SEMINARARBEIT

Der Weisheitskompass

Erfahrung, Wirkung und Erkenntnis im Horizont verantwortlicher Urteilsbildung
Eine integrierte philosophisch-methodologische Untersuchung
mit heuristischem Formalmodell und
möglichem empirischen Forschungsentwurf
Konzeptionelle Entwurfsgrundlagen:

Norbert Rieser

Wissenschaftliche Arbeitsfassung | Oktober 2026
Die Darstellung unterscheidet systematisch zwischen
begrifflicher Begründung, heuristischer Modellbildung
und empirisch möglichem Wirksamkeitsnachweis.

Inhaltsverzeichnis

Inhaltsverzeichnis 2
Zusammenfassung 6
Teil I · Philosophische Grundlegung und Entwicklungsstand 6
Abstract 6
1. Problemstellung und Forschungsfrage 6
2. Quellenbasis, Methode und Geltungsanspruch 7
3. Vom persönlichen Glaubensverständnis zum offenen Orientierungsrahmen 7
4. Weisheit unter begrenztem Wissen: begriffliche Präzisierung 7
5. Die fünf methodischen Schritte 8
6. Der zentrale Erkenntnisgewinn: begründungspflichtige Übergänge 8
7. Dialog, Vertrauen und Resonanz: Bedingungen des Erkenntnisgewinns 8
8. Macht, Wissensasymmetrie und gesellschaftliche Mündigkeit 9
9. Fallanalyse: Einführung eines KI-gestützten Entscheidungssystems 9
10. Einwände und Grenzen 9
10.1 Gefahr der begrifflichen Überdehnung 9
10.2 Gefahr eines unverbindlichen Eklektizismus 10
10.3 Zeitaufwand und asymmetrische Macht 10
10.4 Ungeklärte empirische Überlegenheit 10
11. Ein überprüfbares Forschungsprogramm 10
12. Schluss: Die Entwicklungsstufe des Weisheitskompasses 10
Literatur und Quellen 11
Teil II · Erfahrung, Wirkung und die Grenzen kausaler Schlüsse 13
1. Erkenntnis aus gelebter Erfahrung 13
Die methodologische Unterscheidung zwischen beobachteter Veränderung, kausaler Zuschreibung und verantwortbarer Entscheidung bleibt unabhängig von personenbezogenen Vergleichsansätzen zentral. . 13
3. Das kontrafaktische Problem 13
4. Ein methodischer Dialog als Gedankenexperiment 13
Teil III · Heuristisches mathematisches Modell H5-WK 14
1. Dimensionen und Aggregation 14
2. Urteil, Handlung und Ergebnis 14
3. Lernen und Korrektur 14
4. Wissens- und Machtasymmetrie 14
5. Grenzen der Formalisierung 14

Teil IV · Integrierter Anhang A: Nutzennachweis 15
1. Was mit „Nutzen“ gemeint ist 15
2. Wirkmodell: Wie der Kompass Nutzen erzeugen könnte 15
2.1 Die fünf Schritte als überprüfbare Intervention 15
3. Nutzen auf drei Ebenen 16
Individuum 16
Gruppe und Organisation 16
Gesellschaft und Institution 16
4. Konkrete Indikatoren und Bewertungsraster 16
5. Studiendesign für einen belastbaren ersten Wirksamkeitstest 16
5.1 Forschungsfragen und Hypothesen 16
5.2 Stichprobe und Gruppen 17
5.3 Ablauf und Auswertung 17
6. Fallbeispiel: Einführung eines KI-Systems in einer öffentlichen Einrichtung 18
7. Grenzen und mögliche Gegenbefunde 18
8. Vom Pilotversuch zum gesellschaftlichen Nutzennachweis 18
9. Ergebnis: Was gegenwärtig behauptet werden darf 19
Hinweise zur Quellen- und Forschungsbasis 19
Teil V · Integrierter Anhang B: Operationalisierung und Pilotstudie 20
1. Ziel und wissenschaftlicher Status 20
2. Fünf Dimensionen und zwölf Beobachtungsindikatoren 20
E – Epistemische Prüffähigkeit 20
B – Begründungsqualität 20
D – Dialogfähigkeit 20
V – Verantwortbarkeit 20
K – Korrigierbarkeit 20
3. Bewertungsrubrik je Indikator 20
4. Formale Auswertung 21
5. Pilotstudie: drei Vergleichsbedingungen 21
6. Standardisierte Fallaufgabe 21
7. Prüfung des Nutzens 21
8. Bezug zum wirkungsorientierten Denken 22
9. Falsifizierbarkeit und Weiterentwicklung 22
Teil VI · Schlussbetrachtung 23
Quellen- und Statusvermerk 23

Anhang A · Konzeptionelle Nutzenerwartungen und empirischer Nachweis 24
1. Was mit „Nutzen“ gemeint ist 24
2. Wirkmodell: Wie der Kompass Nutzen erzeugen könnte 24
2.1 Die fünf Schritte als überprüfbare Intervention 24
3. Nutzen auf drei Ebenen 25
Individuum 25
Gruppe und Organisation 25
Gesellschaft und Institution 25
4. Konkrete Indikatoren und Bewertungsraster 25
5. Studiendesign für einen belastbaren ersten Wirksamkeitstest 26
5.1 Forschungsfragen und Hypothesen 26
5.2 Stichprobe und Gruppen 26
5.3 Ablauf und Auswertung 26
6. Fallbeispiel: Einführung eines KI-Systems in einer öffentlichen Einrichtung 27
7. Grenzen und mögliche Gegenbefunde 27
8. Vom Pilotversuch zum gesellschaftlichen Nutzennachweis 28
9. Ergebnis: Was gegenwärtig behauptet werden darf 28
Hinweise zur Quellen- und Forschungsbasis 28
Anhang B · Ursprüngliche Operationalisierung mit zwölf Indikatoren 29
1. Ziel und wissenschaftlicher Status 29
2. Fünf Dimensionen und zwölf Beobachtungsindikatoren 29
E – Epistemische Prüffähigkeit 29
B – Begründungsqualität 29
D – Dialogfähigkeit 29
V – Verantwortbarkeit 29
K – Korrigierbarkeit 29
3. Bewertungsrubrik je Indikator 30
4. Formale Auswertung 30
5. Pilotstudie: drei Vergleichsbedingungen 30
6. Standardisierte Fallaufgabe 30
7. Prüfung des Nutzens 31
8. Bezug zum wirkungsorientierten Denken 31
9. Falsifizierbarkeit und Weiterentwicklung 31
Anhang C · H5-WK-18: Bewertungsmanual der Neun-Indikatoren-Skala 32
C.1 Formalisierung 33

C.2 Standardisierte Fallaufgabe 33
C.3 Bewertung und Gütekriterien 33
Anhang D · Prospektive Fallzahl- und Auswertungsplanung 33
D.1 Sensitivitätsanalyse 34
D.2 Hypothetisches Rechenbeispiel – ausdrücklich keine Studienergebnisse 34
D.3 Pilotstudie und Präregistrierung 34
Schlussbemerkung · Erkenntnisanspruch und Grenzen 34

Zusammenfassung

Die methodologische Unterscheidung zwischen beobachteter Veränderung, kausaler Zuschreibung und verantwortbarer Entscheidung bleibt unabhängig von personenbezogenen Vergleichsansätzen zentral.

Teil I · Philosophische Grundlegung und Entwicklungsstand

Abstract

Die Arbeit rekonstruiert den Weisheitskompass als praxisbezogenen Orientierungsrahmen für Urteile unter Unsicherheit und ungleichen Wissens- und Machtverhältnissen. Ausgangspunkt ist die Unterscheidung zwischen einer persönlichen philosophisch-theologischen Selbstverortung und einem öffentlich anschlussfähigen Prüfverfahren. Die zentrale These: Der mögliche Mehrwert des Weisheitskompasses liegt nicht in neuen Einzelprinzipien, sondern in der expliziten Verknüpfung epistemischer Trennschärfe, begründungspflichtiger Übergänge, dialogischer Korrekturfähigkeit, institutioneller Verantwortlichkeit und revisionsoffener Entscheidungen. Diese Integrationshypothese wird anhand eines hypothetischen KI-Anwendungsfalls konkretisiert und in überprüfbare Qualitätskriterien übersetzt. Die Arbeit benennt Einwände und entwickelt ein vorläufiges Evaluationsdesign; eine nachgewiesene Wirksamkeit wird derzeit noch nicht behauptet.

Schlüsselbegriffe: Weisheit; begrenztes Wissen; Urteilsbildung; Wissensasymmetrie; Dialog; Resonanz; Macht; Korrigierbarkeit.

1. Problemstellung und Forschungsfrage

Moderne Gesellschaften verfügen über hochspezialisiertes Wissen. Die für Entscheidungen verantwortlichen Personen und die von Entscheidungen Betroffenen können dieses Wissen jedoch häufig weder im gleichen Umfang erwerben noch unabhängig reproduzieren. Aus Spezialisierung entsteht eine unvermeidliche Wissensasymmetrie. Problematisch wird sie dort, wo fachlicher Vorsprung mit institutioneller Macht zusammenfällt und die Bedingungen einer Entscheidung unverständlich oder unzugänglich bleiben. Gleichzeitig müssen Menschen handeln, bevor alle Tatsachen geklärt und alle Folgen sicher absehbar sind.

Die Forschungsfrage lautet daher: Wie kann ein Orientierungsrahmen die Qualität von Urteilen und Entscheidungen unter begrenztem Wissen verbessern, ohne Expertenwissen zu ersetzen oder die eigene Wirksamkeit vorauszusetzen? Drei Teilfragen strukturieren die Untersuchung: (1) Welche Prüfungen sind für begründete Urteile notwendig? (2) Unter welchen Bedingungen wird Dialog erkenntnisfördernd? (3) Wie lassen sich Verantwortlichkeit und spätere Korrektur institutionell absichern?

Die Arbeit vertritt die These, dass Weisheit im hier verwendeten Sinn weder Wissensmenge noch subjektives Wohlbefinden umfasst, sondern eine begründete, maßvolle und korrigierbare Praxis des Urteilens. Entscheidend ist ob sie zum Entscheidungszeitpunkt unter angemessener Berücksichtigung der verfügbaren Gründe und Risiken verantwortbar war; - nicht ob jede Entscheidung im Nachhinein erfolgreich ausfällt.

2. Quellenbasis, Methode und Geltungsanspruch

Die Darstellung stützt sich primär auf zwei vom Autor bereitgestellte Texte: die letzte, als „Überarbeitete Fassung – März 2026“ bezeichnete Version der „Grundlinien eines zeitgemäß reflektierten Glaubens“ (S. 67–68 der 68-seitigen Sammlung) sowie „Der Weisheitskompass: Grundlagen, gesellschaftliche Verantwortung und überprüfbarer praktischer Nutzen“ (Oktober 2026). Frühere Varianten der 68-seitigen Sammlung werden nicht zu einer fiktiven Einheit verschmolzen. Die März-Fassung beschreibt eine philosophisch-theologische Selbstverortung; die Oktober-Fassung formuliert den gesellschaftlichen und methodischen Anspruch.

Methodisch handelt es sich um eine begriffsanalytische und argumentationsrekonstruktive Arbeit. Die Quellenpositionen werden zunächst dargestellt, anschließend auf interne Übergänge, mögliche Einwände und beobachtbare Konsequenzen hin geprüft. Ergänzend werden etablierte Forschungslinien zur reflektierten Urteilsbildung, Weisheit, Hermeneutik und Entscheidung unter Unsicherheit als Vergleichshorizont herangezogen. Dieser Vergleich begründet Anschlussfähigkeit, aber keine empirische Bestätigung des Gesamtmodells.

3. Vom persönlichen Glaubensverständnis zum offenen Orientierungsrahmen

Die letzte Version der „Grundlinien“ beginnt mit drei philosophischen Fragen: Warum ist Wirklichkeit rational verstehbar? Warum besitzen normative Aussagen Geltung? Warum erfährt sich der Mensch als verantwortliches Subjekt? Sie unterscheidet naturwissenschaftliche Beschreibung von weltanschaulicher Deutung und versteht Freiheit als Antwortfähigkeit: als Fähigkeit, Gründe zu prüfen, Stellung zu nehmen und Ansprüche anzuerkennen. Der Begriff eines „tragenden Grundes“ dient in diesem Text als philosophisch-theologische Deutung, die in einer christologischen Perspektive weitergeführt wird (Rieser, März 2026, S. 67–68).

