# Zwischen AI Act und Posthumanismus. Künstliche Intelligenz und ihre ethischen „Radialkräfte“

**Michael Funk**

*Zusammenfassung*

Ethik feiert Konjunktur. Neuerdings dreht sich alles um Künstliche Intelligenz. Die alte Frage „Ethik – wofür braucht man das?“ ist auf dem absteigenden Ast. Längst hat der Run um die Deutungshoheit bei der Regulierung von KI begonnen. Ethik ist Richtlinie, politisches Programm, wirtschaftliches Interesse und/oder eine humanistische Gutenachtgeschichte, die es in der Morgendämmerung intelligenter Maschinen zu überwinden gilt. In vorliegendem Beitrag wird die Metapher der „Radialkraft“ verwendet, um vier ausgewählte Spannungsfelder aufzuzeigen. Was macht Ethik aus? Welche Wissensdiversität hält Ethik in Zeiten Künstlicher Intelligenz im Gleichgewicht? Diese Fragen diskutiere ich anhand von zwei Beispielen: dem EU AI Act sowie der posthumanistischen Kritik an menschenzentrierter Ethik.

## Einleitung: Von Ethiken zur Wissensdiversität

Räumen wir gleich zu Beginn zweierlei ein: Erstens, „Künstliche Intelligenz“ (KI) ist eine schiefe Übersetzung von „artificial intelligence“ und sollte eigentlich „technische Informationsverarbeitung“ meinen. Als solche umfasst sie eine Vielzahl verschiedener Produkte und Verfahren – etwa des *machine learning* als Mittel zum Zweck der Mustererkennung in Daten und anschließender Prognose realer Ereignisse. KI ist Querschnitttechnologie in vielen Lebenslagen. In vorliegendem Beitrag soll es nicht um die technischen Details gehen, sondern um die ethischen Radialkräfte durch deren Einsatz. Dabei werde ich der eingebürgerten Redeweise von KI folgen (vgl. auch die Beiträge in diesem Band).

Theo Hug, Petra Missomelius, Heike Ortner (Hg.):
Künstliche Intelligenz im Diskurs: Interdisziplinäre Perspektiven zur Gegenwart und Zukunft von KI-Anwendungen
© 2024 innsbruck university press, ISBN 978-3-99106-139-7, DOI 10.15203/99106-139-7

Zweitens: So wie es nicht *die* KI gibt, so gibt es auch nicht *die* Ethik. Hier treffen sich zwei Metaphern. Zum einen KI als Bild für Technik, die wie Menschen oder Tiere „intelligent“ sein könnte. Zum anderen klingt der Titel des vorliegenden Beitrages doch arg nach Computersimulation von Maschinenbauteilen, also nach dem Bild der „Radialkräfte“ die im Verbund mit „Fliehkräften“ auf einen rotierenden Körper wirken. Was da rotiert, nennen wir Ethik. Vermutlich entspricht das Bild der Fliehkräfte eher der Erwartung: KI erzeugt Fragen und Probleme, welche die Ethik aus ihrem altehrwürdigen Zentrum katapultieren – z.B. moralische Maschinen oder künstliches Bewusstsein als Grundlage für Robotergesetze. Man kann das aber auch herumdrehen: Vielleicht sind es ja doch Radialkräfte, die Ethik in Zentrumnähe halten – z.B. durch Regulierung und Kodizes oder die Entmystifizierung technologischer Verheißungen (auch KI ist „nur“ technische Informationsverarbeitung). Dabei kann Ethik in dreifacher Hinsicht auf bekannten Unterscheidungen aufbauen[1]:

1. **Moral/Sitte** im Sinne gewohnter Handlungen, die sich stark zwischen menschlichen Gruppen unterscheiden können, einschließlich taubem Fanatismus oder der Sitte, die Möglichkeit von Moral überhaupt zu leugnen, bis hin zu Fragen nach den Grenzen und der Verbindlichkeit individueller oder gemeinschaftlicher sozialer Orientierungen (z.B. Hoerster 2022; Somek 2021; Williams 1986)

2. **Ethik**
a) als rationale **Moralwissenschaft**, Suche nach der Begründung des in sich Guten – im Gegensatz zum bloß zweckdienlich Nützlichen (z.B. Horster 2009; Murdoch 2023; Tugendhat 1984) – bzw. des spezifisch menschlich Guten und dessen Natur (z.B. Fischer 2022; Foot 2023; Kallhoff 2022), sowie über das Konzept einer modernen Wissenschaft hinaus:
b) als **angewandte Ethik und ethische Praxis** jenseits theoretischer Begründungsdebatten oder formaler Metaethik (Johnson/Toulmin 1988; Schwemmer 1980; Thurnherr 2010)
c) und noch stärker als **Lebenskunst/Klugheitslehre** des glücklichen individuellen und gemeinsamen Lebens, wie schon in der Antike oder außereuropäischen Traditionen (Hossenfelder (Hrsg.) 2013; Masek 2023; Nussbaum 2020)

3. **Kodizes/Ethikrichtlinien**, in denen Handlungsnormen explizit ausformuliert sind – wie im hippokratischen Eid oder den Zehn Geboten – im Unter-

1 Zur vertieften Analyse mit Blick auf die Roboter- und KI-Ethik siehe Funk (2022a).

schied zu moralischem Handlungswissen, das auch implizit durch Bräuche und Rituale ausgedrückt wird.

Die Frage des vorliegenden Beitrags lässt sich entsprechend zusammenfassen:

> *Wie verändert der Umgang mit „technischer Informationsverarbeitung" die Ethik im dreifachen Sinne: hinsichtlich moralischer Gewohnheiten, moralwissenschaftlicher Reflexion (einschließlich Lebenskunst sowie ethischer Praxis) und Ethikrichtlinien?*

Ich möchte diese Frage zuerst mit einer Übersicht zu vier „Radialkräften" bearbeiten und anschließend anhand von zwei Beispielen vertiefen: dem AI Act sowie der kritischen Diskussion über die Menschenzentriertheit der Ethik. Dabei schließe ich an meinen Beitrag *Nichtwissen ist Macht! Künstliche Intelligenz und ihre diversen Ethiken* (vgl. Funk 2024) unmittelbar an, wo zwei Formen der Wissensdiversität unterschieden wurden:

**Wissensdiversität I:** Vielfalt moralischer Lebensstile (oben 1.) und ethischer Lebenskünste (2.c)

- bildet den Fundus personalen Wissens und zwischenmenschlicher Fertigkeiten: erlernt durch Imitation, Tradition, zwischenmenschliche Interaktionen, Trial and Error, pragmatische Bewährung durch wiederholt gelingenden Umgang, Riten, Sitten, Bräuche, Intuition, Moral Sense, Umgehen mit Emotionen, leiblichen Impulsen etc.
- Kritische Infrastruktur freiheitlich-demokratischer Gesellschaften
- wird auch in der Moralphilosophie, Metaethik und Moralbegründung stillschweigend vorausgesetzt
- Prinzipielle Grenze der KI: Maschinen können das (höchstens teilweise) simulieren, aber nicht authentisch leben

