„Artificial Sepsis“ – Leitlinien einer salutogenetischen Bot-Nutzung

Matthias O. Rath

Zusammenfassung

Der Beitrag beleuchtet die systemischen Risiken generativer KI-Systeme am Beispiel *Großer Sprachmodelle* wie ChatGPT. Er führt die Metapher *Artificial Sepsis* ein, um epistemische Selbstvergiftung durch rückgekoppelte KI-Inhalte zu beschreiben, und entwickelt ein Modell salutogenetischer Bot-Nutzung. Ethische Überlegungen zur viralen Verbreitung und zu den globalen Chancen und Herausforderungen generativer KI ergänzen die Beobachtung einer algorithmisch induzierten Selbstreferenz. Ziel des Beitrags ist eine paradigmatische Verschiebung der gängigen Bewertung des internetbasierten KI-Einsatzes hin zu epistemischer Resilienz, digitaler Kohärenz und verantwortungsgeleiteter Technikgestaltung.

1. Einleitung

Die rapide Ausbreitung generativer Künstlicher Intelligenz (KI) in Form Großer Sprachmodelle (*Large Language Models*, LLMs) wie ChatGPT transformiert gegenwärtig Kommunikationspraktiken, Informationsräume und das Verständnis von Autor:innenschaft. Basierend auf dem Konzept der „Transformer architecture” von Vaswani et al. (2017) wurden „neural networks“ mit riesigen Textkorpora darauf trainiert, jedes Wort oder Subwort (ein *token*, vgl. Vaswani et al. 2017) im Kontext aller anderen *tokens* zu verarbeiten. LLMs entwickeln dadurch zwar unerwartete Fähigkeiten, erzeugen aber auch Inkonsistenzen, die durch Korrekturen in den Trainingsdaten und im ‚Prompting‘, den Anweisungen (*prompts*) an die KI, optimiert werden müssen (vgl. Zhao et al. 2021). So zeigen Weidinger und andere (2021), dass Daten- und Prompting-abhängige Inkonsistenzen von LLMs bei ungeübter und unkontrollierter Nutzung gravierende methodische Schwächen mit medienethischen Folgen und erheblichen gesellschaftlichen Risiken generieren, etwa datengestützt die Verstärkung von Vorurteilen

und Stereotypen (*bias*), Desinformation und Vertrauensprobleme sowie in Bezug auf das Prompting die Problematik fehlender Transparenz (Opazität, vgl. Zhao et al. 2024).

Galten diese Inkonsistenzen und ihre Folgen noch bis vor kurzem als dystopische Vision, so sind diese inzwischen als Herausforderungen in alltägliche Schreib-, Forschungs- und Bildungsprozesse eingedrungen. Die öffentliche und wissenschaftliche Auseinandersetzung kreist dabei vorrangig um Fragen der Autor:innenschaft, der Täuschungspotenziale und der juristischen Verantwortbarkeit KI-generierter Texte (vgl. Dwivedi et al. 2023). Doch diese Fokussierungen greifen zu kurz. Denn das Risiko besteht nicht allein in Fälschungen oder Plagiaten, sondern in einer tiefgreifenden *epistemischen Transformation*, die sich aus selbstreferenziellen Rückkopplungen generativer Systeme speist. Welche Qualität haben Erkenntnisprozesse, an denen KI beteiligt ist? Welche Strukturen entstehen, wenn maschinell erzeugte Texte zum Ausgangspunkt weiterer maschineller Generierung werden? Dieser Beitrag greift bewusst den medizinischen Begriff der *Sepsis* metaphorisch auf, um auf diese Problematik aufmerksam zu machen.

Theoretisch wird der Beitrag von drei zentralen Perspektiven getragen:

1. Methodisch vom Konzept der *Futures Literacy* (vgl. Millers 2011; UNESCO 2018) als Reflexionsrahmen für antizipative Ethik,
2. inhaltlich von der Diskurs- und Kommunikationsethik von Jürgen Habermas als normativem Anspruch an KI-gestützte Kommunikation beziehungsweise KI-basierte Textgenerierung sowie
3. metapherntheoretisch von der *Conceptual Metaphor Theory* (vgl. Lakoff 1993) und der *Deliberate Metaphor Theory* (vgl. Steen 2023) als ethisch-epistemisches Gegenmodell zu dysfunktionalen Informationsökologien.

2. „Sepsis" als Metapher

Digitale Fehlfunktionen sind anhand *medizinischer* Metaphern besonders anschaulich beschreibbar. Ausdrücke wie ‚Virus' werden bereits seit den 1980er Jahren zur Beschreibung softwarebasierter Angriffe genutzt (vgl. Cohen 1987). Die hier eingeführte Metapher der ‚Artificial Sepsis' verweist jedoch auf eine qualitativ neue Störungsebene: die Selbstvergiftung eines informationsverarbeitenden Systems durch die zirkuläre Rezeption KI-generierter Inhalte.

Metaphern können als „*Sprachspiele*" (Wittgenstein 1999: 241) angesehen werden, die einer bestimmten abstrakten Denkstruktur (vgl. Bertau 1996)

entsprechen und Wörter eines „Ausgangsbereichs" mit ihrer Bedeutung in einen „Zielbereich" (Schmale 2019: 5) übertragen. Für die Philosophie und in der Folge die Theologie ist „Krankheit als Metapher" (Bendemann 2022: 88) schon seit der Antike konzeptionell geübt, um „Störungen" und ihre Überwindung strukturell nachvollziehbar zu machen.