Für die Weiterentwicklung ist eine präzise Differenz wesentlich: Der reflektierte Glaube bezeichnet eine mögliche persönliche Antwort auf Fragen nach Sinn und Wirklichkeitsgrund. Der Weisheitskompass hingegen soll als Verfahren auch für Menschen verwendbar sein, die diese Antwort nicht teilen. Seine öffentliche Geltung kann deshalb nicht von einer bestimmten Gottesvorstellung abhängen. Die Trennung schützt die Prüfbarkeit des Modells vor weltanschaulicher Vorentscheidung und stellt keine Abwertung von Glauben dar.

Die im Oktober 2026 vorgelegte Fassung setzt einen neuen Akzent: „Prüffähigkeit trotz Wissensungleichstand“ wird zum Leitgedanken. Fachwissen soll nicht nivelliert, sondern hinsichtlich Herkunft, Unsicherheit, Interessen, Folgen und Legitimation befragbar werden (Rieser, Oktober 2026, S. 1–4). Damit erweitert sich die Perspektive von individueller Lebensführung auf gesellschaftliche Entscheidungsprozesse.

4. Weisheit unter begrenztem Wissen: begriffliche Präzisierung

Ein epistemisch realistischer Weisheitsbegriff beginnt mit der Anerkennung, dass Urteile fehlbar sind. Fehlbarkeit rechtfertigt weder Beliebigkeit noch Handlungsverzicht. Vielmehr verlangt sie die angemessene Kalibrierung von Überzeugungen: Welche Gründe sprechen für eine Annahme, welche dagegen, und wie groß ist die verbleibende Unsicherheit? In diesem Sinn lässt sich Weisheit als Fähigkeit verstehen, begrenzte Erkenntnis mit Perspektivenprüfung, Wertklärung und verantwortlichem Handeln zu verbinden.

Ein solcher Begriff berührt Aristoteles' Unterscheidung von theoretischem Wissen und praktischer Klugheit (phronēsis), ist mit dieser aber nicht identisch: Der Weisheitskompass ergänzt die individuelle Urteilspraxis ausdrücklich um Macht- und Institutionenprüfung. Forschung zum „reflective judgment" beschreibt die Entwicklung begründeter Urteile in schlecht strukturierten Problemsituationen; psychologische Weisheitsforschung untersucht unter anderem Perspektivenwechsel, intellektuelle Demut und den Umgang mit Unsicherheit (King & Kitchener, 2004; Grossmann, 2017). Diese Bezüge sind konzeptionell einschlägig, ersetzen aber keinen Wirksamkeitstest des konkreten Kompasses.

Die normative Kurzdefinition lautet: Weisheit ist die Fähigkeit, unter begrenztem Wissen relevante Perspektiven und Werte kritisch abzuwägen, die eigene Entscheidung angemessen zu begründen, ihre Folgen verantwortlich zu tragen und auf neue Gründe hin zu revidieren.

5. Die fünf methodischen Schritte

Die Abfolge ist heuristisch und iterativ: Neue Befunde können eine erneute Unterscheidung verlangen, Einwände eine Begründung verändern und unerwartete Folgen eine Entscheidung zurücknehmen. Die Schritte erzeugen keine algorithmische „richtige Antwort". Sie strukturieren Rechenschaftspflichten und reduzieren die Gefahr unbemerkter Kurzschlüsse.

6. Der zentrale Erkenntnisgewinn: begründungspflichtige Übergänge

Der methodisch anspruchsvollste Teil liegt zwischen den einzelnen Aussageebenen. Aus Daten folgt nicht automatisch eine Deutung; aus einer Deutung folgt nicht unmittelbar eine moralische Wertung; aus einer Wertung ergibt sich selten genau eine Handlung. Ebenso begründet technische Expertise für sich genommen keine demokratische Entscheidungskompetenz. Der Weisheitskompass verlangt, jeden Übergang durch zusätzliche Gründe auszuweisen (Rieser, Oktober 2026, S. 4).

Hier zeigt sich der Unterschied zwischen bloßer „Perspektivenvielfalt" und methodischer Urteilskritik. Perspektiven dürfen nicht um der Harmonie willen vermischt werden. Ein empirisch widerlegter Sachverhalt gewinnt durch seinen existenziellen Sinn keinen Tatsachenstatus; umgekehrt entscheidet eine Messung nicht allein über die Legitimität eines Eingriffs. Das Verfahren verlangt ausdrücklich, Unvereinbares als unvereinbar zu benennen.

7. Dialog, Vertrauen und Resonanz: Bedingungen des Erkenntnisgewinns

Dialog beinhaltet für das Modell eine doppelte Funktion. Er erschließt Informationen, die einer einzelnen Person fehlen, und er konfrontiert Deutungen mit ernsthaften Gegenargumenten. Vertrauen erleichtert die Bereitschaft, Unsicherheit einzugestehen; es darf jedoch nicht in Kritikverzicht umschlagen. Resonanz bezeichnet hier nicht Einverständnis, sondern die Erfahrung, dass eine fremde Perspektive die eigene Frage- und Urteilslage produktiv verändert.

In der philosophischen Hermeneutik wird Verstehen als geschichtlich situiert und für neue Horizonte offen beschrieben (Gadamer, 1960). Diskurstheoretische Ansätze unterstreichen die Bedeutung von Gründen und öffentlicher Rechtfertigung (Habermas, 1981). Rosas Resonanzbegriff kann den relationalen Aspekt des Gesprächs beleuchten, ist aber kein Wahrheitskriterium (Rosa, 2016). Diese drei Ansätze dürfen deshalb nicht gleichgesetzt werden: Verstehen, Rechtfertigen und Sich-angesprochen-Fühlen sind unterscheidbare Vorgänge.

Erkenntnisfördernd wird Dialog nur unter Bedingungen, die über Gesprächsfreundlichkeit hinausgehen: (a) begründeter Widerspruch ist möglich; (b) Beteiligte dürfen ihre Position ohne Sanktion ändern; (c) Wissens- und Machtgefälle werden benannt; (d) Tatsachenbehauptungen bleiben überprüfbar; (e) Dissens darf bestehen bleiben. Ein Gespräch, das lediglich gegenseitige Bestätigung erzeugt, kann emotional wertvoll sein, ohne die Urteilsqualität zu verbessern.

8. Macht, Wissensasymmetrie und gesellschaftliche Mündigkeit

Die individuelle Fähigkeit, Argumente zu prüfen, reicht nicht aus, wenn Daten, Infrastruktur, Fachsprache und Entscheidungswege institutionell abgeschirmt sind. Der gesellschaftliche Anspruch des Weisheitskompasses besteht daher darin, Transparenz, Zuständigkeit, Beteiligung und Widerspruchsmöglichkeiten mitzudenken. Expertise ist unverzichtbar; sie legitimiert jedoch nicht automatisch politische oder ethische Vorentscheidungen. Wer Wissen kontrolliert, muss die Reichweite seiner Aussagen und die Bedingungen ihrer Nutzung offenlegen, soweit rechtlich und praktisch möglich.

Daraus folgt eine institutionelle Ergänzung: Eine Entscheidung kann individuell sorgfältig begründet sein und dennoch an fehlender demokratischer Legitimation oder unzureichenden Beschwerdewegen scheitern. Der Kompass muss deshalb zwischen Erkenntnisqualität, normativer Rechtfertigung und Verfahrensgerechtigkeit unterscheiden. Das Oktober-Dokument verweist hierzu auf europäische Diskurse über KI-Governance, Menschenrechte und Rechenschaft (Rieser, Oktober 2026, S. 5–6); diese thematische Anschlussfähigkeit ist keine institutionelle Anerkennung des Modells.

9. Fallanalyse: Einführung eines KI-gestützten Entscheidungssystems

Ein hypothetischer öffentlicher Träger erwägt, Anträge mithilfe eines KI-Systems vorzusortieren. Der Anbieter verspricht kürzere Bearbeitungszeiten; zugleich sind die verwendeten Trainingsdaten nur teilweise zugänglich. Das Beispiel dient der Illustration, nicht als empirischer Fallbericht.

Unterscheiden: Zunächst werden nachgewiesene Leistungswerte, Anbieterprognosen und normative Erwartungen getrennt. Begründen: Die Organisation verlangt Angaben zu Fehlerquoten, Testpopulationen, möglichen Verzerrungen und Grenzen der Übertragbarkeit. Verbinden: Fachleute, Beschäftigte, betroffene Bürgerinnen und Bürger sowie Datenschutz- und Rechtsverantwortliche prüfen unterschiedliche Auswirkungen. Handeln: Eine befristete, kontrollierte Erprobung wäre nur vertretbar, wenn Zuständigkeiten, menschliche Aufsicht, Einspruchsrechte und Abbruchkriterien feststehen. Überprüfen: Fehlerverteilung, Beschwerden, Bearbeitungsdauer und unerwartete Benachteiligungen werden anhand vorab definierter Indikatoren ausgewertet.

Der mögliche Gewinn besteht nicht darin, dass der Kompass eine technische Auditmethodik ersetzt. Er macht vielmehr sichtbar, warum technische Leistungsfähigkeit allein weder die Wertentscheidung noch die Legitimation des Einsatzes trägt. Besonders wichtig ist die Frage, wer die Folgen einer fehlerhaften Empfehlung zu verantworten hat und wie Betroffene eine Korrektur erreichen.

10. Einwände und Grenzen

10.1 Gefahr der begrifflichen Überdehnung

Der Begriff „Weisheit“ könnte eine methodische Checkliste mit überhöhtem Anspruch versehen. Dem lässt sich nur begegnen, wenn klar zwischen normativer Zielvorstellung, konkreten Prüfschritten und

beobachtbaren Ergebnissen unterschieden wird. Eine vollständig ausgefüllte Checkliste beweist noch keine Weisheit.

10.2 Gefahr eines unverbindlichen Eklektizismus

Das Modell integriert Ansätze aus Erkenntnistheorie, Ethik, Systemdenken und Dialogtheorie. Ihre Verbindung ist nur dann überzeugend, wenn die jeweiligen Geltungsbereiche erkennbar bleiben. Nicht jede Theorie bleibt kompatibel; nicht jede Perspektive hat dieselbe Evidenz. „Verbinden“ muss daher ebenso das begründete Nicht-Verbinden erlauben.

10.3 Zeitaufwand und asymmetrische Macht

In dringlichen Situationen kann die vollständige Durchführung aller Schritte unpraktikabel sein. Unter starkem Machtgefälle reicht Gesprächsoffenheit möglicherweise nicht aus. Deshalb sind abgestufte Verfahren, externe Prüfung, dokumentierte Zuständigkeiten und wirksame Einspruchswege nötig. Der Kompass ist kein Ersatz für rechtliche Schutzmechanismen.

10.4 Ungeklärte empirische Überlegenheit

Dass einzelne Bestandteile durch Forschung gestützt werden, beweist nicht, dass ihre Kombination bessere Entscheidungen hervorbringt als etablierte Alternativen. Ein solcher Schluss wäre ein Kompositionsfehler. Die Gesamtwirksamkeit muss vergleichend untersucht werden.

11. Ein überprüfbares Forschungsprogramm

Ein Pilotprojekt könnte Teilnehmenden vergleichbare, mehrdeutige Entscheidungsszenarien vorlegen. Eine Gruppe arbeitet mit dem Weisheitskompass, eine andere mit einer etablierten Reflexionsmethode. Vorab definierte Kriterien bewerten anonymisierte Antworten möglichst verblindet. Ein bloßer Zufriedenheitsfragebogen wäre dafür unzureichend.

Die Auswertung soll neben möglichen Verbesserungen auch Nullbefunde, Übungseffekte und Zeitaufwand erfassen. Entscheidend wird ein Vergleich mit einer aktiven Alternative, nicht nur mit „gar keiner Methode“. Auch eine höhere Punktzahl für schriftliche Begründungen belegt noch keine besseren realweltlichen Folgen. Langfristige Untersuchungen sollten prüfen, ob Entscheidungen tatsächlich transparenter, fairer und korrigierbarer werden.

12. Schluss: Die Entwicklungsstufe des Weisheitskompasses

Der aktuelle Stand lässt sich in drei Sätzen zusammenfassen. Erstens: Der Weisheitskompass entstand aus einer persönlichen Auseinandersetzung mit Lebenserfahrung, Wissen, Glauben und verantwortlicher Freiheit. Zweitens: Seine weiterentwickelte Fassung ist ein weltanschaulich offener Orientierungsrahmen, der individuelle Urteilskraft mit der kritischen Prüfung von Wissens- und Machtverhältnissen verbindet. Drittens: Sein spezifischer Anspruch liegt in der geordneten Verknüpfung von Prüffähigkeit, dialogischer Korrekturfähigkeit und verantwortlicher Entscheidung; ob daraus ein messbarer Mehrwert entsteht, ist Gegenstand künftiger Prüfung.