**Wissensdiversität II:** Vielfalt der Wissensformen, die in der Ethik (2.) wirksam werden

- s.o. Wissensdiversität I (denn Ethik wird von lebenden Menschen gemacht)
- Wissensprozesse, epistemische Praxen: Verfahrenswissen der angewandten Ethik (Kasuistik, Einzelfallanalysen, Case Studies etc.), Problemlösungsstrategien ethischer Praxis (logisch-ableitend, analogisch-vergleichend, top-down, bottom-up etc.) (2.b)

- propositionales, formales Wissen der Metaethik: Sprachanalyse ethischer Argumente und moralischer Urteile, formale Logik ethischer Begründungen (2.a)
- Kenntnis der Positionen der Moralbegründung: deontologische Ethik, Utilitarismus, Diskursethik, Tugendethik etc. (2.a)
- propositionales Wissen ethischer Theorien: Konzepte, Werke, AutorInnen auswendig lernen (2.a)
- Bsp.: Umgang mit Fehlschlüssen (z.B. Sein-Sollen-Fehlschluss oder humanitaristischer Fehlschluss)
    - auswendig kennen (propositionales Wissen)
    - moralphilosophisch korrekt in Texten oder Argumenten erkennen, aufdecken und korrigieren können
    - im moralischen Alltag wahrnehmen und vermeiden (oder bei nicht erheblicher Problematik gelassen hinnehmen) können

Die These, an die in vorliegendem Beitrag angeschlossen wird, lautet:

> *Ethik Künstlicher Intelligenz ist für leiblich-sinnliche Menschen, für emotionale Wesen im kommunikativen Gemeinschaftsleben, nicht jedoch für mathematisch reduzierte Massenpunkte oder Vektoren in Computermodellen.*

Diesen Unterschied zu erkennen, setzt eine Haltung voraus, die in den aktuellen Debatten um KI-Ethik nicht selbstverständlich ist.[2] Eine wesentliche Ursache ist die Vereinnahmung kritischer Ethik durch industrielle Interessen im so bezeichneten Überwachungskapitalismus. Insofern finden wir hier eine der stärksten Radialkräfte auf die aktuelle Ethik. Zu den Vorurteilen im Umgang mit KI gehört, die Diversität ethischen und moralischen Wissens zu unterlaufen. Das passiert z.B., wenn Ethik auf Guidelines reduziert wird, also „Checklistenökonomie“, wonach moralische Verantwortung im Moment des Abhakens von Kästchen zur Fairness und Vertrauenswürdigkeit erledigt ist (erstes Beispiel). Der „Export“ europäischer Ethikrichtlinien ist nicht wertfrei und könnte ungewollt zur Wiederholung kolonialer Handlungsmuster führen. Dazu gehört das Ausblenden moralischen Handlungswissens indigener Kulturen zugunsten bestimmter Menschen- und Weltbilder. Vielleicht ist die Menschenzentriertheit bestimmter anthropozentrischer Ethiken ein Gift für holistische Alternativen (zweites Beispiel)? In einer Art Prisma sollen **kulturell-soziale Radialkräfte der Ethik** aufgezeigt werden, so wie sie vom Umgang mit KI ausgehen. Dabei muss zur oben genannten Forschungs-

2 Siehe zur kritischen Diskussion z.B.: Coeckelbergh 2022; Dignum 2019 oder Eubanks 2018. Ich schließe auch an meine Analyse in Funk 2023 an.

frage noch ergänzt werden, dass sie natürlich ein Framing vorgibt. Sie ließe sich auch herumdrehen: Welchen Einfluss hat der Umgang mit Ethik (in ihren drei Bedeutungen) auf KI? Beide Perspektiven gehören zwar zusammen. Jedoch gebe ich der Frage nach dem Einfluss von KI auf Ethik hier den Vorzug, da sie im Anbetracht wirtschaftlich-technologischer Tatsachen und hinterherhinkender Debatte über das, was wir wollen, die drängendere zu sein scheint.

## Radialkräfte der Ethik

Sehen wir uns einige der Radialkräfte an. Bildlich schauen wir auf die Kufen einer Eiskunstläuferin und die wirkende Energie. Welche Kraft hält sie in der Spur, wenn sie sich virtuos in die Kurven legt? Ganz ähnlich ist es – metaphorisch – mit der Ethik. Und alle hoffen, dass sie nicht stolpert und sich erst recht nicht dabei verletzt. Ich möchte vier dieser Spannungsfelder herausheben – sicherlich lassen sich noch mehr finden. Sie beschreiben eine allgemeine Verfasstheit und werden hier mit Blick auf KI wiedergegeben.

I. Unter dem Titel **„Nehmt der Industrie die Ethik weg!“** (Metzinger 2019) lässt sich eine erste Radialkraft zusammenfassen. Das Problem liegt darin, dass Ethik durch den industriellen Umgang mit KI völlig aus dem Gleichgewicht gerissen wird. Damit sind weniger neue Fragen im Anbetracht neuer technologischer Möglichkeiten gemeint. Denn diese geben ja der ethischen Diskussion Schwung. Auch sollten Beiträge zur ethischen Bewertung aus der technischen, industriellen Praxis heraus ausdrücklich erwünscht sein! Es wäre doch vermessen, Ethik zum elitären akademischen Sport zu machen, jenseits des Alltagsverstandes. Problematisch ist jedoch die Umnutzung von Ethik zur Interessenvertretung. Es wirken also zwei Pole aufeinander ein:

1. Fliehkraft: industrielle oder politische Lobbyarbeit, „ethics washing“ zur Akzeptanzbeschaffung neuer KI, Ausspielen von Ethik gegen politische Regulierung, Unternehmen mit marktbeherrschender Stellung nutzen Regulierung (juristische und/oder ethische), um MitbewerberInnen klein zu halten

⇨ Umnutzen von Ethik, Ethik als zweckdienliches Mittel zur Interessenvertretung

2. Gegenpol: ethischer Elfenbeinturm

⇨ **Radialkraft:** Balanceakt der KI-Ethik zwischen bodenständiger Praxisrelevanz sowie autonomer Suche nach in sich Gutem und Schutz vor Missbrauch

- ⇨ Angewandte Ethik (2.b): Vermittlung ethischen Fachwissens und alltäglichen Handlungswissens

Eine Lösungsstrategie besteht in wirksamer Wissensdiversität durch authentische transdisziplinäre Integration: industrielles, technisches Wissen im Dialog mit politischem Wissen und wissenschaftlichem, ethischem Fachwissen. Die Diversität ethischen Wissens, von gelebter Moral (1.) bis hin zur rationalen Argumentation (2.) sollte nicht durch Lobbyismus entwertet werden.