Moderne Metapherntheorien (vgl. Semino/Demjén 2016), vor allem aus dem Bereich der kognitiven Linguistik (vgl. Geeraerts 2006; Putterer 2022), unterscheiden mehrere Metapherntypen, unter anderen „Strukturmetaphern". Lakoff und Johnson (2003: 22) bezeichnen damit in ihrer *Conceptual Metaphor Theory* (vgl. Lakoff 1993) jene „Fälle, in denen ein Konzept von einem anderen Konzept her metaphorisch strukturiert wird". Darüber hinaus gehend stellen die hier genutzten Metaphern jedoch nicht nur rhetorische Figuren oder konzeptionelle Übertragungen dar, sondern sie dienen im Sinne der *Deliberate Metaphor Theory* (vgl. Steen 2023) einem bewusst herbeigeführten Perspektivwechsel: „I propose that a metaphor is used deliberately when it is expressly meant to change the addressee's perspective on the referent or topic that is the target of the metaphor, by making the addressee look at it from a different conceptual domain or space, which functions as a conceptual source" (Steen 2008: 222).

Der medizinische Ausdruck „Sepsis" ist definiert als „life-threatening organ dysfunction resulting from a dysregulated host response to infection" (Singer et al. 2016), also als eine systemische Entgleisung der Immunantwort, bei der sich ein Organismus in der Reaktion auf eine Infektion selbst schädigt. Im Folgenden wird diese Fehlreaktion des Körpers, die das eigene Gewebe und die eigenen Organe bekämpft, als *fachsprachliche Metapher* (vgl. Schmale 2019) auf KI-generierte Kommunikation übertragen und als *Artificial Sepsis* (AS) spezifiziert. AS bezeichnet also eine Struktur, in der sich digitale Systeme gegen ihre eigenen Grundlagen richten, indem sie Inhalte nicht mehr aus menschlicher Erfahrung, dialogischer Interaktion oder sozial validierten Quellen schöpfen, sondern zunehmend aus eigenen Outputs – mit der Folge einer systemischen Selbstvergiftung.

3. Digitale Autointoxikation: Zur Diagnose der AS

In der Informatikgeschichte haben sich metaphorisierende Konzepte etabliert, um softwaretechnische oder netzwerkbezogene Fehlfunktionen zu beschreiben. Begriffsbildend wurde vor allem der Ausdruck *bug* als Be-

zeichnung des angeblich ersten Computer-Hardware-Fehlers durch eine Motte am 9. September 1947 in einem Relais eines *Harvard Mark II* Rechners (vgl. Lunduke 2022). Ab den 1980er Jahren sind vor allem metaphorische Ausdrücke wie „Virus" (Cohen 1987) oder „viral attack" (Ross 1990) geläufig. Gemeinsam ist diesen Metaphern die Vorstellung eines *externen* Eingriffs: Fehlfunktionen sind, ähnlich wie im bildspendenden Ausgangsbereich der hier verwendeten Metapher, der Medizin, von außen verursacht, durch externe Faktoren wie ‚Viren' oder ähnliches.

Die hier eingeführte Metapher ‚AS' will jedoch das Augenmerk auf den gegenteiligen Sachverhalt richten. Der problematische Faktor kommt nicht von außen, also als Eindringling, sondern von innen, nämlich durch die Präsenz KI-generierter Texte (im weitesten Sinne) im Netz. Nochmals mit dem fachsprachlichen Ausgangsbereich unserer Metapher AS, der Medizin, gesprochen: Der medizinische Normalfall ist der Einbruch des Fremden durch die Infektion, ein Keim, eine Mikrobe, ein Virus befällt einen Wirtsorganismus. Der medizinische Notfall hingegen tritt auf, wenn sich die Immunantwort des Körpers gegen sich selbst wendet und der Körper sich selbst vergiftet.

Ziel dieses Beitrags ist, die metaphorisch eingeführte *AS* als Fehlorientierung des digitalen Systems nach „innen" zu konstatieren und dann, wieder metaphorisch gesprochen, nach *Heilung* für diese Selbstvergiftung zu suchen. Denn die Gefahr digitaler Desinformation – so die These dieses Beitrags – liegt weniger in externen Aktoren, sondern in strukturell bedingten *Selbstreferenzdynamiken*. Diese Dynamiken lassen sich als Ausdruck einer ‚epistemischen Pathologie' verstehen – einer Störung des digitalen Erkenntnissystems, bei der sich Informationen zunehmend von ihren normativen und erfahrungsbezogenen Quellen lösen. Der metaphorisch aufgegriffene Begriff der Sepsis verweist in seinen „ko(n)textuellen und semantischen Informationen" (Schmale 2019: 9), die er definitorisch zur Verfügung stellt, auf eine *systemische Dysfunktion*: Unter den Bedingungen artifizieller Intelligenz, wie sie zum Beispiel ChatGPT zur Verfügung stellt, wird Wissen nicht mehr produziert, sondern rekombiniert; Urheberschaft wird entgrenzt statt verortet; und die Verifikation von Information wird ersetzt durch statistische Plausibilität (vgl. Rath 2018).

Bis Ende 2024 lernten LLMs wie GPT-3.5 und GPT-4 anhand riesiger Textmengen, darunter auch Daten, die maschinell erzeugt wurden. Bender und andere (2021) haben gezeigt, dass diese Modelle beginnen, ihre eigenen Outputs als Trainingsgrundlage zu nutzen – etwa durch Webcrawling öffentlich verfügbarer Inhalte, die selbst wieder von generativen Syste-

men stammen. Zusätzliche Dynamik erhält diese Entwicklung durch die *ChatGPT Search*-Funktion, die seit Dezember 2024 (OpenAI 2024) allen Usern zur Verfügung steht – das *ChatGPT 4 Model* und höher sucht nun auch im Internet. Dieser Rückkopplungseffekt bedroht die epistemische Diversität: Je mehr maschinell generierte Inhalte zirkulieren und rezykliert werden, desto stärker verengt sich der semantische Horizont, auf dem neue Inhalte beruhen. Shumailov und andere (2023: 2) beschreiben diesen Prozess daher als „model collapse", der zu einem algorithmisch induzierten „data poisoning" (Shumailov et al. 2023: 3) führt.