Die Leitthese: Weisheit heißt, unter begrenztem Wissen prüffähig zu bleiben, begründeten Widerspruch im Dialog ernst zu nehmen und so zu entscheiden, dass Gründe, Folgen und Korrekturmöglichkeiten nachvollziehbar bleiben.

Meine weiteren ergänzten Weisheitskompass-Orientierungsmodelle sind in diesem Verständnis keine konkurrierenden Grundsysteme, sondern sachbezogene Ausdifferenzierungen und anwendungsorientierte Fortführungen des offenen Rahmens. Sie müssen sich an denselben Anforderungen messen lassen: Klarheit der Begriffe, Tragfähigkeit der Gründe, Offenheit für Gegenargumente und Verantwortlichkeit gegenüber den Betroffenen.

Literatur und Quellen

Rieser, Norbert (März 2026): Grundlinien eines zeitgemäß reflektierten Glaubens. Überarbeitete Fassung. In: ARBEITSVERSIONEN – Stand 01–2026, S. 67–68. Auf Homepage veröffentlichtes Arbeitsmanuskript.

Rieser, Norbert (Oktober 2026): Der Weisheitskompass. Grundlagen, gesellschaftliche Verantwortung und überprüfbarer praktischer Nutzen. Auf Homepage veröffentlichtes Arbeitsmanuskript, 8 S.

Aristoteles: Nikomachische Ethik, Buch VI. Verschiedene Ausgaben.

Gadamer, Hans-Georg (1960): Wahrheit und Methode. Tübingen: Mohr.

Grossmann, Igor (2017): Wisdom in Context. Perspectives on Psychological Science, 12(2), 233–257.

Habermas, Jürgen (1981): Theorie des kommunikativen Handelns. Frankfurt am Main: Suhrkamp.

King, Patricia M.; Kitchener, Karen S. (2004): Reflective Judgment: Theory and Research on the Development of Epistemic Assumptions Through Adulthood. Educational Psychologist, 39(1), 5–18.

Rosa, Hartmut (2016): Resonanz. Eine Soziologie der Weltbeziehung. Berlin: Suhrkamp.

Lempert, Robert J. (2019): Robust Decision Making. In: Marchau et al. (Hg.), Decision Making under Deep Uncertainty. Cham: Springer.

Hinweis zur Zitierweise: Die beiden Manuskripte sind die Primärquellen dieser Rekonstruktion. Die übrigen Titel dienen der begrifflichen Kontextualisierung; eine vollständige systematische Literaturrecherche oder Quellenverifikation ist nicht Gegenstand dieser Arbeitsfassung.

Schritt	Operative Leitfrage	Erwartetes Arbeitsergebnis
Unterscheiden	Was ist Befund, Deutung, Prognose, Wertung oder offene Frage?	Begriffs- und Evidenzklärung; markierte Unsicherheit.
Begründen	Welche Gründe tragen die Aussage; welche Übergänge sind strittig?	Offengelegte Annahmen, Quellen, Gegenargumente.
Verbinden	Welche Perspektiven ergänzen sich; wo bestehen echte Widersprüche?	Vergleich von Alternativen und Betroffenenpositionen.
Handeln	Welcher Schritt ist angesichts von Risiken, Rechten und Zuständigkeiten vertretbar?	Begründete Entscheidung samt Verantwortungszuordnung.
Überprüfen	Woran erkennen wir Irrtum, Nebenfolgen oder	Indikatoren, Fristen, Einspruchs-

	Revisionsbedarf? und Korrekturwege.
Übergang	Zu prüfende Brücke
Befund → Interpretation	Welche Modelle, Messgrenzen und Alternativerklärungen liegen zugrunde?
Interpretation → Bewertung	Welche Normen, Interessen und Betroffenenperspektiven werden eingeführt?
Bewertung → Entscheidung	Welche Optionen, Schäden, Rechte und Unsicherheiten sind abzuwägen?
Expertise → Legitimation	Wer ist entscheidungsbefugt, wer kann widersprechen und kontrollieren?
Entscheidung → Fortsetzung	Welche Beobachtung führt zu Anpassung, Stopp oder Rücknahme?
Dimension	Beobachtbarer Indikator
Epistemische Präzision	Trennung von Befund, Prognose, Annahme und Wertung.
Umgang mit Unsicherheit	Benennung fehlender Informationen und Geltungsgrenzen.
Perspektiven- und Machtprüfung	Relevante Gegenpositionen, Betroffene und Abhängigkeiten.
Argumentative Qualität	Explizite Begründungsbrücken und ernsthafte Einwände.
Verantwortung	Benannte Zuständigkeiten, Rechte und verhältnismäßige Optionen.
Revision	Konkrete Messgrößen, Termine und Korrekturauslöser.

Teil II · Erfahrung, Wirkung und die Grenzen kausaler Schlüsse

1. Erkenntnis aus gelebter Erfahrung

Erfahrungen besitzen eine eigentümliche Unmittelbarkeit: Sie widerfahren konkreten Menschen in unwiederholbaren Lebenslagen und können bislang unbemerkte Zusammenhänge erschließen. Doch die Evidenz eines Erlebnisses ist nicht identisch mit der Evidenz seiner Erklärung. Zwischen dem Satz „Nach der Veränderung trat eine Besserung ein“ und der weiterreichenden Behauptung „Die Veränderung verursachte die Besserung“ liegt ein anspruchsvoller inferenzieller Übergang. Gerade dieser Übergang bedarf einer eigenständigen Prüfung.

Die methodologische Unterscheidung zwischen beobachteter Veränderung, kausaler Zuschreibung und verantwortbarer Entscheidung bleibt unabhängig von personenbezogenen Vergleichsansätzen zentral.

Die methodologische Unterscheidung zwischen beobachteter Veränderung, kausaler Zuschreibung und verantwortbarer Entscheidung bleibt unabhängig von personenbezogenen Vergleichsansätzen zentral.

Der produktive methodische Widerstreit lässt sich in einer Frage verdichten: Was verändert sich tatsächlich, wodurch verändert es sich, und welche Folgerungen dürfen wir angesichts verbleibender Unsicherheit verantworten?

3. Das kontrafaktische Problem

Ein beobachtetes Ergebnis kann durch die Handlung selbst, durch Kontextbedingungen oder durch unbeobachtete Einflüsse zustande kommen. Der kausale Effekt ist deshalb von der bloßen zeitlichen Folge zu unterscheiden. Im kontrafaktischen Denkmodell wird für denselben Fall das Ergebnis mit einer Intervention dem hypothetischen Ergebnis ohne Intervention gegenübergestellt:

$$\tau_i = Y_i(1) - Y_i(0)$$

Da für denselben Menschen zur selben Zeit gewöhnlich nur eines dieser Ergebnisse beobachtbar ist, kann ein einzelner Selbstversuch die kontrafaktische Differenz nicht unmittelbar identifizieren. Wiederholte Beobachtung, vorab definierte Zielgrößen, geeignete Kontrollen und gegebenenfalls randomisierte Designs können die Aussagekraft erhöhen; sie beseitigen jedoch nicht sämtliche Unsicherheit.

4. Ein methodischer Dialog als Gedankenexperiment

Norbert: „Die Erfahrung eröffnet die Frage, aber sie entscheidet noch nicht über ihre Erklärung.“ – Franz (fiktive Gesprächsfigur): „Eine Erklärung muss sich dennoch an ihren Wirkungen bewähren.“ – Norbert: „Gewiss. Doch müssen wir unterscheiden, ob eine Wirkung beobachtet, einer Maßnahme ursächlich zugeschrieben oder als Handlungsgrund verallgemeinert wird.“ – Franz (fiktiv): „Dann wäre ein gelungenes Verfahren eines, das Veränderung ermöglicht und seine eigenen Wirkungsannahmen überprüfen lässt.“ Die hier Franz zugeschriebenen Sätze sind ausdrücklich keine dokumentierten Zitate, sondern eine philosophische Rekonstruktion möglicher Einwände.

Teil III · Heuristisches mathematisches Modell H5-WK

1. Dimensionen und Aggregation

Das H5-WK beschreibt fünf unterscheidbare Qualitäten: epistemische Prüffähigkeit E, Begründungsqualität B, Dialogfähigkeit D, Verantwortbarkeit V und Korrigierbarkeit K. Alle Größen werden für die heuristische Darstellung auf das Intervall [0,1] normiert. Eine mögliche Aggregation ist das gewichtete geometrische Mittel:

$$Q_t = E_t^{\alpha} \cdot B_t^{\beta} \cdot D_t^{\gamma} \cdot V_t^{\delta} \cdot K_t^{\varepsilon}$$

Bei $\alpha + \beta + \gamma + \delta + \varepsilon = 1$ ergibt sich eine normierte gewichtete Form. Die multiplikative Verknüpfung formuliert die normative Annahme, dass schwerwiegende Defizite einer Dimension nicht vollständig durch Stärken anderer Dimensionen kompensiert werden dürfen. Die Gewichtung ist vor einer empirischen Prüfung festzulegen; ihre Angemessenheit ist selbst Gegenstand der Validierung.

$$Q_t = (E_t \cdot B_t \cdot D_t \cdot V_t \cdot K_t)^{(1/5)}$$

2. Urteil, Handlung und Ergebnis

Die Qualität des Urteilens Q ist nicht mit dem späteren Handlungsergebnis Y gleichzusetzen. Ein wohlbegründetes Urteil kann unter ungünstigen Umständen scheitern; ein unzureichend begründetes Urteil kann zufällig erfolgreich sein.

$$Y_{t+1} = f(A_t, C_t, U_t)$$

A bezeichnet die Handlung, C die Kontextbedingungen und U nicht kontrollierte Einflüsse. Die Wirkung der Handlung ist erst dann kausal zuordenbar, wenn alternative Erklärungen mit geeigneten Methoden geprüft werden.

3. Lernen und Korrektur

$$S_{t+1} = \mathcal{L}(S_t, A_t, Y_{t+1}, F_t)$$

S bezeichnet den jeweiligen Erkenntniszustand, $\mathcal{L}$ eine Lernregel und F Rückmeldungen aus Kritik, Dialog und Beobachtung. Erkenntnisfortschritt kann auch in der begründeten Verringerung subjektiver Gewissheit bestehen.

4. Wissens- und Machtasymmetrie

$$R_t(\text{erforderlich}) = r_0 + \lambda \cdot M_t$$

M bezeichnet die Stärke einer relevanten Wissens- und Machtasymmetrie. Die Gleichung formuliert keine empirisch bestätigte Gesetzmäßigkeit, sondern eine überprüfungsbedürftige normative Leitregel: Mit zunehmender Asymmetrie steigen die Anforderungen an Transparenz, Rechenschaft, Beteiligung und institutionell gesicherte Korrektur.

5. Grenzen der Formalisierung

Die numerische Darstellung darf keine Scheingenauigkeit erzeugen. Neben Q sind stets die fünf Einzelwerte, die zugrunde liegenden Textbelege und die Unsicherheiten der Bewertung auszuweisen. Der Index ersetzt weder praktische Klugheit noch demokratische Legitimation.

Teil IV · Integrierter Anhang A: Nutzennachweis

1. Was mit „Nutzen“ gemeint ist

Der Weisheitskompass beansprucht nicht, automatisch richtige Entscheidungen hervorzubringen. Sein zunächst überprüfbarer Nutzen besteht darin, die Qualität des Urteilens und der Entscheidungsbegründung zu verbessern. Ein gut begründetes Urteil kann sich wegen unvorhersehbarer Entwicklungen als sachlich unzutreffend erweisen; umgekehrt kann eine schlecht begründete Entscheidung zufällig erfolgreich sein. Ergebnisqualität und Verfahrensqualität sind deshalb getrennt zu erfassen.

Als Nutzennachweis gelten nicht die bloße Zustimmung zum Modell, subjektive Begeisterung nach einem Workshop oder die Zahl verwendeter Fachbegriffe. Aussagekräftiger ist, ob Personen an neuen Fällen beobachtbar besser zwischen Befund, Deutung und Wertung unterscheiden, relevante Gegenargumente erkennen, Wissens- und Machtasymmetrien benennen, Alternativen begründen und überprüfbare Korrekturbedingungen formulieren.

