle Folgen haben.

$$S = (O, A, C, U, T, St)$$

- O = Ziele (Objectives)
- A = Handlungsalternativen (Alternatives)
- C = Kontextbedingungen (Context)
- U = Unsicherheit (Uncertainty)
- T = Zeit / Zeithorizont (Time)
- St = Stakeholder / betroffene Anspruchsgruppen



Notation und Bedeutung

$W(S)$	=	Weisheitskompass als Orientierungsrahmen für die Situation S
F	=	Fall bzw. Entscheidungssituation
$\Pi(F)$	=	perspektivische Erschließung des Falles
B	=	Brückenregeln (ethisch, rechtlich, fachlich, institutionell)
M	=	Wirk- und Vermittlungs-zusammenhänge
K	=	Kontextbedingungen
U	=	Unsicherheits- und Revisionsbedingungen
$E(F)$	=	integriertes Erkenntnisergebnis
a	=	konkrete Handlungsoption
J_t	=	Urteil zum Zeitpunkt t
V_t	=	relevante Wertansprüche zum Zeitpunkt t
G_t	=	Gewissensreflexion zum Zeitpunkt t
R_t	=	Beziehungen und Vertrauensbedingungen zum Zeitpunkt t
A_t	=	verfügbare Handlungsalternativen zum Zeitpunkt t
K_t	=	situative Bedingungen zum Zeitpunkt t
Γ	=	heuristische Integrationsfunktion
Φ	=	heuristische Verknüpfungsfunktion
t	=	betrachteter Entscheidungszeitpunkt

Heuristisches Modell

Zweistufiger Orientierungs- und Urteilsprozess

I Integrierte Erkenntnisbildung

Verknüpfung von Perspektiven, Regeln, Kontexten und Unsicherheit.

$$E(F) = \Gamma(\Pi(F), B, M, K, U)$$

II Urteil und Entscheidung

Abwägung von Evidenz, Werten, Gewissen und Handlungsalternativen.

$$J_t = \Phi(E(F_t), V_t, G_t, R_t, A_t \mid K_t)$$

$$Anordnung(a) \neq moralische\ Rechtfertigung(a)$$

Eine institutionelle Anordnung impliziert nicht automatisch eine moralische Rechtfertigung.

Interpretation

Zentrale Einsichten für verantwortete Urteilskraft

1. Werte allein erzeugen noch kein verantwortliches Urteil.
2. Urteilskraft verbindet Evidenz, Gewissen und Folgenabschätzung.
3. Kontextbedingungen verändern die Reichweite von Entscheidungen.
4. Institutionelle Anforderungen entbinden nicht von persönlicher Verantwortung.
5. Der Weisheitskompass ordnet Perspektiven, ersetzt aber nicht menschliche Verantwortung.

Krisendynamik *Misstrauen → defensives Verhalten → Bedrohungswahrnehmung → Eskalation*

Beobachtungen und Deutungen sind sorgfältig zu unterscheiden.