Konnte dieser Kollaps 2023 noch als nur drohend angesehen werden (vgl. Rath 2024), so zeigt der aktuelle *Bad Bot Report* (Imperva 2025), dass dieser Kollaps faktisch schon gegeben ist. Bereits 51 Prozent aller Webaktivitäten stammen von automatisierten Bots, die Inhalte generieren, die ihrerseits wiederum indexiert und von KI-Systemen als vermeintlich legitime Quellen genutzt werden. Besonders problematisch ist dies bei KI-Systemen, die für medizinische Beratung, psychologische Unterstützung oder juristische Einschätzungen eingesetzt werden – Bereiche, in denen semantische Präzision, Kontextsensitivität und ethische Verantwortung unabdingbar sind.

Damit entsteht durch den Rückgriff auf quasi unbegrenzt zur Verfügung stehende, artifiziell erzeugte Inhalte eine Art *ontologischer Kurzschluss*: Die LLMs generieren sprachliche Artefakte, die sich im nächsten Zyklus als ‚Realität' ausgeben – „models do not forget previously learned data, but rather start misinterpreting what they believe to be real, by reinforcing their own beliefs" (Shumailov et al. 2023: 3). In der Folge verschwimmen die Grenzen zwischen Simulation und Bezugnahme. Die Gefahr besteht dann nicht nur darin, Fehler zu übernehmen, sondern in der schleichenden Entwertung epistemischer, weil semantischer Standards. Was als „Content" zirkuliert, verliert seine Verankerung in realweltlicher Erfahrung, in sozialen Diskursen, in verantwortlicher Urheberschaft. AS beschreibt diesen Kollaps – eine Entkopplung digitaler Inhalte von ihren erkenntnisstiftenden Quellen. Das Konstrukt der AS fordert daher eine ethische Antwort, die über technische Korrekturmechanismen hinausweist: eine neue Kultur zukunftszugewandter *epistemischer Verantwortung*. Methodisch verweist dies auf die Erträge einer *Futures Literacy*.

4. Futures Literacy: Antizipieren, um zu verantworten

Das Konzept der *Futures Literacy* (vgl. Miller 2011; UNESCO 2018) wurde entwickelt, um die Fähigkeit zur aktiven Gestaltung der Zukunft durch kritische Reflexion zu fördern. Es geht nicht um deterministische Vorhersagen, sondern um die imaginative Erschließung möglicher, plausibler und wünschbarer Zukünfte. In Zeiten zunehmender Unsicherheit und technologischer Beschleunigung eröffnet *Futures Literacy* einen Reflexionsraum, der das ethische Potenzial antizipativer Urteilskraft betont.

Futures Literacy ist damit mehr als ein Planungstool. Sie ist eine epistemische Kompetenz – eine Kulturtechnik, mit der Gegenwart durch Zukunft verstehbar gemacht wird. Riel Miller unterscheidet in diesem Zusammenhang „‚used' futures" – implizite, unreflektierte Zukunftsbilder – und „open futures" – bewusste, plurale und veränderliche Imaginationen (vgl. UNESCO 2018: 54, 163). Der ethische Imperativ besteht darin, Zukunft nicht bloß zu konsumieren, sondern verantwortlich zu entwerfen.

Ein praktisches Beispiel ist der Aufbau sogenannter *sandboxes*, kontrollierte Umgebungen, in denen KI-Systeme unter Beobachtung getestet werden, bevor sie in reale Kontexte eingeführt werden (vgl. OECD 2023). Sie dienen nicht nur der technischen Optimierung und der prospektiven Abklärung von Haftungsfragen (Truby et al. 2022), sondern auch der ethischen Bewertung (vgl. Undheim et al. 2023): Wie reagiert ein KI-System auf Ambiguität? Welche Vorannahmen sind in die Trainingsdaten eingeschrieben? Wie können Diskriminierungen, Verzerrungen oder epistemische Ausschlüsse erkannt und adressiert werden?

Auch der Aufbau von Delphi-basierten Dialogforen in Bildung, Wissenschaft und Verwaltung kann als *futures*-literates Handeln verstanden werden. Diese Verfahren integrieren Expertenwissen, gesellschaftliche Erwartungen und normatives Urteilsvermögen in eine strukturierte Reflexion über die Zukunft. Dabei zeigt sich ein Spannungsverhältnis zwischen antizipativer Verantwortung und demokratischer Legitimität: Wer bestimmt, welche Zukunftsoptionen als wünschbar gelten? Welche Akteure sind sichtbar, welche marginalisiert? *Futures Literacy* muss deshalb mit dem Anspruch verbunden sein, partizipativ und dekolonial (vgl. Bourgeois et al. 2024) sowie reflexiv (vgl. Mangnus 2021) zu operieren.