2. Wirkmodell: Wie der Kompass Nutzen erzeugen könnte

Die vermutete Wirkungskette lautet: strukturierte Reflexionsschritte → explizite Prüfung von Begründungsübergängen → verbesserte Qualität dokumentierter Urteile → höhere Nachvollziehbarkeit, Kritikfähigkeit und Revidierbarkeit gemeinsamer Entscheidungen. Der letzte Übergang zu besseren gesellschaftlichen Folgen bleibt eine weitergehende Hypothese.

2.1 Die fünf Schritte als überprüfbare Intervention

Unterscheiden – Was ist belegt, was wird interpretiert, was ist Werturteil? Wie sicher sind Quellen und Befunde?

Begründen – Welche Gründe tragen eine Behauptung oder Entscheidung? Wo liegen unbegründete Übergänge?

Verbinden – Welche Perspektiven, Interessen, Betroffenen und Gegenargumente fehlen? Wer verfügt über Wissen und Entscheidungsmacht?

Handeln – Welche Alternativen sind unter Unsicherheit verhältnismäßig? Wer trägt die Folgen und die Verantwortung?

Überprüfen – An welchen Indikatoren, zu welchem Zeitpunkt und durch wen wird die Entscheidung erneut beurteilt?

Die Eigenleistung des Modells ist als Integrationshypothese zu verstehen: Die einzelnen Verfahren sind in der Forschung nicht neu; zu prüfen ist der Zusatznutzen ihrer geordneten Verbindung, besonders der expliziten Betrachtung von Wissensverteilung, Macht, Legitimation und Korrektur.

3. Nutzen auf drei Ebenen

Individuum

Erwartet wird eine präzisere Unsicherheitswahrnehmung, geringere Neigung zu vorschnellen Schlussfolgerungen und eine besser begründete Wahl zwischen Alternativen. Messbar ist dies anhand schriftlicher Fallbearbeitungen, nicht anhand einer Selbsteinschätzung „Ich denke jetzt kritischer“.

Gruppe und Organisation

Erwartet werden dokumentierte Gegenpositionen, erkennbare Verantwortlichkeiten und klare Auslöser für die Überprüfung einer Entscheidung. Messbar sind etwa die Qualität von Sitzungsprotokollen, die Berücksichtigung von Minderheitsargumenten und die Zahl tatsächlich nachverfolgter Prüfaufträge. Mehr Dokumentation ist jedoch nicht automatisch bessere Entscheidungskultur.

Gesellschaft und Institution

Ein möglicher Nutzen liegt darin, Expertisen auch für Nichtfachleute kritisierbar zu machen, Betroffene sichtbar einzubeziehen und Entscheidungsbefugnisse von Sachkompetenz zu unterscheiden. Institutionelle Transparenz, Rechtsschutz und demokratische Legitimation kann der Weisheitskompass unterstützen, aber nicht ersetzen. Aussagen über gesellschaftliche Wirkung erfordern eigene längerfristige Untersuchungen.

4. Konkrete Indikatoren und Bewertungsraster

Für eine standardisierte Fallaufgabe werden sechs Dimensionen auf einer Skala von 0 bis 3 Punkten beurteilt. Zwei geschulte, gegenüber der Gruppenzugehörigkeit verblindete Personen bewerten unabhängig voneinander. Die Gesamtskala reicht von 0 bis 18 Punkten; zusätzlich werden die Einzeldimensionen berichtet.

Bewertungsanker je Dimension: 0 = fehlt oder sachlich verfehlt; 1 = angedeutet, aber unspezifisch; 2 = sachlich zutreffend und nachvollziehbar; 3 = differenziert, fallbezogen und mit relevanten Grenzen bzw. Gegenargumenten. Vor Studienbeginn sind Beispielantworten für alle Stufen zu entwickeln und zu erproben.

Ergänzende Messgrößen: Bearbeitungszeit; Verständlichkeit für Laien; Qualität einer mündlichen Begründung; Bereitschaft, auf neue Evidenz hin die Entscheidung anzupassen. Die Bearbeitungszeit ist wichtig, weil ein Verfahren unter Zeitdruck praktisch untauglich sein kann, obwohl es im Seminar gute Ergebnisse erzielt.

5. Studiendesign für einen belastbaren ersten Wirksamkeitstest

5.1 Forschungsfragen und Hypothesen

Primäre Forschungsfrage: Verbessert eine kurze Schulung im Weisheitskompass die Qualität der Begründung unbekannter Entscheidungsfälle stärker als eine zeitgleiche aktive Vergleichsschulung?

H1: Die Interventionsgruppe erreicht nach der Schulung höhere, vorab definierte Werte im Gesamtbewertungsraster als die aktive Vergleichsgruppe, unter Kontrolle der Ausgangsleistung.

H2: Besonders ausgeprägt sind Unterschiede bei der Identifikation von Wissens- und Machtasymmetrien sowie bei expliziten Korrekturbedingungen.

H3 (explorativ): Der Zugewinn bleibt in einer erneuten Fallaufgabe nach vier bis acht Wochen teilweise erhalten.

H4 (explorativ): Die Qualität steigt nicht nur durch längere Bearbeitungszeit; Zeitaufwand und Nutzen werden gemeinsam berichtet.

5.2 Stichprobe und Gruppen

Als Pilotstudie kommen beispielsweise 60–90 erwachsene Teilnehmende aus Bildung, Verwaltung oder Zivilgesellschaft infrage, zufällig auf zwei oder drei Gruppen verteilt. Diese Größenordnung ist ein Planungsbeispiel, keine statistisch begründete Fallzahl. Für eine bestätigende Studie muss die Stichprobengröße anhand der erwarteten Effektgröße, der Streuung und der gewünschten Teststärke vorab berechnet werden.

Gruppe A erhält eine standardisierte Einführung in den Weisheitskompass mit Übungsfällen. Gruppe B erhält eine gleich lange, ernstzunehmende Einführung in ein etabliertes allgemeines Reflexions- oder Entscheidungsverfahren. Optional kann Gruppe C eine reine Informationsveranstaltung erhalten. Nur der Vergleich mit Gruppe B prüft den spezifischen Zusatznutzen gegenüber vorhandenen Methoden.

5.3 Ablauf und Auswertung

T0: Ausgangstest mit einer unbekannten, standardisierten Fallvignette; Erfassung relevanter Vorkenntnisse.

Intervention: gleiche Dauer, gleichwertige Lehrpersonenmaterialien und vergleichbare Übungsintensität in allen Gruppen.

T1: neuer, im Schwierigkeitsgrad möglichst vergleichbarer Fall unmittelbar danach; keine identischen Aufgaben wiederholen.

T2: optionaler Nachtest nach vier bis acht Wochen mit drittem Fall zur Prüfung des Transfers.

Verblindete Bewertung durch zwei unabhängige Rater; Prüfung ihrer Übereinstimmung; Konfliktklärung nach festgelegter Regel.

Vorabregistrierte primäre Zielgröße und statistische Auswertung (z. B. Gruppenvergleich der T1-Werte unter Kontrolle von T0); Effektgrößen und Konfidenzintervalle berichten.

Wesentlich sind zufällige Zuteilung, eine aktive Kontrollbedingung, möglichst vergleichbare Fälle, transparente Ausfallquoten und eine Auswertung, die auch ausbleibende oder negative Effekte sichtbar macht. Die Untersuchenden sollten die eigene Rolle als Entwickler oder Befürworter offenlegen.

6. Fallbeispiel: Einführung eines KI-Systems in einer öffentlichen Einrichtung

Ausgangsfall: Eine Behörde erwägt ein KI-gestütztes System zur Priorisierung eingehender Anträge. Ein Anbieter verspricht schnellere Bearbeitung; die zugrunde liegenden Daten, Fehlerquoten und die Auswirkungen auf unterschiedliche Bevölkerungsgruppen sind nur teilweise bekannt. Der Fall ist eine didaktische Vignette, kein Bericht über eine konkrete Behörde.

Eine schwache Antwort übernimmt das Effizienzversprechen, empfiehlt die Einführung und nennt „laufende Kontrolle“, ohne deren Inhalt festzulegen. Eine starke Antwort trennt überprüfte Leistungsdaten von Anbieterbehauptungen; verlangt eine unabhängige Prüfung der Fehlerverteilung; benennt die Informationsmacht des Anbieters und die Verantwortung der Behörde; berücksichtigt Betroffene und Rechtsbehelfe; vergleicht Pilotbetrieb, Verzicht und Alternativen; definiert vorab Abbruchkriterien, Zuständigkeiten und einen Evaluationstermin.

Ein solcher Unterschied ist anhand des Bewertungsrasters beobachtbar. Er belegt jedoch noch nicht, dass die tatsächlich getroffene Behördenentscheidung gerechter, rechtmäßiger oder wirtschaftlicher ausfällt. Dazu wären reale Prozess- und Ergebnisdaten notwendig.

7. Grenzen und mögliche Gegenbefunde

Selektionsverzerrung: Besonders reflektierte Personen melden sich eher freiwillig.

Lehrpersonen- und Erwartungseffekte: Die Begeisterung für das Modell kann die Antworten beeinflussen.

Training auf das Raster: Teilnehmende lernen möglicherweise nur, erwartete Schlüsselwörter zu verwenden. Deshalb neue Transferfälle und unabhängige Bewertungen einsetzen.

Soziale Erwünschtheit: Das Benennen von „Macht“ oder „Dialog“ kann zur Formel werden, ohne eine konkrete Analyse zu verbessern.

Zeitkosten: Eine höhere Punktzahl bei unverhältnismäßig großem Aufwand kann den praktischen Nutzen begrenzen.

Kontextgrenzen: Ein Instrument individueller Reflexion kann institutionelle Abhängigkeiten, fehlende Einsichtsrechte oder unfaire Machtstrukturen nicht selbst beseitigen.

Ein ausbleibender Vorteil gegenüber einer guten aktiven Vergleichsmethode wäre ein ernst zu nehmendes Ergebnis: Dann wäre der Weisheitskompass möglicherweise eine didaktisch nützliche Darstellung bekannter Verfahren, aber kein nachweislich überlegenes Verfahren. Auch dieser Befund könnte zur Präzisierung des Modells beitragen.

8. Vom Pilotversuch zum gesellschaftlichen Nutzennachweis

Eine stufenweise Evaluation ist angemessen: (1) Verständlichkeit und Akzeptanz prüfen; (2) unmittelbare Verbesserung der Argumentationsqualität in kontrollierten Fallaufgaben testen; (3) Transfer in neue Fälle und Nachhaltigkeit untersuchen; (4) in Organisationen dokumentierte Entscheidungen und Revisionspraxis über mehrere Monate vergleichen; (5) erst danach mögliche Folgen für Fairness, Beteiligung, Vertrauen und Fehlentscheidungen untersuchen. Für jede Stufe müssen geeignete Vergleichsmaßstäbe und datenschutzkonforme Erhebungen festgelegt werden.

Ein überzeugender gesellschaftlicher Nachweis verlangt mehr als ein gutes Seminarergebnis: Es müsste gezeigt werden, dass das Verfahren unter realen Bedingungen von Machtungleichheit, knapper Zeit und kontroversen Interessen praktikabel ist und zusätzliche Verbesserungen gegenüber der bisherigen Praxis bewirkt.

9. Ergebnis: Was gegenwärtig behauptet werden darf

Bereits begründbar ist der konzeptionelle Nutzen: Der Weisheitskompass macht kritische Übergänge zwischen Wissen, Deutung, Wertung und Handlung sichtbar und verbindet sie mit Dialog, Machtkritik und Revisionsbereitschaft. Plausibel, aber noch unbewiesen, ist seine lernförderliche Wirkung. Empirisch offen sind seine Überlegenheit gegenüber Alternativen und seine längerfristigen gesellschaftlichen Folgen.

Die wissenschaftlich redliche Leitformel lautet: Der Weisheitskompass ist ein prüfbares Verfahren zur Förderung begründeter und korrigierbarer Urteilsbildung; ob und in welchem Umfang es einen Zusatznutzen erzeugt, ist Gegenstand eines angelegten Forschungsprogramms.

Hinweise zur Quellen- und Forschungsbasis

Dieser Anhang entwickelt das in einer formalen Seminararbeit skizzierte Forschungsprogramm methodisch weiter. Er berichtet keine eigenen empirischen Daten und beansprucht keinen bereits erbrachten Wirksamkeitsnachweis. Fachliche Anschlussstellen sind Forschung zu Metakognition, reflektiertem Urteilen, intellektueller Bescheidenheit, Perspektivenübernahme, Wertklärung sowie adaptiver Entscheidungsfindung; die bibliografischen Angaben der Hauptarbeit gelten fort. Für eine tatsächliche Studie sind Messinstrumente, Fallvignetten, Bewertungsanker, Analyseplan und Ethik-/Datenschutzkonzept vor Beginn vollständig zu dokumentieren.