II. Die zweite Radialkraft ist verbunden mit **„Checklistenethik“** (Nowotny 2023, S. 214). Dabei geht es weniger um den Missbrauch ethischer Etiketten zur Akzeptanzbeschaffung, sondern mehr um die Reduktion ethischen Wissens auf Kodizes (3.):

1. Fliehkraft: Reduktion von Ethik auf KI-Guidelines, die top-down vorgegeben werden und dann „nur“ noch abgehakt werden müssen (Regulierung von oben)
- ⇨ Reduktion von Ethik auf eine spezifische Wissensform
2. Gegenpol: bottom-up Reduktion von Ethik auf das je individuelle Befinden
- ⇨ **Radialkraft:** Balanceakt der KI-Ethik zwischen gemeinschaftlich verbindlichen Guidelines (3.) und konkreter Lebensführung (1., 2.c): auch Guidelines wollen von Menschen gelebt sein
- ⇨ Angewandte Ethik: Vermittlung von universellen Richtlinien und spezifischen Alltagsfällen (Case Studies, Kasuistik)

Auch hier ergeben sich Lösungsstrategien durch transdisziplinäre Integration. Vielleicht ist das offensichtlicher, wenn von technischen oder geisteswissenschaftlichen Fächern die Rede ist. Jedoch gibt es auch subtilere Formen der Integration, konkret von Wissensdiversität II (s.o.): propositionales Wissen (explizite Formulierungen von Guidelines) in Verbindung mit moralischem Alltagswissen, ethisch-reflexivem Fachwissen sowie Ethik als Lebenskunst. Transdisziplinär ist eine solche Integration weniger wegen der Berufsabschlüsse der beteiligten Menschen, sondern wegen der verschiedenen Erfahrungshorizonte.

III. Bleibt noch der **„Elfenbeinturm“** philosophischen, weltabgewandten ExpertInnentums, der eine eigene Hybris ausprägt. Da ist zum einen die weltabgewandte Träumerei oder Unverständlichkeit verkopfter Glasperlenspiele, die zuweilen Hand in Hand geht mit Angst vor dem wahren, sinnlich-körperlichen Leben. So folgt entweder die Furcht vor überlegenen Maschinen oder die schlichte Belanglosigkeit, wenn es um das Lösen

realer Herausforderungen geht. Zum anderen wirkt die philosophische Selbstbeschränkung fatal, wenn sie im formal-geistigen Kunstraum nur noch das als Ethik übriglässt, was durch KI besonders gut ersetzt werden kann – z.B. metaethische Theorien:

1. Fliehkraft: ethisches ExpertInnentum in einer eigenen Bubble über den Köpfen anderer gesellschaftlicher Gruppen und/oder skeptisches Alleszerzweifeln
⇨ Angst vor KI und dem Gespenst überlegener Maschinen
2. Gegenpol: Entwissenschaftlichung von Ethik, unreflektiertes Aufeinanderprallen moralischer Vorstellungen und/oder blinder, sturer Dogmatismus bis hin zur KI-Ideologie, Transhumanismus als Mythologie und Weltbild, Glaube an Superintelligenz
⇨ **Radialkraft:** selbstkritische Reflexion auf die Chancen und Grenzen wissenschaftlichen Wissens und Ethik als fluider, lebenslanger, zwischenmenschlicher Lernweg (Wissen als Klugheit (2.c)) gegen den Kokon des Formalismus oder dogmatische Elitendünkel
⇨ Angewandte Ethik: skeptisches Prüfen und authentische Haltung im Umgang mit KI verbinden

Transdisziplinäre Integration kann sich dann sogar als überlebenswichtig für die akademische, philosophische Ethik herausstellen. Schafft sie es, sich konstruktiv in den Debatten zum Umgang mit KI einzubringen, oder reduziert sie sich selbst auf einen ethischen Automaten, der früher oder später von einer entsprechend leistungsfähigen Maschine – völlig zu Recht – ersetzt wird?

IV. Nicht weit weg vom Elfenbeinturm ist das vierte Spannungsfeld der **„narzisstischen Bettgeschichten“** (Barad 2020, S. 97). Es geht also um die kritische Auseinandersetzung mit den Menschenbildern des Humanismus – die vielleicht viel zu europäisch und männlich geprägt sind – zugunsten alternativer Sichtweisen. Posthumanismus meint nicht die Überwindung der menschlichen Natur mit technologischen Mitteln (das wäre Transhumanismus), sondern die Überwindung des Humanismus:

1. Fliehkraft: dogmatische weltanschauliche Anthropozentrik und ein sich darin erschöpfender Humanismus, weiterhin Eurozentrik und Androzentrik
⇨ „narzisstische“ Wissensform, „blind“ für Alteritäten
2. Gegenpol: Elimination jeder Anthropozentrik, die Menschenzentriertheit der Ethik beenden bzw. das europäische und/oder männliche Menschenbild überwinden

- ⇨ **Radialkraft:** skeptische Aufklärung und kritischer Dialog
- ⇨ Angewandte Ethik: Fingerspitzengefühl, wirkliche Diskriminierung zu erkennen, sich nicht von Strohmannargumenten ablenken lassen, trotzdem Menschen, ihre Endlichkeit und Verletzlichkeit im Zentrum der Ethik zu erkennen

Im Sinne einer transdisziplinären Integration bietet sich hier der skeptische Dialog als methodische Arbeit an. Es passiert ja in der Tat viel zu schnell, dass unabsichtlich zu enge Menschenbilder zu Schäden führen. Ziel könnte moralische Klugheit sein, also authentisches, sinnlich-leibliches Wissen im Umgang mit Alterität sowie selbstreflexives Wissen im Umgang mit der eigenen Sprache. Es ließen sich weitere Radialkräfte ergänzen, die Problemfelder sind komplex. Doch sehen wir uns im Fortgang einfach zwei Beispiele an.

## Erstes Beispiel: AI Act und Verantwortung als Richtlinie

Im Mittelpunkt des EU AI Act steht die juristische Regulierung von **inakzeptablen Hochrisiko-KI-Systemen** unter der **Maßgabe vertrauenswürdiger und fairer Technologieentwicklung**. Mit den Grundrechten der Bevölkerung soll gleichzeitig die EU als Wirtschaftsstandort gestärkt werden.

> „The aim of the new rules is to foster trustworthy AI in Europe and beyond, by ensuring that AI systems respect fundamental rights, safety, and ethical principles and by addressing risks of very powerful and impactful AI models.“ (European Parliament 2024)

HerstellerInnen dürfen nach einheitlichen Maßstäben nur Produkte unter Achtung europäischer Prinzipien und Werte auf den Markt bringen (Schmid et al. 2023, S. 237-238). Auf Grundlage Harmonisierter Europäischer Normen (HEN) durchlaufen Produkte ein „Konformitätsbewertungsverfahren“, sodass auch für HerstellerInnen auf dem europäischen Binnenmarkt Rechtssicherheit gegeben ist. Regulatorische Hürden sollen innovative Technologiegestaltung nicht behindern. In den Gesetzestexten sind folglich auch Unterstützungsmaßnahmen enthalten. Ziel ist „eine gute Balance zwischen Risiken und Chancen“ sowie „ein Werkzeug einer im positiven Sinn nachhaltigen Entwicklung“ (ebd., S. 252). Grundlage ist die **Transparenz und Erklärbarkeit** funktionaler Eigenschaften von KI-Systemen, um Vertrauenswürdigkeit auf NutzerInnenseite zu ermöglichen. Drei Spezifika der Technikfolgenabschätzung werden übernommen (Grunwald 2003; Grunwald 2017):

1. Die Bewertung erfolgt **anwendungsspezifisch**. Zur Analyse der konkreten Anwendung werden klassifikatorische Merkmale vorgeschlagen, die sich an der Ersetzung menschlicher Handlungen durch konkrete KI-Methoden orientieren (Schmid et al. 2023, S. 93-102, S. 109-121).