In ethischer Perspektive lassen sich Parallelen zur Verantwortungsethik von Hans Jonas ziehen: Zukunft ist nicht nur Möglichkeitsraum, sondern auch Verpflichtungsraum. Die Prognose des Möglichen, vor allem des *worst case*, wird als „*Heuristik* der Furcht" (Jonas 1979: 392; vgl. Rath 1988) zur

Begründung gegenwärtiger Verantwortung. Ethische Orientierung muss in der digitalen Transformation antizipativ erfolgen – als heuristische Rahmung des noch nicht Wirklichen. *Futures Literacy* ist somit kein methodischer Luxus, sondern eine notwendige Voraussetzung dafür, dass KI-Entwicklung nicht blind, sondern verantwortet verläuft. Sie schafft jene semantische Tiefenschärfe, die notwendig ist, um AS zu vermeiden: durch präventive, plurale und reflektierte Zukunftsdeutungen.

5. Kommunikationsethik: Geltungsansprüche und epistemische Verantwortung

Jürgen Habermas formulierte in seiner *Theorie des kommunikativen Handelns* vier Geltungsansprüche, die in jedem Verständigungsakt implizit präsent sind – sie strukturieren die Möglichkeit legitimer Kommunikation (vgl. Habermas 1981: 439). Wir rekurrieren hier auf drei davon: *Wahrheit* – die Aussage soll inhaltlich zutreffen, *Wahrhaftigkeit* – die Sprechende sollen subjektiv ehrlich sein, und *Richtigkeit* – das Gesagte soll in einem sozialen Kontext als normativ akzeptabel gelten.

Im Kontext generativer KI stellen sich diese Geltungsansprüche neu: Kann ein KI-System „wahr" sprechen, ohne Erfahrung? Kann es „wahrhaftig" sein, ohne subjektives Bewusstsein? Kann es „richtig" kommunizieren, ohne normative Einbettung? Die Antwort lautet: nur simulativ. Generative Sprachmodelle wie ChatGPT produzieren Aussagen, die so wirken, als seien sie intentional, obwohl sie es nicht sind. Ihre Kommunikation ist performativ, aber nicht intentional.

Gunkel und Bryson (2014) schlagen vor, diese Differenz ernst zu nehmen. In ihrer Theorie der Maschinenmoral (*Machine Morality*) unterscheiden sie zwischen *moral agents* (Handlungsträgern mit Verantwortung) und *moral patients* (Wesen mit moralischem Anspruch auf Schutz). Generative KI-Systeme passen in keine dieser Kategorien vollständig. KI-Systeme erfordern eine neue Ethik relationaler Verantwortung: nicht, weil sie Subjekte sind, sondern weil Menschen mit ihnen interagieren, als wären sie es (vgl. für den Bildungsbereich Rath [im Druck]). Diese *as-if-Kommunikation* hat praktische Konsequenzen: Viele Nutzer:innen entwickeln eine anthropomorphe Beziehung zu KI-Systemen – sie erwarten Relevanz, Konsistenz, sogar Empathie. Diese Erwartungen treffen jedoch im Moment noch auf Systeme, die über keine Weltbindung verfügen. Daraus entsteht eine kommunikative Dissonanz, die Habermas' Modell nicht auflöst, aber normativ

rahmt: Wenn Geltungsansprüche nicht erfüllbar sind, müssen sie kenntlich gemacht werden.

Daraus ergibt sich ein *Design-Imperativ*, KI-Systeme so zu gestalten, dass ihre epistemische Statuslage transparent bleibt. Epistemische Label, die eine schlussfolgernd vermeintliche oder naheliegende Notwendigkeit markieren (vgl. Moon et al. 2016), aber auch Antwort-Quellenangaben, Unsicherheitsindikatoren oder stilistische Markierungen können dazu beitragen, die Differenz zwischen Mensch und Maschine kommunikationsethisch produktiv zu machen.

Habermas' Theorie bleibt somit normativer Maßstab in einer Welt, in der Kommunikation nicht mehr ausschließlich zwischen Menschen stattfindet, um kommunikative Rationalität auch unter Bedingungen algorithmischer Produktion aufrechtzuerhalten.

6. Informationsethik und digitale Kompetenz

Die ethischen Herausforderungen generativer KI lassen sich aber nicht allein durch technische Standards oder gesetzliche Regularien bewältigen. Mindestens ebenso entscheidend ist die Kompetenz der Nutzer:innen im Umgang mit digitalen Systemen. Die Diskussion um „Digital Literacy" (Buckingham 2015) hat sich in den vergangenen Jahren zum Konzept einer „Critical GenAI Literacy" (Rapanta et al. 2025) weiterentwickelt, das neben funktionalem Wissen auch ethische (vgl. Capurro 2010; Floridi 2013), reflexive und pädagogisch-didaktische Dimensionen umfasst.

Floridi (2013: 102–133) betont in diesem Zusammenhang die epistemische Verantwortung für die „infosphere" mediatisierter Gesellschaften (vgl. Kalina et al. 2018): Wo die Unterscheidung zwischen Information, Kommunikation und Handlung verschwimmt, tragen Menschen Verantwortung nicht nur für ihre Daten, sondern auch für ihre epistemischen Positionierungen. „Critical GenAI Literacy" wird damit zu einer Disziplin der Lebensführung, einer Frage der Urteilskraft in einem von Maschinen mitgestalteten Bedeutungsraum.

Beispiel Hochschullehre: Viele Studierende nutzen ChatGPT zur Ideenfindung, zum Formulieren wissenschaftlicher Texte oder zur Simulation von Prüfungsgesprächen. Lehrende stehen vor der Herausforderung, diese Nutzung nicht pauschal zu verbieten, sondern kompetenzorientiert zu begleiten – als Differenzierung zwischen zulässiger Unterstützung und un-

zulässiger Substitution, Förderung reflexiver Textarbeit und Stärkung der Fähigkeit zur Quellenkritik.