Dimension	Beobachtbares Kriterium
Epistemische Trennung	Befund, Annahme, Deutung und Wertung korrekt unterscheiden
Unsicherheitskalibrierung	Quellenlage, Grenzen und eigene Sicherheit angemessen benennen
Perspektiven und Macht	Betroffene, Gegenpositionen, Informationsvorsprünge und Interessen erkennen
Begründungsbrücken	Übergänge von Daten zu Deutungen, Werten und Handlungen ausdrücklich begründen
Handlungsalternativen	Mindestens zwei realistische Optionen mit Folgen und Zuständigkeiten vergleichen
Korrigierbarkeit	Prüfindikator, Termin, Verantwortliche und Änderungsauslöser festlegen

Teil V · Integrierter Anhang B: Operationalisierung und Pilotstudie

1. Ziel und wissenschaftlicher Status

Das H5-WK erfasst nicht „Weisheit“ als Persönlichkeitseigenschaft, sondern die beobachtbare Qualität einer begründeten Entscheidung unter Unsicherheit. Fünf Dimensionen werden anhand schriftlicher Fallbearbeitungen durch unabhängige Beurteilende bewertet. Das Instrument ist eine zu prüfende Neuentwicklung, kein validierter Test.

2. Fünf Dimensionen und zwölf Beobachtungsindikatoren

E – Epistemische Prüffähigkeit

E1: Trennt Befunde, Deutungen und Wertungen ausdrücklich.

E2: Benennt Unsicherheit, Evidenzlücken und Grenzen eigener Expertise.

E3: Prüft Herkunft, Qualität und Gegenbelege zentraler Aussagen.

B – Begründungsqualität

B1: Macht die Brücke von Befund zu Bewertung und Handlung explizit.

B2: Vergleicht mindestens zwei realistische Alternativen.

D – Dialogfähigkeit

D1: Gibt eine Gegenposition fair und nachvollziehbar wieder.

D2: Reagiert auf ein starkes Gegenargument mit Begründung oder Revision.

V – Verantwortbarkeit

V1: Legt betroffene Werte und Zielkonflikte offen.

V2: Identifiziert Betroffene, Interessen sowie Wissens- und Machtasymmetrien.

V3: Benennt Zuständigkeiten, Risiken und Grenzen legitimer Entscheidungsmacht.

K – Korrigierbarkeit

K1: Formuliert beobachtbare Erfolgskriterien und einen Prüfzeitpunkt.

K2: Benennt konkrete Auslöser für Anpassung, Abbruch oder Neubewertung.

3. Bewertungsrubrik je Indikator

Jeder Indikator erhält 0–3 Punkte. Dimensionen werden als Mittelwert ihrer Indikatoren auf 0–1 normiert. Neben dem Profil der fünf Dimensionen wird ein Gesamtwert nur explorativ berichtet. Bei fehlender Beurteilbarkeit wird „nicht bewertbar“ statt 0 vergeben; fehlende Werte werden gesondert ausgewiesen.

4. Formale Auswertung

Für Dimension j mit n_j bewertbaren Indikatoren gilt: x_j = (Summe der Indikatorpunkte)/(3 · n_j). Der vorläufige H5-WK-Index lautet Q = (E · B · D · V · K)^(1/5). Die Gleichgewichtung ist eine Annahme; additive Alternativen und Einzelprofile werden parallel ausgewertet. Ein Nullwert führt im Produkt zu Q = 0: Dieser starke Engpass-Effekt muss empirisch und normativ überprüft werden.

Primärer Studienendpunkt sollte NICHT allein Q sein, sondern eine vorab definierte Gesamtbewertung anhand der zwölf Indikatoren, vorzugsweise ergänzt um die fünf Dimensionswerte. Sonst könnte die mathematische Form die Ergebnisse stärker bestimmen als die beobachtete Urteilsqualität.

5. Pilotstudie: drei Vergleichsbedingungen

Gruppe A: Einführung in den Weisheitskompass mit fünf Schritten und Übungen.

Gruppe B: zeitgleiche Schulung mit einem etablierten allgemeinen Reflexionsschema.

Gruppe C: Bearbeitung derselben Aufgaben ohne strukturierte Methode (gleicher Zeitaufwand).

Geeignete erste Pilotierung: etwa 15–25 Personen je Gruppe als Machbarkeitsstudie, nicht als belastbare Wirksamkeitsprüfung. Zuteilung möglichst randomisiert; Bewertende kennen die Gruppenzugehörigkeit nicht. Vor- und Nachtest verwenden unterschiedliche, hinsichtlich Schwierigkeit vorgeprüfte Fälle; ein verzögerter Transfertest nach 4–8 Wochen ist sinnvoll. Eine spätere Hauptstudie benötigt eine vorab berechnete Stichprobengröße.

6. Standardisierte Fallaufgabe

Eine Gemeinde erwägt ein KI-System zur Priorisierung sozialer Unterstützungsangebote. Der Anbieter verspricht höhere Effizienz, legt aber Trainingsdaten und Fehlerraten nur teilweise offen. Betroffene befürchten Benachteiligung. Das Budget ist begrenzt. Entscheiden Sie über Pilotierung, Ablehnung oder alternative Vorgehensweise und begründen Sie Ihre Entscheidung.

Arbeitsauftrag: (a) Befunde/Annahmen/Werte trennen; (b) zwei Alternativen darstellen; (c) stärkstes Gegenargument formulieren; (d) Macht und Betroffene identifizieren; (e) nächsten verantwortbaren Schritt wählen; (f) Prüfkriterien, Zuständigkeit und Abbruchbedingung festlegen.

7. Prüfung des Nutzens

Präregistrierte Hauptfrage: Verbessert die H5-WK-Gruppe ihre blind bewertete Urteilsqualität gegenüber Gruppe B stärker? Ergänzend wird Gruppe C verglichen. Die Veränderungsdifferenz kann als (Nachher–Vorher)_A – (Nachher–Vorher)_B beschrieben und mit geeigneten statistischen Modellen, Unsicherheitsintervallen und Sensitivitätsanalysen geschätzt werden. Eine positive Differenz allein belegt ohne robustes Design noch keine Kausalität.

Zusätzlich prüfen: Übereinstimmung der Bewertenden (z. B. gewichtetes Kappa bzw. ICC je nach Skala), Verständlichkeit, Bearbeitungszeit, Abbruchquote, Belastung und Übertragbarkeit auf neue Fälle. Verbesserte Fallargumentation ist nicht gleichbedeutend mit besseren realen Folgen.

8. Bezug zum wirkungsorientierten Denken

Bezug zum Der Ansatz des wirkungsorientierten Denkens von Prof. DDr. Helmut Retzl verdient als wertvoller gedanklicher Impuls ausdrückliche Würdigung. Die vorliegende Gegenüberstellung greift diese Perspektive mit Respekt und Anerkennung auf, verdeutlicht inhaltliche Berührungspunkte und wahrt zugleich die Eigenständigkeit sowie die unterschiedlichen Schwerpunktsetzungen beider Ansätze.

9. Falsifizierbarkeit und Weiterentwicklung

Das Modell wäre zu revidieren, wenn die Rubrik unzuverlässig ist, die fünf Dimensionen nicht unterscheidbar sind, der zusätzliche Nutzen gegenüber etablierten Verfahren ausbleibt oder der Zeitaufwand unverhältnismäßig hoch ist. Solche Befunde wären Erkenntnisgewinn, kein Scheitern wissenschaftlicher Redlichkeit.

Leitsatz: Ein Verfahren verantwortlicher Urteilsbildung muss die Prüfung seiner eigenen Wirksamkeit zulassen.

Punkte	Beobachtbare Ausprägung
0	Nicht vorhanden oder sachlich verfehlt.
1	Angedeutet, aber unklar, einseitig oder nicht begründet.
2	Nachvollziehbar und überwiegend sachgerecht, mit erkennbaren Lücken.
3	Präzise, differenziert, fallbezogen und kritisch reflektiert.

Teil VI · Schlussbetrachtung

Der Weisheitskompass gewinnt seine philosophische Kontur nicht aus einem Anspruch auf Unfehlbarkeit, sondern aus der sorgfältig geregelten Beziehung von Erfahrung, Begründung, Dialog und Verantwortung. Seine praktische Bewährungsprobe liegt darin, ob er unter realen Bedingungen differenziertere Urteile hervorbringt, relevante Machtasymmetrien sichtbar macht und Entscheidungen für begründete Korrekturen offenhält. Die mathematische Heuristik macht diese Erwartungen explizit; die empirische Prüfung muss erst erweisen, ob sie tragfähig sind. Gerade die Möglichkeit des Scheiterns gehört zu einem ernsthaft wissenschaftlichen Verständnis des Modells.

Quellen- und Statusvermerk

Die methodologische Unterscheidung zwischen beobachteter Veränderung, kausaler Zuschreibung und verantwortbarer Entscheidung bleibt unabhängig von personenbezogenen Vergleichsansätzen zentral.

Teil V · Integrierte Forschungsanhänge

Die folgenden Anhänge bündeln die früheren Forschungsentwürfe und die jüngste Präzisierung der 18-Punkte-Skala. Unterschiedliche Rubriken werden als Entwicklungsstufen ausgewiesen, nicht als austauschbare validierte Messinstrumente.

Anhang A – Der Nutzennachweis des Weisheitskompasses

Konzeptionelle Wirkannahmen, messbare Kriterien und ein empirisches Prüfdesign

Ergänzung zur Seminararbeit „Der Weisheitskompass – Vom persönlichen Orientierungsmodell zum prüfbaren Rahmen verantwortlicher Urteilsbildung“ · Arbeitsstand Oktober 2026

1. Was mit „Nutzen“ gemeint ist

Der Weisheitskompass beansprucht nicht, automatisch richtige Entscheidungen hervorzubringen. Sein zunächst überprüfbarer Nutzen besteht darin, die Qualität des Urteilens und der Entscheidungsbegründung zu verbessern. Ein gut begründetes Urteil kann sich wegen unvorhersehbarer Entwicklungen als sachlich unzutreffend erweisen; umgekehrt kann eine schlecht begründete Entscheidung zufällig erfolgreich sein. Ergebnisqualität und Verfahrensqualität sind deshalb getrennt zu erfassen.

Als Nutzennachweis gelten nicht eine bloße Zustimmung zum Modell, subjektive Begeisterung nach einem Workshop oder die Zahl verwendeter Fachbegriffe. Aussagekräftiger wird, ob Personen an neuen Fällen beobachtbar besser zwischen Befund, Deutung und Wertung unterscheiden, relevante Gegenargumente erkennen, Wissens- und Machtasymmetrien benennen, Alternativen begründen und überprüfbare Korrekturbedingungen formulieren.

2. Wirkmodell: Wie der Kompass Nutzen erzeugen könnte

Die vermutete Wirkungskette lautet: strukturierte Reflexionsschritte → explizite Prüfung von Begründungsübergängen → verbesserte Qualität dokumentierter Urteile → höhere Nachvollziehbarkeit, Kritikfähigkeit und Revidierbarkeit gemeinsamer Entscheidungen. Der letzte Übergang zu besseren gesellschaftlichen Folgen bleibt eine weitergehende Hypothese.

2.1 Die fünf Schritte als überprüfbare Intervention

Unterscheiden – Was ist belegt, was wird interpretiert, was ist Werturteil? Wie sicher sind Quellen und Befunde?

Begründen – Welche Gründe tragen eine Behauptung oder Entscheidung? Wo liegen unbegründete Übergänge?

Verbinden – Welche Perspektiven, Interessen, Betroffenen und Gegenargumente fehlen? Wer verfügt über Wissen und Entscheidungsmacht?

Handeln – Welche Alternativen sind unter Unsicherheit verhältnismäßig? Wer trägt die Folgen und die Verantwortung?

Überprüfen – An welchen Indikatoren, zu welchem Zeitpunkt und durch wen wird die Entscheidung erneut beurteilt?

Die Eigenleistung des Modells wird als Integrationshypothese verstanden:

Die einzelnen Verfahren sind in der Forschung nicht neu; zu prüfen ist der Zusatznutzen ihrer geordneten Verbindung, der expliziten Betrachtung von Wissensverteilung, Macht, Legitimation und Korrektur.