2. Der eingeschränkte Satz funktionaler Optionen steht von vorneherein einer Operationalisierung von Ethik in einer KI entgegen. Im AI Act nimmt entsprechend die **menschliche Aufsicht bzw. Überwachung** der Technologien eine zentrale Rolle ein. Im Mittelpunkt steht der Schutz von Grundrechten (ebd., S. 238-239). Dahinter steckt die in der Technikphilosophie bekannte Einsicht:

„Im moralischen Sinne wird das computerisierte Entscheidungssystem, auch wenn es hinreichend lernfähig ist, dadurch nicht zu einer vollverantwortlichen moralischen Person als Träger der moralisch-praktischen Vernunft gemacht werden können.“ (Lenk 1994, S. 82-83)

3. **Risikomanagement** und **Risikoabschätzung** dienen der Erhöhung der Produktsicherheit unter Einbezug von **Zweck-Mittel-Analysen** (Anwendungsziele, Umnutzungs- bzw. Missbrauchspotenziale) (Grunwald 2017, S. 339-350; Grunwald (Hrsg.) (1999), S. 155-174; Lenk 2007 S. 166-197). Die Regulierung durch den AI Act zielt besonders auf Technologien mit „inakzeptablen“ Schadensrisiken für einzelne Menschen, Gemeinschaften oder Sachwerte ab. Geringe oder minimale Risiken soll mit allgemeinen Transparenzanforderungen und Verhaltenskodizes begegnet werden (Schmid et al. 2023, S. 128-136, S. 239-242).[3]

Wie schon bei den Ethikrichtlinien von 2019 (European Union 2019) bewirkt der AI Act eine stärkere Wahrnehmung ethischer Themen. Ethik lässt sich leicht rechtfertigen mit Verweis auf die Vorgaben der EU. Das alte Argument – „Ethik, wofür braucht man das?“ – scheint obsolet. Nach der UNESCO 2021 (Unesco 2022) hat auch die UN im März 2024 eine Resolution zu sicherer, vertrauenswürdiger und nachhaltiger KI – für alle Menschen während des kompletten Lifecycles – verabschiedet (United Nations 2024). Dabei handelt es sich um top-down verfasste Kompromisspapiere, die Empfehlungen aussprechen und für Kritik sorgen:

3 Je nach Darstellung werden 3 bis 5 Risikoklassen unterschieden. Entscheidend sind die Kriterien, anhand derer die Einstufung vorgenommen wird. Siehe auch: https://digital-strategy.ec.europa.eu/en/policies/regulatory-framework-ai [Stand vom 28-07-2024].

> „Selbstredend haben die UN-Diplomaten darauf geachtet, ihren Regierungen keine Steine in den Weg zum Schlachtfeld zu legen: Die KI-Resolution gilt ausdrücklich nicht für KI im Militärbereich. Immerhin steht auf der Tagesordnung der nächsten Generalversammlung eine UN-Resolution zu ‚Tödlichen Autonomen Waffensystemen'. Ein Konsensbeschluss dazu ist unwahrscheinlich. Schon im vorbereitenden Ausschuss haben unter anderem Russland und Indien dagegengestimmt." (Sokolov 2024)

Mit Blick auf den AI Act wird kritisiert, dass zwar Risiken im Mittelpunkt stehen, jedoch die Reflexion auf **wünschenswerte Zukünfte** zu kurz kommt:

> „The legislation's strong focus on risk also neglects the potential benefits of a technology. Regulators should aim to increase the overall value that the use of a technology creates for the public, so both risks and benefits should matter." (Prainsack/Forgò 2024)

Zu den generellen Bedenken bei Top-down-Richtlinien – über den Köpfen der Betroffenen – gehört die **globale Vereinnahmung durch industrielle Interessen**:

> „Self-regulating ethical frameworks allow companies to choose how to deploy technologies and, by extension, to decide what ethical AI means for the rest of the world. […] To understand what is at stake, we must focus less on ethics and more on power." (Crawford 2019, S. 224, siehe auch Gerdes 2022)

Die Problemlage reicht bis hin zu wirtschaftskritischen, postkolonialen Themen (Mbembe 2021) und dem Anliegen, koloniale Missstände sowie Imperialismus nicht beim **„Export" von Ethikrichtlinien** zu wiederholen. So wird schon seit Längerem die Bedeutung eines interkulturellen Kanons, feministischer Forschung und „minority studies" als Teil technikphilosophischer Grundlagen behandelt (Ihde 1993, S. 148-151). Hinsehen und Zuhören lässt sich von der Umwelt- und Klimaethik lernen, denn hier sind die interkulturellen Herausforderungen einer Ethik mit globalem Problemaufriss längst bekannt (Birnbacher 2022; Roser/Seidel 2015): **Guidelines hören nicht zu.** Umso wichtiger sind die verschiedenen globalen Stimmen bei der Regulierung von KI (Eke et al. 2023).

Es lassen sich Parallelen zur Diskussion technischer Katastrophen und IngenieurInnenverantwortung der 1980er und 1990er Jahre erkennen (Bhopal, Tschernobyl, Sandoz, Challenger-Unglück in Lenk 1994; sowie Lenk 1982). Zur ethischen Wissensdiversität in den Radialkräften des Umgangs mit KI gehören Guidelines (3.) aber auch ganz

positiv als integraler Teil der IngenieurInnenverantwortung – jedoch im Verbund mit moralischem Können (1.):

> „Moralische Urteilskraft und ehrliches Sich-Orientieren an den Folgerungen des Standesethos im Sinne bestmöglicher Aufgaben- und Pflichterfüllung ist bei allen diesen Abwägungsproblemen und besonders bei Uneinigkeiten über Wertprioritäten sowie bei Interessenkonflikten unerlässlich.“ (Lenk 1994, S. 106)

> „Und weil nur allgemeinmenschliche Interessen zählen, lassen sich die Erfahrungen anderer Kulturen integrieren: Die über Gedanken der Menschenrechte vermittelte Vorstellung von Identität ist im Prinzip flexibel und lernfähig.“ (Höffe 2014, S. 135)

Wer KI-Ethik auf das Abhaken von Guidelines reduziert (Nowotny 2023, S. 214), verstößt gegen diese Flexibilität und schlussendlich gegen Menschenrechte.