Beispiel Medienproduktion: Redaktionen stehen unter Produktionsdruck, KI-Systeme bieten eine scheinbar effiziente Lösung. Doch der Einsatz von automatisierten Textbausteinen, Überschriften oder ganzen Artikeln ohne menschliche Gegenprüfung führt zur Erosion journalistischer Qualitätsstandards. Informationsethik fordert hier institutionelle Sicherung epistemischer Sorgfalt – durch Redaktionsrichtlinien, Faktenchecks, Transparenzpflichten.

Beispiel Verwaltung: Wie geht man mit von KI generierten Verwaltungstexten um? Wer trägt Verantwortung für fehlerhafte Bescheide, diskriminierende Formulierungen oder unzulässige Schlussfolgerungen? Informationsethik bedeutet in diesem Fall: Einführung von Prüfungsschleifen, Dokumentation der Systemverwendung, Schulung des Verwaltungspersonals.

Auf einer übergreifenden Ebene kann man drei Kompetenzdimensionen unterscheiden:

- Technische Kompetenz: Wissen über Funktionsweise, Stärken und Schwächen generativer KI.
- Reflexive Kompetenz: Fähigkeit zur Kontextualisierung, zur Einschätzung von Risiken und zum kritischen Umgang mit KI-generierten Inhalten.
- Normative Kompetenz: Fähigkeit, Entscheidungen auf der Basis von Werten, Rechten und sozialen Folgen zu treffen.

Diese Kompetenzen bilden die Grundlage für *epistemische Resilienz*, die Fähigkeit, auch unter Bedingungen digitaler Ambiguität handlungsfähig zu bleiben. Sie ist das Gegenmodell zur AS: nicht immunisierend, sondern befähigend; nicht abschottend, sondern reflektierend. Und diese Resilienz ist notwendig. Waren 2023 starke KI-Eingriffe in die Infosphäre noch nur Befürchtungen – „ein fauler Apfel verdirbt das ganze Fass" war damals mein Bild (vgl. Rath 2024) –, so sind seit 2024 diese Befürchtungen Realität geworden: Der bereits erwähnte *Bad Bot Report* (Imperva 2025: 2) warnt, dass 37 Prozent des bot-basierten *Traffic* intransparent und schädlich ist, eben durch *bad bots* vollzogen wird. Konkret heißt das z. B., dass folgende Branchen am stärksten von bot-basierten Angriffen auf Accounts, sogenannte *takeovers*, betroffen waren: der Finanzdienstleistungssektor mit 22 Prozent aller Angriffe, gefolgt von Telekommunikation und Internetdienstanbietern mit 18 Prozent sowie Computer- und IT-Branche mit 17 Prozent (vgl. Imperva 2025: 19); und ein Ende ist nicht abzusehen. Das *Copenhagen*

Institute for Future Studies (CIFS 2023) prognostiziert, dass bis 2030 das *Metaverse*, also die Verkopplung von virtueller, erweiterter und physischer Realität, zum größten Teil, wohl über 99 Prozent (vgl. Hvitved 2022), KI-generiert sein wird. Wir werden unser altes Internet nicht wiedererkennen. Dieser Prognose des CIFS muss durch eine proaktive Fragestellung begegnet werden, ganz im Sinne der *Futures Literacy*: Was können und müssten wir tun, um die Erfüllung dieser Erwartung zu vermeiden?

7. Salutogenese als ethisches Gegenmodell

Als ein Gegenmodell zu einer resignativen Haltung oder gar Panik in Bezug auf die KI-Zukunft der Infosphäre wird im Folgenden auf den *salutogenetischen Ansatz* verwiesen. Dies Ansatz wurde ursprünglich von Aaron Antonovsky (1979, 1987) im medizinischen Kontext entwickelt und entspringt damit ebenfalls unserem oben eingeführten fachsprachlichen „Ausgangsbereich" (Schmale 2019: 5) der Metapher AS.

Der Ansatz der Salutogenese soll erklären, warum manche Menschen trotz widriger Lebensbedingungen gesund bleiben. Anstelle der Pathogenese – der Frage nach den Ursachen von Krankheit – fragt Antonovsky: Was hält Menschen gesund? Diese Perspektive lässt sich auf mediale und digitale Kontexte übertragen: Was hält uns *epistemisch* gesund in einer Welt, in der KI unsere Informationsökologie verändert (vgl. Ridder 2024)?

Antonovsky identifiziert drei Schlüsselfaktoren eines sogenannten Kohärenzsinns (*sense of coherence*), der maßgebend sei für die Gesundheit. Es ist ein „durchdringendes, andauerndes und dennoch dynamisches Gefühl des Vertrauens" (Antonovsky 1997: 36) in die grundsätzliche *Verstehbarkeit* (*comprehensibility*), *Handhabbarkeit* (*manageability*) und *Sinnhaftigkeit* (*meaningfulness*) des Lebens (vgl. Eriksson/Lindström 2005). Diese Dimensionen können als ethisch-epistemischer Kompass für den Umgang mit generativer KI fungieren:

- Verstehbarkeit: Digitale Inhalte müssen nachvollziehbar sein. Nutzer*innen können erkennen, wann sie mit KI-generierten Texten konfrontiert sind, wie diese zustande gekommen sind und welche Quellen zugrunde liegen. Transparente Kommunikation ist eine Grundvoraussetzung epistemischer Kohärenz.
- Handhabbarkeit: Der Umgang mit KI darf nicht zu Überforderung führen. Das bedeutet, Werkzeugen, Kriterien und Unterstützungsangeboten zur Bewertung digitaler Inhalte bereitzustellen. In der Bildung heißt das,

dass KI-gestützte Lernprozesse so gestaltet werden, dass Schüler:innen „agency" (Mick 2021) behalten – statt sich der Maschine zu unterwerfen.
– Sinnhaftigkeit: Kommunikation muss in einem Wertekontext verankert sein. Informationen erhalten nicht allein durch ihren Wahrheitsgehalt Relevanz, sondern durch ihre Bedeutung für das Leben, für Teilhabe, für Demokratie. KI-Systeme sollen so konzipiert sein, dass sie nicht bloß kognitive Entlastung, sondern existenzielle Orientierung ermöglichen.