3. Nutzen auf drei Ebenen

Individuum

Erwartet wird eine präzisere Unsicherheitswahrnehmung, geringere Neigung zu vorschnellen Schlussfolgerungen und eine besser begründete Wahl zwischen Alternativen. Messbar ist dies anhand schriftlicher Fallbearbeitungen, nicht anhand einer Selbsteinschätzung „Ich denke jetzt kritischer".

Gruppe und Organisation

Erwartet werden dokumentierte Gegenpositionen, erkennbare Verantwortlichkeiten und klare Auslöser für die Überprüfung einer Entscheidung. Messbar sind etwa die Qualität von Sitzungsprotokollen, die Berücksichtigung von Minderheitsargumenten und die Zahl tatsächlich nachverfolgter Prüfaufträge. Mehr Dokumentation ist jedoch nicht automatisch bessere Entscheidungskultur.

Gesellschaft und Institution

Ein möglicher Nutzen liegt darin, Expertisen auch für Nichtfachleute kritisierbar zu machen, Betroffene sichtbar einzubeziehen und Entscheidungsbefugnisse von Sachkompetenz zu unterscheiden. Institutionelle Transparenz, Rechtsschutz und demokratische Legitimation kann der Weisheitskompass unterstützen, aber nicht ersetzen. Aussagen über gesellschaftliche Wirkung erfordern eigene längerfristige Untersuchungen.

4. Konkrete Indikatoren und Bewertungsraster

Für eine standardisierte Fallaufgabe werden sechs Dimensionen auf einer Skala von 0 bis 3 Punkten beurteilt. Zwei geschulte, gegenüber der Gruppenzugehörigkeit verblindete Personen bewerten unabhängig voneinander. Die Gesamtskala reicht von 0 bis 18 Punkten; zusätzlich werden die Einzeldimensionen berichtet.

Dimension	Beobachtbares Kriterium
Epistemische Trennung	Befund, Annahme, Deutung und Wertung korrekt unterscheiden
Unsicherheitskalibrierung	Quellenlage, Grenzen und eigene Sicherheit angemessen benennen
Perspektiven und Macht	Betroffene, Gegenpositionen, Informationsvorsprünge und Interessen erkennen
Begründungsbrücken	Übergänge von Daten zu Deutungen, Werten und Handlungen ausdrücklich begründen
Handlungsalternativen	Mindestens zwei realistische Optionen mit Folgen

	und Zuständigkeiten vergleichen
Korrigierbarkeit	Prüfindikator, Termin, Verantwortliche und Änderungsauslöser festlegen

Bewertungsanker je Dimension: 0 = fehlt oder sachlich verfehlt; 1 = angedeutet, aber unspezifisch; 2 = sachlich zutreffend und nachvollziehbar; 3 = differenziert, fallbezogen und mit relevanten Grenzen bzw. Gegenargumenten. Vor Studienbeginn sind Beispielantworten für alle Stufen zu entwickeln und zu erproben.

Ergänzende Messgrößen: Bearbeitungszeit; Verständlichkeit für Nichtfachleute; Qualität einer mündlichen Begründung; Bereitschaft, auf neue Evidenz hin die Entscheidung anzupassen. Die Bearbeitungszeit ist wichtig, weil ein Verfahren unter Zeitdruck praktisch untauglich sein kann, obwohl es im Seminar gute Ergebnisse erzielt.

5. Studiendesign für einen belastbaren ersten Wirksamkeitstest

5.1 Forschungsfragen und Hypothesen

Primäre Forschungsfrage: Verbessert eine kurze Schulung im Weisheitskompass die Qualität der Begründung unbekannter Entscheidungsfälle stärker als eine zeitgleiche aktive Vergleichsschulung?

- H1: Die Interventionsgruppe erreicht nach der Schulung höhere, vorab definierte Werte im Gesamtbewertungsraster als die aktive Vergleichsgruppe, unter Kontrolle der Ausgangsleistung.
- H2: Besonders ausgeprägt sind Unterschiede bei der Identifikation von Wissens- und Machtasymmetrien sowie bei expliziten Korrekturbedingungen.
- H3 (explorativ): Der Zugewinn bleibt in einer erneuten Fallaufgabe nach vier bis acht Wochen teilweise erhalten.
- H4 (explorativ): Die Qualität steigt nicht nur durch längere Bearbeitungszeit; Zeitaufwand und Nutzen werden gemeinsam berichtet.

5.2 Stichprobe und Gruppen

Als Pilotstudie kommen beispielsweise 60–90 erwachsene Teilnehmende aus Bildung, Verwaltung oder Zivilgesellschaft infrage, zufällig auf zwei oder drei Gruppen verteilt. Diese Größenordnung ist ein Planungsbeispiel, keine statistisch begründete Fallzahl. Für eine bestätigende Studie soll die Stichprobengröße anhand der erwarteten Effektgröße, der Streuung und der gewünschten Teststärke vorab berechnet werden.

Gruppe A erhält eine standardisierte Einführung in den Weisheitskompass mit Übungsfällen. Gruppe B erhält eine gleich lange, ernstzunehmende Einführung in ein etabliertes allgemeines Reflexions- oder Entscheidungsverfahren. Optional kann Gruppe C eine reine Informationsveranstaltung erhalten. Nur der Vergleich mit Gruppe B prüft den spezifischen Zusatznutzen gegenüber vorhandenen Methoden.

5.3 Ablauf und Auswertung

- T0: Ausgangstest mit einer unbekannten, standardisierten Fallvignette; Erfassung relevanter Vorkenntnisse.

- Intervention: gleiche Dauer, gleichwertige Lehrpersonenmaterialien und vergleichbare Übungsintensität in allen Gruppen.
- T1: neuer, im Schwierigkeitsgrad möglichst vergleichbarer Fall unmittelbar danach; keine identischen Aufgaben wiederholen.
- T2: optionaler Nachtest nach vier bis acht Wochen mit drittem Fall zur Prüfung des Transfers.
- Verblindete Bewertung durch zwei unabhängige Rater; Prüfung ihrer Übereinstimmung; Konfliktklärung nach festgelegter Regel.
- Vorabregistrierte primäre Zielgröße und statistische Auswertung (z. B. Gruppenvergleich der T1-Werte unter Kontrolle von T0); Effektgrößen und Konfidenzintervalle berichten.

Wesentlich sind zufällige Zuteilung, eine aktive Kontrollbedingung, möglichst vergleichbare Fälle, transparente Ausfallquoten und eine Auswertung, die auch ausbleibende oder negative Effekte sichtbar macht. Die Untersuchenden sollten die eigene Rolle als Entwickler oder Befürworter offenlegen.

6. Fallbeispiel: Einführung eines KI-Systems in einer öffentlichen Einrichtung

Ausgangsfall: Eine Behörde erwägt ein KI-gestütztes System zur Priorisierung eingehender Anträge. Ein Anbieter verspricht schnellere Bearbeitung; die zugrunde liegenden Daten, Fehlerquoten und die Auswirkungen auf unterschiedliche Bevölkerungsgruppen sind nur teilweise bekannt. Der Fall ist eine didaktische Vignette, kein Bericht über eine konkrete Behörde.

Eine schwache Antwort übernimmt das Effizienzversprechen, empfiehlt die Einführung und nennt „laufende Kontrolle“, ohne deren Inhalt festzulegen. Eine starke Antwort trennt überprüfte Leistungsdaten von Anbieterbehauptungen; verlangt eine unabhängige Prüfung der Fehlerverteilung; benennt die Informationsmacht des Anbieters und die Verantwortung der Behörde; berücksichtigt Betroffene und Rechtsbehelfe; vergleicht Pilotbetrieb, Verzicht und Alternativen; definiert vorab Abbruchkriterien, Zuständigkeiten und einen Evaluationstermin.

Ein solcher Unterschied wird anhand des Bewertungsrasters beobachtbar. Er belegt noch nicht, dass die tatsächlich getroffene Behördenentscheidung gerechter, rechtmäßiger oder wirtschaftlicher ausfällt. Dazu wären reale Prozess- und Ergebnisdaten notwendig.

7. Grenzen und mögliche Gegenbefunde

- Selektionsverzerrung: Besonders reflektierte Personen melden sich eher freiwillig.
- Lehrpersonen- und Erwartungseffekte: Die Begeisterung für das Modell kann die Antworten beeinflussen.
- Training auf das Raster: Teilnehmende lernen möglicherweise nur, erwartete Schlüsselwörter zu verwenden. Deshalb neue Transferfälle und unabhängige Bewertungen einsetzen.
- Soziale Erwünschtheit: Das Benennen von „Macht“ oder „Dialog“ kann zur Formel werden, ohne eine konkrete Analyse zu verbessern.
- Zeitkosten: Eine höhere Punktzahl bei unverhältnismäßig großem Aufwand kann den praktischen Nutzen begrenzen.
- Kontextgrenzen: Ein Instrument individueller Reflexion kann institutionelle Abhängigkeiten, fehlende Einsichtsrechte oder unfaire Machtstrukturen nicht selbst beseitigen.

Ein ausbleibender Vorteil gegenüber einer guten aktiven Vergleichsmethode wäre ein ernst zu nehmendes Ergebnis: Dann wäre der Weisheitskompass möglicherweise eine didaktisch nützliche

Darstellung bekannter Verfahren, aber kein nachweislich überlegenes Verfahren. Auch dieser Befund könnte zur Präzisierung des Modells beitragen.

8. Vom Pilotversuch zum gesellschaftlichen Nutzennachweis

Eine stufenweise Evaluation ist angemessen: (1) Verständlichkeit und Akzeptanz prüfen; (2) unmittelbare Verbesserung der Argumentationsqualität in kontrollierten Fallaufgaben testen; (3) Transfer in neue Fälle und Nachhaltigkeit untersuchen; (4) in Organisationen dokumentierte Entscheidungen und Revisionspraxis über mehrere Monate vergleichen; (5) erst danach mögliche Folgen für Fairness, Beteiligung, Vertrauen und Fehlentscheidungen untersuchen. Für jede Stufe müssen geeignete Vergleichsmaßstäbe und datenschutzkonforme Erhebungen festgelegt werden.

Ein überzeugender gesellschaftlicher Nachweis verlangt mehr als ein gutes Seminarergebnis: Es müsste gezeigt werden, dass das Verfahren unter realen Bedingungen von Machtungleichheit, knapper Zeit und kontroversen Interessen praktikabel ist und zusätzliche Verbesserungen gegenüber der bisherigen Praxis bewirkt.

9. Ergebnis: Was gegenwärtig behauptet werden darf

Bereits begründbar ist der konzeptionelle Nutzen: Der Weisheitskompass macht kritische Übergänge zwischen Wissen, Deutung, Wertung und Handlung sichtbar und verbindet sie mit Dialog, Machtkritik und Revisionsbereitschaft. Plausibel, aber noch unbewiesen, ist seine lernförderliche Wirkung. Empirisch offen sind seine Überlegenheit gegenüber Alternativen und seine längerfristigen gesellschaftlichen Folgen.

Die wissenschaftlich redliche Leitformel lautet daher: Der Weisheitskompass ist ein prüfbares Verfahren zur Förderung begründeter und korrigierbarer Urteilsbildung; ob und unter welchen Bedingungen es tatsächlich einen Zusatznutzen erzeugt, ist Gegenstand eines ausdrücklich angelegten Forschungsprogramms.

Hinweise zur Quellen- und Forschungsbasis

Dieser Anhang entwickelt das in einer Seminararbeitsform skizzierte Forschungsprogramm methodisch weiter. Er berichtet keine eigenen empirischen Daten und beansprucht keinen bereits erbrachten Wirksamkeitsnachweis. Fachliche Anschlussstellen sind Forschung zu Metakognition, reflektiertem Urteilen, intellektueller Bescheidenheit, Perspektivenübernahme, Wertklärung sowie adaptiver Entscheidungsfindung; die bibliografischen Angaben der Hauptarbeit gelten fort.

Für eine tatsächliche Studie sind Messinstrumente, Fallvignetten, Bewertungsanker, Analyseplan und Ethik-/Datenschutzkonzept vor Beginn vollständig zu dokumentieren.

Anhang B · Ursprüngliche Operationalisierung mit zwölf Indikatoren

H5-WK – Operationalisierung und Pilotstudie

Forschungsanhang zum Weisheitskompass | Arbeitsfassung, Oktober 2026

1. Ziel und wissenschaftlicher Status

Das H5-WK erfasst nicht „Weisheit“ als Persönlichkeitseigenschaft, sondern die beobachtbare Qualität einer begründeten Entscheidung unter Unsicherheit. Fünf Dimensionen werden anhand schriftlicher Fallbearbeitungen durch unabhängige Beurteilende bewertet. Das Instrument ist eine zu prüfende Neuentwicklung, kein validierter Test.