## Zweites Beispiel: „Narzisstische Bettgeschichten“

Radialkraft IV. lässt sich anhand der Diskussion um **Anthropozentrik und Intelligenz** vertiefen. Es geht dabei um die Frage, inwiefern Menschen im Zentrum der Ethik bleiben sollten und ob in weiterer Folge KI ethische *agency* hat – also selbst ein ethischer Akteur ist. Im ersten Schritt steht die Diskussion um Vielfalt menschlicher Intelligenzformen – im Gegensatz zum eingeengten Spektrum von Intelligenztests:

> „In Zukunft werden zahlreiche Disziplinen an der Erarbeitung einer Definition der Intelligenz beteiligt sein und noch weit mehr Interessengruppen an ihrer Vermessung und ihrem Einsatz.“ (Gardner 2002, S. 37)

Es ist ja eine *der* Fragen unserer Zeit, wenn wir „intelligence“ als Informationsverarbeitung ansprechen und gleichzeitig die Doppeldeutigkeit intelligenten Verhaltens in einem Computer zulassen. So wird im zweiten Schritt die Verengung von KI auf „Unternehmens-Intelligenz“ kritisiert. Viele andere Intelligenzformen, die gleichfalls für KI Pate stehen könnten, werden ignoriert (Bridle 2023, S. 21). Hierzu zählen Verflechtungen mit und Beziehungen zu anderen natürlichen Intelligenzen (Pilze, Wälder, Ökosysteme etc.) (ebd., S. 23) bis hin zu *agency*, Handlungsmacht oder „Person-Sein“ der zukünftigen Technologien (Ebd., S. 29).

> „Das abstrakte Denken biologischer Gehirne war die Stütze für die Entstehung aller Kultur und Wissenschaft. Aber diese Tätigkeit […] wird ein kurzer Vorläufer der leistungsfähigeren Erkenntnisvermögen des anorganischen, posthumanen Zeitalters sein.“ (Rees 2017, S. 38)

Es geht also im **Posthumanismus** nicht mehr um die Öffnung des Blicks für bisher weniger beachtete menschliche Intelligenzformen wie bei Gardner, sondern über den Menschen hinaus. James Bridle geht so weit, nicht nur in den Naturwissenschaften, sondern auch in der Ethik die Menschenzentriertheit als Problem darzustellen (Bridle 2023, S. 338-341).

Posthumanistisches Denken bildet als methodisch-aufklärerisches Motiv eine konstruktive Stimme in der Ethik. Es stellt skeptische Kritik zu Verfügung. Gleichzeitig birgt es die Gefahr der zwischenmenschlichen Diskriminierung oder des Dogmatismus. Sehen wir uns hierzu das viel beachtete Konzept des **agentiellen Realismus** von Karen Barad genauer an. Es gehört zu den Entwürfen neueren materialistischen Denkens (***New Materialism***) (Hoppe/Lemke 2021, S. 59-61). Dabei folgt Barad der Kritik an dualistischen Denkweisen, in denen Natur/Kultur, Mensch/Nichtmensch oder Frau/Mann als Gegensätze gesehen werden (Braidotti 2014; Haraway 2018; Tsing 2022). Es geht ihr um Performativität als

> „Infragestellung der unüberprüften Gewohnheiten des Geistes, die der Sprache und anderen Repräsentationsformen mehr Macht bei der Bestimmung unserer Ontologien zugestehen, als sie verdienen“ (Barad 2020, S. 10).

Neue geistige Gewohnheiten sollten Dynamiken der materiellen Welt stärker einschließen. Die Autorin verbindet damit ein posthumanistisches Programm, das sich radikal von Humanismus und Anthropozentrik absetzen will.

Der methodische Gehalt ihres Anliegens lässt sich anhand folgender Passagen nachvollziehen: „Humanistische Umlaufbahnen“ sollen verlassen werden, womit eben der „Griff anthropozentrischer Kräfte“ in „zeitgenössischen Denkmustern“ gemeint ist. „Hochgeschätzte Ionisierungen“ gingen von **poststrukturalistischen AutorInnen** wie Michel Foucault oder Judith Butler aus (siehe auch zu Deleuze und Guatteri: Hubatschke 2024). Jedoch blieben diese gehemmt von „anthropozentrischen Überbleibseln“. Das wiederum soll in Barads „posthumanistischen, performativen“ Neuausrichtungen überwunden werden. Der „Ausnahmezustand des Menschen“ ist „aufs Korn zu nehmen“ (Barad 2020, S. 10-13). In einer „agentiell-realistischen Ontologie“ wird Materie als dynamische Agenten (*agencies*) betrachtet. Sie prägt aktiv die Relationen der Welt so-

wie „intraaktive, diskursive“ Praktiken. Materie „handelt“ und hat „Geschichte“ (ebd., S. 14-22). Der methodische Gehalt liegt in einer **Kritik des isolierten, archimedischen BeobachterInnenstandpunktes**. Barad setzt sich hierzu in einer Fallstudie mit Niels Bohr und den erkenntnistheoretischen Folgen der Quantenphysik auseinander.

Jedoch kann die Autorin zum einen ihren radikalen Wunsch, ohne jede Anthropozentrik auszukommen, nicht durchhalten. Denn sie ist ja eine menschliche Wissenschaftlerin, die in menschlicher Sprache für andere Menschen schreibt, mit dem Ziel, deren Denkgewohnheiten zu verändern. Sie bleibt einer methodisch-sprachkritischen Anthropozentrik verhaftet. Mit Selbstironie kommentiert Barad ihr Vorhaben als Fußnote:

> „Es wäre verwunderlich, wenn mein eigener Versuch, einen erfolgreichen ionisierenden ‚Quantensprung‘ aus der humanistischen-repräsentationalistischen Umlaufbahn heraus zu machen, nicht derselben Anziehungskraft zum Opfer fiele […].“ (ebd., S. 103)

Sie deutet damit die methodische Reichweite ihres reflexiven Versuchs an, zumindest die Rolle menschlicher BeobachterInnen in den Naturwissenschaften kritisch einzuordnen.

Auf der anderen Seite öffnet Barad missverständliche Türen für ein dogmatisches Umkippen ihrer Ideen. So benutzt sie eine **technomorphe Sprache** („Quantensprünge“, „ionisieren“), wodurch Menschen als bloße Naturereignisse angesprochen sind – und das, ohne ihre Metaphern methodisch-kritisch einzuordnen. „Handlung“ und „Geschichte“ fallen aus dem Bereich menschlicher Verantwortung, indem sie zu Merkmalen allgemeiner materieller Prozesse werden. Das ist weder mit der methodischen Trennung von Sein und Sollen vereinbar, noch mit den Grundlagen moderner Rechtsordnungen (Beweisführung auf der Tatsachenebene und individuelle Schuldfrage laut geltender Gesetze). Hinzu tritt die Reduktion von Humanismus auf anthropozentrische Prämissen der modernen Physik. Es wird sogar so pauschal verfahren, dass in einem Handstreich gleichzeitig die normativ-liberalen Kräfte des Humanismus sowie das verbundene Bildungsanliegen zur Ermächtigung politischer Diskurse ausgeblendet werden (ebd., S. 16, S. 29, S. 70). Selbst wenn für Naturwissenschaften die Kritik treffen sollte, bliebe deren Gültigkeit für Gesellschaftswissenschaften eigens zu prüfen. Darüber hinaus schließt die Autorin, dass „wir“ im Vollzug in der Welt „für Markierungen auf Körpern verantwortlich“ sind (ebd., S. 88). Dieser ethische Gemeinplatz kippt in sein Gegenteil an der Stelle, wo Barad das „wir“ ausdrücklich auf nichtmenschliche materielle Agenten ausdehnt: „Leider schloss Bohr jedoch wichtige posthumanistische Implikationen dieser Einsicht mit seinem letztlich humanistischen Verständnis des ‚wir‘ kurz“ – so ihre Kritik am Ende der Fallstudie (ebd., S. 99).