Ein salutogenetisches Ethikmodell der KI-Nutzung bedeutet daher keine technikzentrierte Kontrolle, sondern eine lebensweltorientierte Ermöglichung. In der Hochschuldidaktik etwa könnten KI-Systeme eingesetzt werden, um explorative Lernpfade zu eröffnen – ohne die epistemische Eigenleistung zu entwerten. In der Sozialen Arbeit könnten Bots reflexive Dialoge anregen, um Handlungssicherheit zu fördern – nicht durch Belehrung, sondern durch Co-Konstruktion (vgl. Victor/Goldkind 2025).

Zugleich verweist der Salutogenese-Ansatz auf *epistemische Resilienz*: die Fähigkeit, Widersprüche auszuhalten, Unsicherheit zu navigieren und dennoch zu urteilen. Er steht damit quer zu Modellen, die auf Kontrolle, Verifikation oder Exklusion setzen. Statt AS durch technische Immunsysteme zu bekämpfen, schlägt AS eine andere Strategie vor: die Stärkung des Kohärenzempfindens in offenen, pluralen Diskursen.

Ein salutogenetischer Ausblick legt nahe, dass eine partizipativ gestärkte Medienpraxis die entscheidende Rolle im Umgang mit generativer KI spielen wird. Informationskompetenz und kohärente Kommunikationsräume, sogar Wahrhaftigkeit (Rath 2013) sind dabei keine Eigenschaften der Technologie, sondern Leistungen der Nutzer:innen selbst.

8. Fazit und Ausblick – Von der Sepsis zu Resilienz und Kohärenz

AS bezeichnet eine epistemische Pathologie, die nicht auf externe Desinformation oder böswillige Manipulation zurückzuführen ist, sondern auf eine systemische Rückkopplung interner Outputs in KI-generierten Systemen. Generative LLMs tendieren zu digitaler Selbstvergiftung, indem sie ihre eigenen Inhalte reproduzieren, *re-zirkulieren* und schließlich als vermeintlich autoritative Informationsquellen etablieren.

Demgegenüber plädiert der Beitrag für einen Paradigmenwechsel im Sinne einer digitalen ‚Salutogenese' – weg von reaktiver Abwehr hin zu proaktiver Ermöglichung. Die zentrale These lautet: *Nur durch eine ethische Rekonstruktion unserer Kommunikationsräume, Informationspraktiken*

und Bildungsprozesse lässt sich dieser Pathologie begegnen. Drei theoretische Achsen wurden dazu entfaltet:

1. Die *Futures Literacy* als antizipative Ethik, die mögliche Zukünfte plural reflektiert und partizipativ gestaltet,
2. die kommunikative Rationalität im Anschluss an Habermas, die als normatives Raster zur Beurteilung algorithmischer Äußerungen dient, und
3. die Salutogenese als orientierender Rahmen für eine epistemische Resilienz, die nicht auf Immunisierung, sondern auf Kohärenzbildung zielt.

Diese Triade konvergiert in einem von der Salutogenese inspirierten Begriff, der *digitalen Kohärenz*. Sie bezeichnet die Fähigkeit, auch unter Bedingungen automatisierter Kommunikation Orientierung, Urteilskraft und Anschlussfähigkeit zu bewahren. Digitale Kohärenz entsteht dort, wo Geltungsansprüche nicht verwischt, sondern expliziert werden; wo algorithmische Outputs nicht mystifiziert, sondern kontextualisiert werden; und wo technologische Mittel nicht bloß Effizienz, sondern Verstehbarkeit und Sinn ermöglichen.

In praktischer Hinsicht bedeutet dies, dass KI-Systeme so gestaltet, eingesetzt und bewertet werden müssen, dass sie die Handlungs- und Urteilskompetenz der Beteiligten stärken. Dazu gehören epistemische Labels, partizipative Prüfverfahren, offene Bildungsformate und plural strukturierte Diskursräume. Die Verantwortung für diese Gestaltung liegt nicht allein bei Entwickler:innen oder Nutzer:innen – sondern in einem geteilten, relationalen Raum zwischen Technik, Gesellschaft und Ethik. *Digitale Kohärenz* heißt nicht Anpassung an technische Überforderung, sondern Gestaltung alternativer Möglichkeitsräume. *AS* ist kein Schicksal. Sie ist eine Herausforderung – für unser Wissen, unser Handeln und unsere Vorstellungskraft.

Literatur

Antonovsky, Aaron (1979): Health, stress and coping, San Francisco.

Antonovsky, Aaron (1987): Unraveling the mystery of health: How people manage stress and stay well, San Francisco.

Antonovsky, Aaron (1997): Salutogenese – Zur Entmystifizierung der Gesundheit, Tübingen.

Bendemann, Reinhard von (2022): Christus der Arzt. Frühchristliche Soteriologie und Anthropologie im Lichte antik-medizinischer Konzepte, Stuttgart.

Bender, Emily M. et al. (2021): On the dangers of stochastic parrots: Can language models be too big?, in: Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency, New York, S. 610–623.