2. Fünf Dimensionen und zwölf Beobachtungsindikatoren

E – Epistemische Prüffähigkeit

- E1: Trennt Befunde, Deutungen und Wertungen ausdrücklich.
- E2: Benennt Unsicherheit, Evidenzlücken und Grenzen eigener Expertise.
- E3: Prüft Herkunft, Qualität und Gegenbelege zentraler Aussagen.

B – Begründungsqualität

- B1: Macht die Brücke von Befund zu Bewertung und Handlung explizit.
- B2: Vergleicht mindestens zwei realistische Alternativen.

D – Dialogfähigkeit

- D1: Gibt eine Gegenposition fair und nachvollziehbar wieder.
- D2: Reagiert auf ein starkes Gegenargument mit Begründung oder Revision.

V – Verantwortbarkeit

- V1: Legt betroffene Werte und Zielkonflikte offen.
- V2: Identifiziert Betroffene, Interessen sowie Wissens- und Machtasymmetrien.
- V3: Benennt Zuständigkeiten, Risiken und Grenzen legitimer Entscheidungsmacht.

K – Korrigierbarkeit

- K1: Formuliert beobachtbare Erfolgskriterien und einen Prüfzeitpunkt.
- K2: Benennt konkrete Auslöser für Anpassung, Abbruch oder Neubewertung.

3. Bewertungsrubrik je Indikator

Punkte	Beobachtbare Ausprägung
0	Nicht vorhanden oder sachlich verfehlt.
1	Angedeutet, aber unklar, einseitig oder nicht begründet.
2	Nachvollziehbar und überwiegend sachgerecht, mit erkennbaren Lücken.
3	Präzise, differenziert, fallbezogen und kritisch reflektiert.

Jeder Indikator erhält 0–3 Punkte. Dimensionen werden als Mittelwert ihrer Indikatoren auf 0–1 normiert. Neben dem Profil der fünf Dimensionen wird ein Gesamtwert nur explorativ berichtet. Bei fehlender Beurteilbarkeit wird „nicht bewertbar“ statt 0 vergeben; fehlende Werte werden gesondert ausgewiesen.

4. Formale Auswertung

Für Dimension j mit n_j bewertbaren Indikatoren gilt: x_j = (Summe der Indikatorpunkte)/(3 · n_j). Der vorläufige H5-WK-Index lautet Q = (E · B · D · V · K)^(1/5). Die Gleichgewichtung ist eine Annahme; additive Alternativen und Einzelprofile werden parallel ausgewertet. Ein Nullwert führt im Produkt zu Q = 0: Dieser starke Engpass-Effekt muss empirisch und normativ überprüft werden.

Primärer Studienendpunkt sollte NICHT allein Q sein, sondern eine vorab definierte Gesamtbewertung anhand der zwölf Indikatoren, vorzugsweise ergänzt um die fünf Dimensionswerte. Sonst könnte die mathematische Form die Ergebnisse stärker bestimmen als die beobachtete Urteilsqualität.

5. Pilotstudie: drei Vergleichsbedingungen

- Gruppe A: Einführung in den Weisheitskompass mit fünf Schritten und Übungen.
- Gruppe B: zeitgleiche Schulung mit einem etablierten allgemeinen Reflexionsschema.
- Gruppe C: Bearbeitung derselben Aufgaben ohne strukturierte Methode (gleicher Zeitaufwand).

Geeignete erste Pilotierung: etwa 15–25 Personen je Gruppe als Machbarkeitsstudie, nicht als belastbare Wirksamkeitsprüfung. Zuteilung möglichst randomisiert; Bewertende kennen die Gruppenzugehörigkeit nicht. Vor- und Nachtest verwenden unterschiedliche, hinsichtlich Schwierigkeit vorgeprüfte Fälle; ein verzögerter Transfertest nach 4–8 Wochen ist sinnvoll. Eine spätere Hauptstudie benötigt eine vorab berechnete Stichprobengröße.

6. Standardisierte Fallaufgabe

Eine Gemeinde erwägt ein KI-System zur Priorisierung sozialer Unterstützungsangebote. Der Anbieter verspricht höhere Effizienz, legt aber Trainingsdaten und Fehlerraten nur teilweise offen. Betroffene befürchten Benachteiligung. Das Budget ist begrenzt. Entscheiden Sie über Pilotierung, Ablehnung oder alternative Vorgehensweise und begründen Sie Ihre Entscheidung.

Arbeitsauftrag: (a) Befunde/Annahmen/Werte trennen; (b) zwei Alternativen darstellen; (c) stärkstes Gegenargument formulieren; (d) Macht und Betroffene identifizieren; (e) nächsten verantwortbaren Schritt wählen; (f) Prüfkriterien, Zuständigkeit und Abbruchbedingung festlegen.

7. Prüfung des Nutzens

Präregistrierte Hauptfrage: Verbessert die H5-WK-Gruppe ihre blind bewertete Urteilsqualität gegenüber Gruppe B stärker? Ergänzend wird Gruppe C verglichen. Die Veränderungsdifferenz kann als (Nachher–Vorher)_A – (Nachher–Vorher)_B beschrieben und mit geeigneten statistischen Modellen, Unsicherheitsintervallen und Sensitivitätsanalysen geschätzt werden. Eine positive Differenz allein belegt ohne robustes Design noch keine Kausalität.

Zusätzlich prüfen: Übereinstimmung der Bewertenden (z. B. gewichtetes Kappa bzw. ICC je nach Skala), Verständlichkeit, Bearbeitungszeit, Abbruchquote, Belastung und Übertragbarkeit auf neue Fälle. Verbesserte Fallargumentation ist nicht gleichbedeutend mit besseren realen Folgen.

8. Bezug zum wirkungsorientierten Denken

Eine gedankliche Anschlussstelle besteht zum von Helmut Retzl entwickelten Ansatz des wirkungsorientierten Denkens (Retzl 2013). Während dieser Ansatz Veränderungen und Wirkungszusammenhänge menschlichen Handelns in den Blick nimmt, richtet der Weisheitskompass sein besonderes Augenmerk auf die Unterscheidung zwischen beobachteter Wirkung und begründeter Ursachenzuschreibung, auf die Verlässlichkeit der Erkenntnisgrundlagen sowie auf die ethische Verantwortbarkeit und Korrigierbarkeit daraus abgeleiteter Entscheidungen.

Der Ansatz des wirkungsorientierten Denkens von Prof. DDr. Helmut Retzl verdient als wertvoller gedanklicher Impuls ausdrückliche Würdigung. Die vorliegende Gegenüberstellung greift diese Perspektive mit Respekt und Anerkennung auf, verdeutlicht inhaltliche Berührungspunkte und wahrt zugleich die Eigenständigkeit sowie die unterschiedlichen Schwerpunktsetzungen beider Ansätze.

9. Falsifizierbarkeit und Weiterentwicklung

Das Modell wäre zu revidieren, sollte die Rubrik unzuverlässig sein, die fünf Dimensionen nicht unterscheidbar sind, der zusätzliche Nutzen gegenüber etablierten Verfahren ausbleibt oder der Zeitaufwand unverhältnismäßig hoch wird.

Solche Befunde wären Erkenntnisgewinn, kein Scheitern wissenschaftlicher Redlichkeit. Ein Verfahren verantwortlicher Urteilsbildung sollte auch eine weitere Prüfung seiner Wirksamkeit aufzeigen.

Anhang C · H5-WK-18: Bewertungsmanual der Neun-Indikatoren-Skala

Status: neue, verdichtete Forschungsfassung. Die 18-Punkte-Skala ersetzt für die geplante Hauptstudie nicht stillschweigend die zuvor entwickelten Sechs-Dimensionen- bzw. Zwölf-Indikatoren-Rubriken. Die Auswahl einer Endfassung ist vor der Hauptstudie zu begründen und zu präregistrieren.

Code	Indikator	Bewertungsanker (0 / 1 / 2)
E1	Befund, Deutung und Wertung	0: vermischt; 1: einzelne Trennung; 2: systematisch und fallbezogen getrennt.
E2	Quellen und Unsicherheit	0: ungeprüfte Übernahme; 1: allgemeiner Prüfvorbehalt; 2: konkrete Quellenprüfung und benannte Unsicherheit.
B1	Nachvollziehbare Begründung	0: unbegründete Forderung; 1: teilweise Gründe; 2: explizite Schlusskette von Befund zu Entscheidung.
B2	Alternativen und Gegenargumente	0: keine Alternativen; 1: bloße Erwähnung; 2: begründeter Vergleich realistischer Alternativen.
D1	Perspektivenvielfalt	0: monoperspektivisch; 1: weitere Perspektive genannt; 2: Perspektiven in ihrer sachlichen Relevanz berücksichtigt.
D2	Umgang mit Widerspruch	0: Einwand übergangen; 1: Einwand anerkannt; 2: sachgerecht beantwortet oder Position begründet revidiert.
V1	Werte, Interessen und Macht	0: ausgeblendet; 1: einzelne Aspekte; 2: Zielkonflikte und Asymmetrien explizit geprüft.
V2	Folgen und Betroffene	0: ausgeblendet; 1: einzelne Folgen; 2: konkrete Risiken, Betroffene und Schutzmaßnahmen.
K1	Überprüfung und Revision	0: keine Revision; 1: vage Überprüfung; 2: Kriterien, Zeitpunkt und Revisionsauslöser.

C.1 Formalisierung

$$S_i = \sum_{j=1}^{9} x_{ij}, \quad x_{ij} \in \{0, 1, 2\}, \quad 0 \leq S_i \leq 18$$

$$S_i = E_i + B_i + D_i + V_i + K_i$$

$$E, B, D, V \in [0, 4]; \quad K \in [0, 2]$$

$$\Delta S_i = S_{i,\text{nach}} - S_{i,\text{vor}}$$

Die ungleiche Dimensionengewichtung ist eine bislang nicht validierte Konstruktionsentscheidung. Der Gesamtwert misst dokumentierte Urteilsmerkmale in einer standardisierten Aufgabe, nicht die Weisheit einer Person. Fehlende Evidenz für einen Indikator ist von einem nachweislich unzutreffenden Argument zu unterscheiden; Aufgaben müssen allen Indikatoren eine faire Beobachtungsmöglichkeit eröffnen.

C.2 Standardisierte Fallaufgabe

Eine Gemeinde erwägt die Einführung eines KI-Systems zur Bearbeitung von Bürgeranliegen. Der Anbieter verspricht kürzere Bearbeitungszeiten. Zu prüfen sind Leistungsnachweise, Fehlerquoten, Datenschutz, digitale Zugänglichkeit, Alternativen, Zuständigkeiten und Kriterien für eine spätere Revision. Die Teilnehmenden erhalten identische Fakten und eine vorgegebene Bearbeitungszeit.

C.3 Bewertung und Gütekriterien

Zwei unabhängig arbeitende, möglichst hinsichtlich Gruppenzuteilung und Messzeitpunkt verblindete Beurteilende bewerten anonymisierte Antworten. Interrater-Übereinstimmung, Itemverteilungen, Decken- und Bodeneffekte sowie sprachliche und fachliche Störfaktoren sind in der Pilotstudie zu prüfen. Für den Gesamtwert eignet sich eine Intraklassenkorrelation; ordinale Einzelindikatoren können mittels gewichteter Kappa-Koeffizienten beurteilt werden.

Anhang D · Prospektive Fallzahl- und Auswertungsplanung

Primärer bestätigender Vergleich: Weisheitskompass gegen eine zeitlich und inhaltlich vergleichbare aktive Reflexionsmethode. Zweiseitiges $\alpha = 0{,}05$; Power = 0,80; hypothetischer standardisierter Effekt $d = 0{,}50$; Ausfallreserve = 15 %. Die Annahmen sind keine empirischen Effektbefunde.

$$d = (\mu_{WK} - \mu_{Vergleich}) / \sigma$$

$$n \text{ je Gruppe} \approx 2\,(z_{1-\alpha/2} + z_{1-\beta})^2 / d^2$$

$$n_{rekrutiert} = \lceil n_{auswertbar} / (1 - 0{,}15) \rceil$$

Bei $d = 0{,}50$: 64 auswertbare bzw. 76 zu rekrutierende Personen je Gruppe.