Selbst wenn dieses pantheistische (?), jedenfalls materialistisch-physikalistische Plädoyer für Verantwortung gut gemeint sein mag, trägt es doch eine gefährliche Last in sich: Von Menschen hergestellte Maschinen wären dann mindestens genauso wichtig wie andere Menschen.

> *Folglich führen Barads neue „geistige Gewohnheiten" dazu, Verantwortung an der Elektrik von Computern zu orientieren und reale Ungerechtigkeiten gegenüber anderen Menschen zu übersehen. Die auch als Feministin bekannte Autorin öffnet damit Denkweisen Tür und Tor, die doch eigentlich nicht in ihrem Interesse liegen dürften.*

Verantwortung wird zu Beliebigkeit und die Rolle der Politik in ihrem Szientismus zu schwach reflektiert (Hoppe/Lemke 2021, S. 74-79). Verschärft wird diese Hybris durch die vorsorgliche Abfertigung alternativer Konzepte:

> „Die einzigen Möglichkeiten [zur Konzeptualisierung der Materie] scheinen die Naivität des Empirismus oder dieselben alten **narzisstischen Bettgeschichten** zu sein [außer man vertrete eine] posthumanistische Sichtweise von Performativität [...]." (Barad 2020, S. 97, Hervorhebung v. M.F.)

An Punkten wie diesen setzt sich Barad ähnlicher Kritik aus wie AutorInnen, die sie in den Schubladen des Humanismus, der Aufklärung oder Anthropozentrik pauschal anprangert: Stigmatisierung durch fehlendes Hinsehen und Zuhören, schlussendlich Ausblenden der verwickelten Vielfältigkeiten humanistischer Lebenspraxen.

Auch Rosi Braidotti gilt als prägende Vertreterin des **feministischen, kritischen Posthumanismus** (Hoppe/Lemke 2021, S. 83). Ihr geht es im Besonderen um den Widerspruch gegen Eurozentrismus, Imperialismus, Kolonialismus und die Annahme eines einheitlichen Subjekts. Sie formuliert einen monistischen Ansatz, demzufolge ein „Paradigmenwechsel" in „Begriffen, Methoden und politischen Praktiken" der „Gesellschaftstheorien" erfolgen muss. Im Anschluss an Bruno Latour soll ihre „Agenda der posthumanen Verwicklung" einen methodischen Anspruch einlösen und als Instrumentarium zum Verständnis menschlicher wie nichtmenschlicher Akteure im lebendigen Ökosystem der Erde dienen (Braidotti 2014, S. 6-11). Zentral ist hierfür die Verabschiedung des Anthropozentrismus aus den Geisteswissenschaften (ebd., S. 61-72, S. 147-149). Die binäre Logik aus Selbst und Anderem soll überwunden werden. Im antihumanistischen Programm ist kein Platz für „*den* Menschen" (ebd., S. 19-31, S. 172-176), dementsprechend ist Posthumanismus darüber hinaus mit der Suche nach alternativen Formen des menschlichen Subjekts befasst (ebd., S. 42-50).

Es geht um die politische Vision der „Wiederherstellung eines panhumanen kosmopolitischen Bands“, das in Folge eines „selbstherrlichen“, „eurozentrischen“ Humanismus verloren gegangen sei (ebd. S. 57-59). Hierzu ist eine „radikale Entfremdung von den Begriffen moralischer Rationalität, einheitlicher Identität, transzendentalen Bewusstseins oder angeborener und allgemeiner moralischer Werte“ nötig. „Postanthropozentrischer Posthumanismus“ meint eine „Ökosophie“, „die die vielfältigen Schichten des Subjekts von innen nach außen über alles, was dazwischen liegt, transversal zu durchziehen sucht“ (ebd, S. 97). Als kritische Sicht auf die Gefahren selbstherrlicher Menschenbilder eröffnet Braidottis Entwurf wichtige Perspektiven und kann als Mahnung zur Erarbeitung einer nicht eurozentrischen KI-Ethik verstanden werden. Jedoch besteht auch hier die Gefahr, dass durch die Inklusion von technischen Mitteln wie Computern in einen ökosophischen Monismus eben gerade nicht tolerantes und offenes politisches Verhalten folgt:

> „Die so verstandene ‚Maschinenwerdung‘ bezeichnet und realisiert die relationalen Kräfte eines Subjekts, das nicht mehr in einen dualistischen Rahmen gezwängt ist, sondern in einer bestimmten Verbindung zu vielfältigen Anderen steht und in seiner technologisch vermittelten planetarischen Umwelt aufgeht.“ (ebd., S. 96)

Die Verwicklung mit Maschinen kann dann – wenn Menschen sich nicht als ihre HerstellerInnen herausheben dürfen – zum fatalen Kurzschluss führen:

> *Es wäre niemand mehr für die Technologien verantwortlich, Ungerechtigkeit zwischen Menschen wäre kein normatives Übel mehr, sondern eine bloße relationale Kraft, der nicht mehr mit rationaler Kritik und zwischenmenschlicher Courage begegnet werden könnte.*

Die destruktiven Potentiale materieller, technologischer Prozesse kann die Autorin durch ihren pauschalen Wurf nicht einfangen (Hoppe/Lemke 2021, S. 96-97). So erfrischend methodischer Posthumanismus auch ist, ohne selbstkritische Skepsis, methodisch-sprachkritische Anthropozentrik und das Abarbeiten an humanistischen Idealen beißt sich die Katze in den Schwanz. KI-Ethik kann jedoch an der Reflexion dieser kritischen Einwände wachsen. Darum sollte differenziert zwischen den methodischen Chancen solcher Kritik und den dogmatischen Irrwegen unterschieden werden. Es geht dabei um die Wahrnehmung ethischer Probleme. Offensichtlich ist das Sichern menschlicher, politischer Autonomie viel wichtiger in Zeiten bedrohter Privatheit – durch Google oder Facebook – (Pauen/Welzer 2015, S. 273-286) als der energiezehrende Streit für autonome und gleichberechtigte – übrigens von Menschen hergestellte – technische

Güter. Gleiches gilt für das **permanente Ringen um Menschenrechte und Menschenwürde**, in deren ethischem Zentrum nun einmal Menschen stehen (Pollmann 2022). Hinzu treten Probleme durch weltweite Kriege und Umweltzerstörung, Beeinflussung demokratischer Wahlen durch Deep Fakes oder unabsichtliche, aber reale Diskriminierung durch intransparente Modellbildung in machine learning-Systemen.