Bertau, Marie-Cécile (1996): Sprachspiel Metapher: Denkweisen und kommunikative Funktion einer rhetorischen Figur, Wiesbaden.

Bourgeois, Robin / Karuri-Sebina, Geci / Feukeu, Kwamou Eva (2024): The future as a public good: decolonising the future through anticipatory participatory action research, in: Foresight 26 (4/2024), S. 533–549. https://doi.org/10.1108/FS-11-2021-0225

Buckingham, David (2015): Defining digital literacy – what do young people need to know about digital media?, in: Nordic Journal of Digital Literacy 10 (4/2015), S. 21–35.

Capurro, Rafael (2010): Informationsethik. Ein interdisziplinärer Ansatz, in: Information – Wissenschaft & Praxis 61 (6/2010), S. 315–320.

CIFS (2023): Metaverse Delphi Study. A delphi study on the development of the metaverse towards 2030. Copenhagen Institute for Futures Studies (CIFS), Kopenhagen (online unter: https://veri-media.io/wp-content/uploads/2023/03/metaverse-delphi-study.pdf – letzter Zugriff: 25.6.2025).

Cohen, Fred (1987): Computer viruses: theory and experiments, in: Computers & Security 6 (1/1987), S. 22–35.

Dwivedi, Yogesh K. et al. (2023): So what if ChatGPT wrote it? Multidisciplinary perspectives on opportunities, challenges and implications of generative conversational ai for research, practice and policy, in: International Journal of Information Management 71 (August 2023), 102642.

Eriksson, Monica / Lindström, Bengt (2005): Validity of Antonovsky's sense of coherence scale: a systematic review Journal of Epidemiology & Community Health 59 (6/2005), S. 460–466.

Floridi, Luciano (2013): The ethics of information, Oxford.

Geeraerts, Dirk (2006): Introduction. A rough guide to cognitive linguistics, in: Dirk Geeraerts (Hg.), Cognitive Linguistics: Basic readings, Berlin, New York, S. 1–28.

Gunkel, David J. / Bryson, Joanna (2014): Introduction to the special issue on machine morality: The machine as moral agent and patient, in: Philosophy & Technology 27 (März 2014), S. 5–8.

Habermas, Jürgen (1981): Theorie des kommunikativen Handelns. Band 1. Frankfurt am Main.

Hvitved, Sofie (2022): What if 99 % of the Metaverse is made by AI? 24.2.2022 (online unter: https://cifs.dk/news/what-if-99-of-the-metaverse-is-made-by-ai – letzter Zugriff: 25.6.2025).

Imperva (2025): 2025 Bad Bot Report. The Rapid rise of bots and the unseen risk for business (online unter: https://www.imperva.com/resources/resource-library/reports/2025-bad-bot-report/ – letzter Zugriff: 26.9.2025).

Jonas, Hans (1979): Das Prinzip Verantwortung. Versuch einer Ethik für die technologische Zivilisation, Frankfurt am Main.

Kalina, Andreas et al. (Hg.) (2018): Mediatisierte Gesellschaften. Medienkommunikation und Sozialwelten im Wandel (Tutzinger Studien zur Politik), Baden-Baden.

Lakoff George (1993): The contemporary theory of metaphor, in: Andrew Ortony (Hg.), Metaphor and Thought. 2. Aufl., Cambridge, S. 202–251.

Lakoff, George / Johnson, Mark (2003): Leben in Metaphern. Konstruktion und Gebrauch von Sprachbildern. 3. Aufl., Heidelberg.

Lunduke, Bryan (2022): The story of the first 'computer bug'... is a pile of lies, in: The Lunduke Journal of Technology, 19.8.2022 (online unter: https://lunduke.substack.com/p/the-story-of-the-first-computer-bug – letzter Zugriff: 25.6.2025).

Mangnus, Astrid C. et al. (2021): Futures literacy and the diversity of the future, in: Futures 132 (September 2021), article 102793. https://doi.org/10.1016/j.futures.2021.102793

Mick, Carola (2021). Das Agency-Paradigma, in: Ullrich Bauer / Uwe H. Bittlingmayer / Albert Scherr (Hg.), Handbuch Bildungs- und Erziehungssoziologie, Wiesbaden, S. 1–15.

Miller, Riel (2011): Futures literacy — embracing complexity and using the future, Ethos 10 (October 2011), S. 23–28.

Moon, Lori / Kirvaitis, Patricija / Madden, Noreen (2016): Selective annotation of modal readings: Delving into the difficult data, in: Linguistic Issues in Language Technology 14 (6/2016) (online unter: https://aclanthology.org/2016.lilt-14.6 – letzter Zugriff: 25.6.2025).

OECD (2023): Regulatory sandboxes in artificial intelligence. OECD digital economy Papers, July 2023, No. 356 (online unter: https://www.oecd.org/content/dam/oecd/en/publications/reports/2023/07/regulatory-sandboxes-in-artificial-intelligence_a44aae4f/8f80a0e6-en.pdf – letzter Zugriff: 25.6.2025).

OpenAI (2024): Introducing ChatGPT search, 16. Dezember 2024 (online unter: https://openai.com/index/introducing-chatgpt-search/ – letzter Zugriff: 9.7.2025).