Die z-basierte Formel ist eine Näherung; die gerundete Zielgröße von 64 je Gruppe beruht auf einer nichtzentralen t-Test-Planung. Bei zwei Gruppen sind 152 Personen zu rekrutieren. Bei drei Gruppen und zwei gleichrangigen primären Vergleichen mit Bonferroni-Korrektur ist die konservative Planungsgröße etwa 276 Personen (92 je Gruppe).

D.1 Sensitivitätsanalyse

Angenommenes d	Auswertbar je Gruppe	Rekrutierung, 2 Gruppen / 15 % Ausfall
0,30	176	416
0,40	100	236
0,50	64	152
0,60	45	106
0,80	26	62

Für die Hauptanalyse wird eine ANCOVA des Nachher-Wertes unter Adjustierung des Ausgangswertes empfohlen. Eine entsprechend optimierte Fallzahlberechnung benötigt zusätzlich die empirische Vorher-Nachher-Korrelation und verlässliche Streuungsschätzungen. Der minimale praktisch bedeutsame Unterschied muss unabhängig von der erwarteten Effektgröße fachlich festgelegt werden.

D.2 Hypothetisches Rechenbeispiel – ausdrücklich keine Studienergebnisse

Beispielhafte Mittelwerte: Weisheitskompass 9,5 → 13,0; aktive Vergleichsgruppe 9,7 → 11,7; unstrukturierte Gruppe 9,6 → 10,5. Daraus ergeben sich rein rechnerisch Differenzen der Veränderungen von 1,5 bzw. 2,6 Punkten. Ohne reale Einzelwerte und Streuungen ist weder ein Signifikanztest noch ein Wirksamkeitsnachweis möglich.

$$\Delta WK\text{–}B = (13{,}0 - 9{,}5) - (11{,}7 - 9{,}7) = 1{,}5 \text{ Punkte}$$

D.3 Pilotstudie und Präregistrierung

Eine Pilotstudie mit etwa 60–90 Personen dient der Prüfung von Machbarkeit, Bewertungsreliabilität, Streuung, Aufgabenäquivalenz und Ausfallraten, nicht der bestätigenden Wirksamkeitsprüfung. Vor der Hauptstudie werden Endpunkt, Rubrikfassung, Ausschlussregeln, Hypothesen, Auswertungsmodell, Umgang mit fehlenden Werten und Testkorrektur verbindlich präregistriert

Schlussbemerkung · Erkenntnisanspruch und Grenzen

Die hier integrierte Darstellung führt die philosophische Argumentation, die heuristische Modellbildung und die empirische Prüfplanung zu einem kohärenten Forschungsentwurf zusammen. Die Skalenvarianten dokumentieren einen weiteren Entwicklungsprozess.

Eine wissenschaftliche Bestätigung wird unabhängige Messungen, nachvollziehbare Güteprüfungen und einen methodisch belastbaren Gruppenvergleich erfordern.

Der Weisheitskompass kann Glaubwürdigkeit gewinnen, indem er die eigene Korrigierbarkeit erweitert.

DER WEISHEITSKOMPASS

Verantwortlich urteilen. Gemeinsam entscheiden. Zukunftsfähig handeln.

Ein wissenschaftlich fundierter, mathematisch modellierter und empirisch überprüfbarer Ansatz für bessere Entscheidungen unter Unsicherheit und in komplexen Zusammenhängen.

1 AUSGANGSLAGE

Warum wir einen Weisheitskompass brauchen

- **Begrenztes Wissen** — Unvollständige Informationen und unsichere Prognosen.
- **Wissensasymmetrien** — Unterschiedlicher Zugang zu Wissen, Macht und Perspektiven.
- **Komplexe Zusammenhänge** — Vielschichtige Ursachen und weitreichende Folgen.
- **Zielkonflikte** — Ökonomische, ökologische, soziale und ethische Spannungen.
- **Hohe Verantwortung** — Entscheidungen betreffen gegenwärtige und zukünftige Generationen.

2 DIE FÜNF SCHRITTE

Struktur verantwortlicher Urteilsbildung

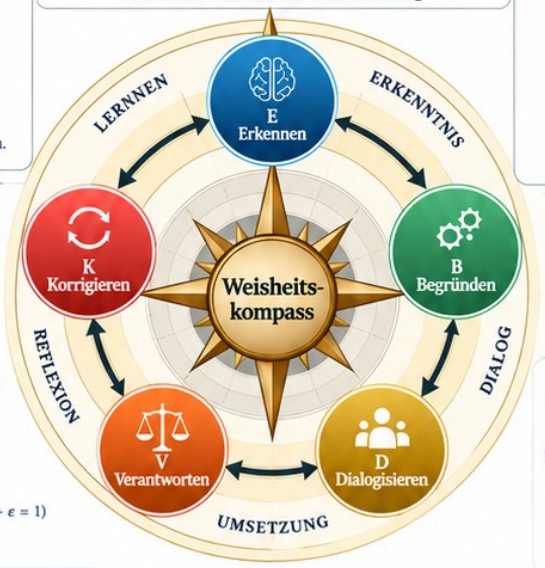
1 Unterscheiden	2 Begründen	3 Verbinden	4 Handeln	5 Überprüfen
Befund, Deutung und Wertung trennen	Schlussfolgerungen transparent machen	Perspektiven einbeziehen und dialogfähig bleiben	Verantwortbar entscheiden und umsetzen	Wirkungen evaluieren und korrigieren

Vom Problem zur lernenden Entscheidung

3 DIE FÜNF QUALITÄTSDIMENSIONEN

Was ein weises Urteil auszeichnet

- **E – Epistemische Prüffähigkeit** — Umgang mit Evidenz, Quellen und Unsicherheit
- **B – Begründungsqualität** — Logische und nachvollziehbare Argumentation
- **D – Dialogfähigkeit** — Einbezug unterschiedlicher Perspektiven
- **V – Verantwortbarkeit** — Berücksichtigung von Werten, Interessen und Folgen
- **K – Korrigierbarkeit** — Offenheit für Überprüfung und Revision



4 DAS HEURISTISCHE MATHEMATISCHE MODELL H5-WK

Die Qualität eines Urteils entsteht aus dem Zusammenspiel der fünf Dimensionen.

$$Q_t = \left(E_t^{\alpha} \cdot B_t^{\beta} \cdot D^{\gamma} \cdot V^{\delta} \cdot K_t^{\epsilon}\right)^{\frac{1}{5}}$$

Q_t	Qualitätsindex der Urteilsbildung (0–1)
E_t	epistemische Prüffähigkeit
B_t	Begründungsqualität
D_t	Dialogfähigkeit
V_t	Verantwortbarkeit
K_t	Korrigierbarkeit
$\alpha, \beta, \gamma, \delta, \epsilon$	Gewichtungsfaktoren ($\alpha + \beta + \gamma + \delta + \epsilon = 1$)

$E_t, B_t, D_t, V_t, K_t \in [0, 1]$

DYNAMISCHER LERNPROZESS

Der Weisheitskompass ist ein wiederkehrender Zyklus aus Urteilen, Handeln und Lernen.

$$S_{t+1} = \mathcal{L}(S_t, A_t, Y_{t+1}, F_t)$$

GEDANKLICHER ANSCHLUSS: WIRKUNGSORIENTIERTES DENKEN

Wirkungsorientiertes Denken fragt nach den Folgen unseres Handelns. Der Weisheitskompass betrachtet zusätzlich die Verlässlichkeit der Erkenntnisse und die verantwortliche Einordnung.

Vgl. Retzl, Helmut (2013): Wirkungsorientiertes Denken. Ein neuer Weg zur Gesundung. Goldegg Verlag.

5 DAS MESSINSTRUMENT H5-WK-18

Neun Indikatoren, 0–18 Punkte

Dimension	Indikator	Beschreibung	Punkte
E Epistemische Prüffähigkeit	E1	Befund, Deutung und Wertung unterscheiden	0-2
	E2	Quellen und Unsicherheit beurteilen	0-2
B Begründungs-qualität	B1	Schlussfolgerungen nachvollziehbar begründen	0-2
	B2	Alternativen und Gegenargumente prüfen	0-2
D Dialogfähigkeit	D1	Unterschiedliche Perspektiven berücksichtigen	0-2
	D2	Auf begründeten Widerspruch eingehen	0-2
V Verantwortbarkeit	V1	Werte, Interessen und Machtverhältnisse erkennen	0-2
	V2	Folgen und Betroffene berücksichtigen	0-2
K Korrigierbarkeit	K1	Bedingungen für Überprüfung und Revision festlegen	0-2
		Gesamtwert S	**0–18**

BEWERTUNG PRO INDIKATOR

- **0 — Nicht nachgewiesen** — Merkmal fehlt oder bleibt unklar.
- **1 — Ansatzweise nachgewiesen** — Merkmal erkennbar, aber unvollständig.
- **2 — Überzeugend nachgewiesen** — Merkmal konkret, sachgerecht und nachvollziehbar ausgeführt.

MATHEMATISCHE AUSWERTUNG

Gesamtwert einer Person i:

$$S_i = \sum_{j=1}^{9} x_{ij} \quad (0 \le S_i \le 18)$$

Dimensionswerte:

$$E_i = E_{i1} + E_{i2} \quad (0–4)$$
$$B_i = B_{i1} + B_{i2} \quad (0–4)$$
$$D_i = D_{i1} + D_{i2} \quad (0–4)$$
$$V_i = V_{i1} + V_{i2} \quad (0–4)$$
$$K_i = K_{i1} \quad (0–2)$$

$$S_i = E_i + B_i + D_i + V_i + K_i$$

Veränderung (Vorher–Nachher):

$$\Delta S_i = S_{i,\text{nach}} - S_{i,\text{vor}}$$

Gruppenvergleich (ANCOVA):

$$S_{\text{nach}} = \beta_0 + \beta_1 \cdot \text{Gruppe} + \beta_2 \cdot S_{\text{vor}} + \epsilon$$

6 STUDIENDESIGN

Randomisierte kontrollierte Untersuchung

Teilnehmende Personen
Ausgangsmessung (Vorher)

- **Gruppe A** — Weisheitskompass Anwendung der fünf Schritte
- **Gruppe B** — Aktive Vergleichsmethode Strukturierte Reflexion
- **Gruppe C** — **Kontrollgruppe** Unstrukturierte Bearbeitung

Nachher-Messung und Transfermessung
Anonymisierte Bewertung mit H5-WK-18 durch unabhängige Beurteilende

7 FALLZAHLPLANUNG

Für den Nachweis eines mittleren Effekts (d = 0,50)

Zwei unabhängige Gruppen (zweiseitiger Test)

$$n = 2 \cdot \frac{\left(z_{1-\alpha/2} + z_{1-\beta}\right)^2 \cdot \sigma^2}{\Delta^2}$$

Annahmen:

$\alpha = 0{,}05$ $(z_{1-\alpha/2} = 1{,}96)$
$1 - \beta = 0{,}80$ $(z_{1-\beta} = 0{,}84)$
$d = 0{,}50$ $\Rightarrow \Delta = d \cdot \sigma$
$\sigma = 3$ (Streuung der Gesamtpunkte)

Ergebnis:

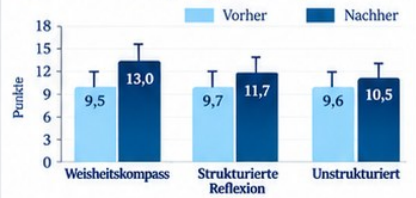
$n \approx 64$ Personen je Gruppe (auswertbar)
Bei 15 % Ausfällen: $n_{\text{rekrutierung}} \approx 76$ je Gruppe

Drei Gruppen (zwei primäre Vergleiche):

$n_{\text{gesamt}} \approx 276$ Personen (inkl. 15 % Ausfälle)

8 BEISPIELHAFTE ERGEBNISSE (HYPOTHETISCH)

Mittlere Urteilsqualität (0–18 Punkte)



Zusätzliche Verbesserung:

$\Delta_{WK-B} = (13{,}0 - 9{,}5) - (11{,}7 - 9{,}7) = 1{,}5$
$\Delta_{WK-C} = (13{,}0 - 9{,}5) - (10{,}5 - 9{,}6) = 2{,}6$

Hinweis: Dies sind hypothetische Werte zur Illustration. Ein empirischer Nachweis steht noch aus.

EIN ORIENTIERUNGSRAHMEN FÜR BESSERE ENTSCHEIDUNGEN – IN WISSENSCHAFT, POLITIK, WIRTSCHAFT UND GESELLSCHAFT.