> *Unterlassung ist auch eine Handlung: Es ist darum eigens zu rechtfertigen, warum diese realen Probleme zugunsten posthumanistischer Gleichmacherei von Menschen und Maschinen hintangestellt werden sollten.*

Als Radialkraft, die unsere ethischen Kufen in der Bahn hält, wirkt wohl eine Vermittlung beider Sichtweisen: Ethik ist immer anthropozentrisch und KI ist nach wie vor schlicht von Menschen hergestellte „technische Informationsverarbeitung". Jedoch hat Technik – einschließlich KI – auch so etwas wie „Ethik" oder „Politik", jedoch im übertragenen Sinne: sie verkörpert Werte, Normen, Menschen- und Weltbilder derjenigen, die sie hergestellt haben (Winner 1980). Ethische Wissensdiversität entfaltet sich in diesen reflexiven Wechselbewegungen. Daraus nun zu folgern, dass eine KI selbst ethisch handelt, kann zwar die zwischenmenschliche Debatte anfeuern oder vergiften. Jedenfalls steigt damit das Risiko zwischenmenschlicher Diskriminierung und des Rückfalls hinter Errungenschaften der Menschenrechte und Menschenwürde. Eine Rechtfertigung, diesen Preis zu zahlen und in eine selbstverschuldete Unmündigkeit zu fallen, steht aus …

## Zusammenfassung

Vorliegender Beitrag schließt unmittelbar an meine Analysen zur Wissensdiversität (vgl. Funk 2024) an. Dabei wurden vier Spannungsfelder einschließlich – bildlich gesagt – Radialkräften der Ethik exemplarisch unterschieden (Abb. 1). Hierzu zählen das Ringen gegen

I. die Vereinnahmung von Ethik durch industrielle Interessen,
II. die Beschränkung auf top-down vorgegebene „Checklistenethik",
III. den lebensfremden fachphilosophischen Elfenbeinturm sowie
IV. posthumanistische Irrtümer.

Es lassen sich von der Orientierung individueller Lebensentwürfe bis hin zur globalen Geostrategie zahlreiche Spannungsfelder ergänzen. Anhand von zwei Beispielen wurden die vier ausgewählten vertieft. Zuerst ging es um den AI Act und Ethikricht-

linien. Top-down-Regulierung läuft Gefahr, Ethik auf Guidelines zu reduzieren und dabei sowohl moralische Alltagspraxis, also auch Bottom-up-Entwicklungspotenziale zu übersehen. Das kann umso größere Probleme nach sich ziehen, wenn beim weltweiten „Export“ europäischer Richtlinien auch noch koloniale Fehler wiederholt werden. Industrielle Interessen zielen auf eine möglichst weiche, rechtlich bindende Regulierung zugunsten ethischer Selbstverpflichtungen ab. Ethik wird also gegen geltendes Recht ausgespielt. Auf der anderen Seite fürchten Unternehmen in Europa um Wettbewerbsnachteile, insbesondere wenn der KI-Markt ohnehin von global agierenden Firmen vor allem der US-Westküste beherrscht wird. Im AI Act wird das komplexe Problemfeld zu sehr auf Risikomanagement reduziert. Ethik ist mehr, sie ist auch prüfendes Vorausdenken darüber, was wir eigentlich im positiven Sinne von KI erwarten – also Debatte über den wünschenswerten „public benefit“ (Prainsack/Forgó 2024). Dafür bedarf es im authentischen Sinne des Wortes **ethisch-demokratischer Lebenskunst statt Multipe-Choice-Bürokratie**.

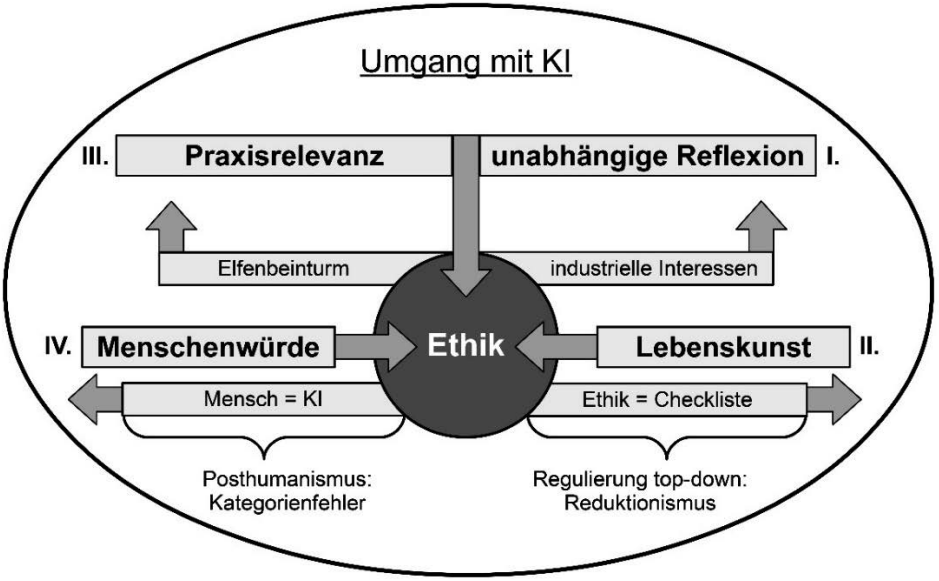


*Abb. 1: Vier Radialkräfte der Ethik*

Das zweite Beispiel rückte posthumanistische Kritik an menschenzentrierter, humanistischer Ethik in den Mittelpunkt. Dabei traten folgende Probleme auf:

1. Es gibt verschiedene Humanismen, Humanismus ist mehr als nur Anthropozentrik, es gibt diverse Anthropozentriken (z.B. weltanschauliche oder methodische Anthropozentrik).
2. Physische Objekte einschließlich von Menschen hergestellter technischer Mittel werden nicht nur deskriptiv, sondern normativ als Alteritäten auf einer Stufe mit Menschen anerkannt.

⇨ **Posthumanistischer Fehlschluss:** technische Informationsverarbeitung (KI) wird als moralischer Akteur auf eine Stufe mit Menschen gestellt; damit wird ein rationalistisch-reduktionistisches Weltbild aufgewertet (formale Sprache, propositionales Wissen, Zweck-Mittel-rational in von Menschen hergestellte technische Güter implementiert), das doch eigentlich als „anthropozentrischer Humanismus" an den Pranger gestellt und überwunden werden soll – denn mit Humanismus ist auch die Epoche moderner Natur- und Technikwissenschaften angesprochen.

⇨ Dieser Kategorienfehler wirkt toxisch auf die kritische Infrastruktur wirksamer Wissensdiversität I.

3. Der Blick wird von realen zwischenmenschlichen Ungerechtigkeiten abgelenkt, weiterhin von Problemen politischer Teilhabe sowie Herausforderungen interkultureller Wahrnehmungen und respektvollen Dialogs.
4. Auflösung rationaler Debatten in relativistische Beliebigkeiten, Stigmatisierung von Wahrheitssuche (statt Öffnung hin zur Diversität verschiedener Wissensformen).

⇨ **Verstärkung der Gefahren** von Fakes, Propaganda, Gefährdung von Menschenrechten.