Putterer, Elisabeth (2022): Von der Conceptual Metaphor Theory zur Deliberate Metaphor Theory: Theoretische Annahmen, Kritikpunkte und Klärungsversuche, in: Initium 4 (1/2022), S. 112–127. https://doi.org/10.33934/initium.2022.4.9

Rapanta, Chrysi et al. (2025): Critical GenAI literacy: Postdigital configurations, in: Postdigital Science and Education (2. Juli 2025). https://doi.org/10.1007/s42438-025-00573-w

Rath, Matthias (1988): Intuition und Modell. Hans Jonas' „Prinzip Verantwortung" und die Frage nach einer Ethik für das wissenschaftliche Zeitalter, Frankfurt am Main.

Rath, Matthias (2013): Authentizität als Eigensein und Konstruktion – Überlegungen zur Wahrhaftigkeit in der computervermittelten Kommunikation, in: Martin Emmer / Alexander Filipovic / Jan-Hinrik Schmidt / Ingrid Stapf (Hg.), Echtheit, Wahrheit, Ehrlichkeit. Authentizität in der Online-Kommunikation, Weinheim, S. 16–27.

Rath, Matthias (2018): Data Science – die neue Leitwissenschaft?, in: Thomas Knubben / Erich Schöls / Uli Braun (Hg.), Weltkulturatlas – Kultur in Zeit der Globalisierung. Daten, Geschichten, Grafiken, Stuttgart, S. 21–37.

Rath, Matthias (2024): 'To find the 'rotten apple' – information ethical requirements for the information literacy of autonomous writing engines, in: Serap Kurbanoğlu / Sonja Špiranec / Joumana Boustany / Yurdagül Ünal / İpek Şencan / Denis Kos / Esther Grassian / Diane Mizrachi / Loriene Roy (Hg.), Information Experience and Information Literacy. 8th European Conference on Information Literacy, ECIL 2023, Revised Selected Papers, Part II. Cham, S. 129–139.

Rath, Matthias [im Druck]: Peer-to-Peer Learning mit „Artificial Companions". Hochschuldidaktische Überlegungen zur mediatisierten Lehre in digitalen Kontexten, in: Carolyn Blume / Gudrun Marci-Boehncke / Patricia Ronan (Hg.), Peer-to-Peer-Konzepte in der Sprachendidaktik. Theorie und Praxis für die Hochschullehre, Bielefeld.

Ridder, Jeroen de (2024): online illusions of understanding, in: Social Epistemology 36 (6/2024), S. 727–742.

Ross, Andrew (1990): hacking away at the counterculture, in: Postmodern Culture 1 (1/1990). https://dx.doi.org/10.1353/pmc.1990.0011

Schmale, Günter (2019): Mögliche Metaphern in der Fachsprache, in: ELAD-SILDA Études de Linguistique et d'Analyse des Discours – Studies in Linguistics and Discourse Analysis 2 (Oktober 2019), S. 1–32.

Semino, Elena / Demjén, Zsófia (Hg.) (2016): The Routledge handbook of metaphor and language. London. https://doi.org/10.4324/9781315672953

Shumailov, Ilia et al. (2023): The curse of recursion: Training on generated data makes models forget, in: arXiv:2305.17493. https://doi.org/10.48550/arXiv.2305.17493

Singer, Mervyn et al. (2016): The third international consensus definitions for sepsis and septic shock (Sepsis-3), in: JAMA 315 (8/2016), S. 801–810.

Steen, Gerard J. (2008): The paradox of metaphor: Why we need a three-dimensional model of metaphor, in: Metaphor and Symbol 23 (4/2008), S. 213–241. https://doi.org/10.1080/10926480802426753

Steen, Gerard J. (2023:) Thinking by metaphor, fast and slow: Deliberate Metaphor Theory offers a new model for metaphor and its comprehension, in: Frontiers in Psychology 14 (5. September 2023), 1242888. https://doi.org/10.3389/fpsyg.2023.1242888

Truby, Jon et al. (2022): A sandbox approach to regulating high-risk artificial intelligence applications, in: European Journal of Risk Regulation 13 (2/2022), S. 270–294.

Undheim, Kristin / Erikson, Truls / Timmermans, Bram (2023): True uncertainty and ethical AI: regulatory sandboxes as a policy tool for moral imagination, in: AI Ethics 3 (August 2023), S. 997–1002.

UNESCO (2018): Transforming the future: anticipation in the 21st century. London (online unter: https://unesdoc.unesco.org/ark:/48223/pf0000264644 – letzter Zugriff: 25.6.2025).

Vaswani, Ashish et al. (2017): Attention is all you need, in: arXiv: 1706.03762. https://doi.org/10.48550/arXiv.1706.03762

Victor, Bryan G. / Goldkind, Lauri (2025): The therapist in the machine: Confronting AI's Challenge to clinical social work, in: Journal of Technology in Human Services 43 (2/2025), S73–81. https://doi.org/10.1080/15228835.2025.2500827

Weidinger, Laura et al. (2021): Ethical and social risks of harm from Language Models, in: arXiv:2112.04359v1. https://doi.org/10.48550/arXiv.2112.04359

Wittgenstein, Ludwig (1999): Philosophische Untersuchungen, in: Gertrude Elizabeth Margaret Anscombe / Rush Rhees /Georg Henrik von Wright (Hg.), Ludwig Wittgenstein Werkausgabe. Band 1, Frankfurt am Main, S. 231–485.

Zhao, Haiyan et al. (2024): Explainability for large language models: A survey, in: ACM Transactions on Intelligent Systems and Technology 15 (2/2024), Article 20. https://doi.org/10.1145/3639372

Zhao, Zihao et al. (2021): Calibrate before use: Improving few-shot performance of language models, in: Proceedings of Machine Learning Research 139, 12697–12706 (online unter: https://proceedings.mlr.press/v139/zhao21c.html – letzter Zugriff: 25.6.2025).