Die Ursache liegt im Defizit differenzierter Humanismus- und Anthropozentrikforschung. Stattdessen wird diese im pauschalen Handstreich verdammt. „Narzisstische Bettgeschichten" (Barad) können nicht nur von Menschen erzählt werden, die als „humanistisch" oder „anthropozentrisch" bezeichnet werden, sondern auch von Menschen, die sich selbst „posthumanistisch" nennen. Relationalismus braucht ein methodisch anthropozentrisches Fundament (einschließlich Sprachkritik; Funk 2022b, S. 23-52), um nicht unabsichtlich politisch toxische Diskriminieren oder Silencing zu befördern. Auf der anderen Seite können gerade die vielfältigen Stimmen feministischer, postkolonialer oder posthumanistischer Kritik einen wichtigen Beitrag zur (menschlichen) Wahrnehmung gelebter Wissensdiversität leisten und schlussendlich die kritischen Infrastrukturen demokratischer Gesellschaften stärken. Es bleibt ein Balanceakt auf glattem Eis …

## Literatur

Barad, Karen (2020): *Agentieller Realismus.* Berlin: Suhrkamp.
Birnbacher, Dieter (2022): *Klimaethik.* Stuttgart: Reclam.
Braidotti, Rosi (2014): *Posthumanismus.* Frankfurt a.M./New York: Campus.
Bridle, James (2023): *Die unfassbare Vielfalt des Seins.* München: C.H. Beck.
Coeckelbergh, Mark (2022): *The Political Philosophy of AI.* Cambridge u.a.: Polity.
Crawford, Kate (2019): *Atlas of AI.* New Haven/London: YUP.
Dignun, Virginia (2019): *Responsible artificial intelligence.* Cham: Springer.
Eke, Damian; Wakunuma, Kutamo & Akintoye, Simisola (Hrsg.) (2023): *Responsible AI in Africa.* Cham: Springer.
Eubanks, Virginia (2018): *Automating Inequality.* New York: St. Martin's.
European Parliament (2024): AI Act https://digital-strategy.ec.europa.eu/en/policies/regulatory-framework-ai [Stand vom 28-07-2024]
European Union (2019): Ethics guidelines for trustworthy AI, published 8 April 2019 https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai [Stand vom 28-07-2024].
Fischer, Johannes (2022): *Die Zukunft der Ethik.* Tübingen: Mohr Siebeck.
Foot, Philippa (2023): *Die Natur des Guten.* Berlin: Suhrkamp.
Funk, Michael (2022a): *Roboter- und KI-Ethik.* Wiesbaden: Springer.
Funk, Michael (2022b): *Angewandte Ethik und Technikbewertung.* Wiesbaden: Springer.
Funk, Michael (2023): *Ethik künstlicher Intelligenz.* Wiesbaden: Springer.
Funk, Michael (2024): Nichtwissen ist Macht! Künstliche Intelligenz und ihre diversen Ethiken. In: Beinsteiner, Andreas; Dittrich, Ann-Kathrin & Hug, Theo (Hrsg.): *Medien – Wissen – Bildung: Wissensdiversität und formatierte Bildungsräume*. Innsbruck: iup, S. 41–58.
Gardner, Howard (2002): *Intelligenzen. Die Vielfalt des menschlichen Geistes.* Stuttgart: Klett-Cotta.
Gerdes, Anne (2022): The tech industry hijacking of the AI ethics research agenda and why we should reclaim it. *Discov Artif Intell* 2:25.
Grunwald, Armin (Hrsg.) (1999): *Rationale Technikfolgenbeurteilung. Konzeption und methodische Grundlagen.* Heidelberg: Springer.
Grunwald, Armin (Hrsg.) (2003): *Technikgestaltung zwischen Wunsch und Wirklichkeit.* Heidelberg: Springer.
Grunwald, Armin (2017): *Technik und Politikberatung.* Frankfurt a.M.: Suhrkamp.
Haraway, Donna (2018): *Unruhig bleiben.* Frankfurt a.M./New York: Campus.
Hoerster, Norbert (2022): *Was ist Moral?* Stuttgart: Reclam.
Höffe, Otfried (2014): *Die Macht der Moral im 21. Jahrhundert.* München: C.H. Beck.
Hoppe, Katharina & Lemke, Thomas (2021): *Neue Materialismen.* Hamburg: Junius.

Horster, Detlef (2009): *Ethik.* Stuttgart: Reclam.
Hossenfelder, Malte (Hrsg.) (2013): *Antike Glückslehren.* Stuttgart: Kröner.
Hubatschke, Christoph (2024): *Minoritäre Technologien.* Frankfurt a.M./New York: Campus.
Ihde, Don (1993): *Postphenomenology. Essays in the Postmodern Context.* Evanston, Illinois: Northwestern University Press.
Jonsen, Albert & Toulmin Stephen (1988): *The abuse of casuistry.* Berkeley: University of Chicago Press.
Kallhoff, Angela (2022): *Der Mensch – das moralische Tier.* Berlin: Suhrkamp.
Lenk, Hans (1982): *Zur Sozialphilosophie der Technik.* Frankfurt a.M.: Suhrkamp.
Lenk, Hans (1994): *Macht und Machbarkeit der Technik.* Stuttgart: Reclam.
Lenk, Hans (2007): *Global TechnoScience and Responsibility.* Berlin: LIT.
Masek, Michaela (2023): *Antike Glücksethik.* Wien: facultas.
Mbembe, Achille (2021): *Kritik der schwarzen Vernunft.* Berlin: Suhrkamp.
Metzinger, Thomas (2019): Nehmt der Industrie die Ethik weg! Gastbeitrag im *Tagesspiegel* vom 08.04.2019 https://www.tagesspiegel.de/politik/nehmt-der-industrie-die-ethik-weg-5548855.html [Stand vom 28-07-2024].
Murdoch, Iris (2023): *Die Souveränität des Guten.* Berlin: Suhrkamp.
Nowotny, Helga (2023): *Die KI sei mit euch.* Berlin: Matthes & Seitz.
Nussbaum, Marta (2020): *Gerechtigkeit oder Das gute Leben.* Frankfurt a.M.: Suhrkamp.
Pauen, Michael & Welzer, Harald (2015): *Autonomie. Eine Verteidigung.* Frankfurt a.M.: Fischer.
Pollmann, Arnd (2022): *Menschenrechte und Menschenwürde.* Berlin: Suhrkamp.
Prainsack, Barbara & Forgó, Nikolaus (2024): New AI regulation in the EU seeks to reduce risk without assessing public benefit. *Nat Med*, Comment 18 March 2024. https://doi.org/10.1038/s41591-024-02874-2 [Stand vom 28-07-2024].
Rees, Martin (2017): Die organische Intelligenz hat keine langfristige Zukunft. In: Brockman, John (Hrsg.): *Was sollen wir von Künstlicher Intelligenz halten?* Frankfurt a.M.: Fischer, S. 36–39.
Roser, Dominic & Seidel, Christian (2015): *Ethik des Klimawandels.* Darmstadt: WBG.
Schmid, Thomas; Hildesheim, Wolfgang & Holoyad, Taras (2023): *Künstliche Intelligenz managen und verstehen.* Berlin u. a.: Beuth.
Schwemmer, Oswald (1980): *Philosophie der Praxis.* Frankfurt a.M.: Suhrkamp.
Sokolov, Daniel AJ (2024): UN-Generalversammlung: Mehr KI für Gutes, keine KI für Schlechtes, *heise* vom 22.03.2024 https://heise.de/-9662662 [Stand vom 28-07-2024].
Somek, Alexander (2021): *Moral als Bosheit.* Tübingen: Mohr Siebeck.
Thurnherr, Urs (2010*): Angewandte Ethik.* Hamburg: Junius.
Tsing, Anna (2022): *Der Pilz am Ende der Welt.* Berlin: Matthes & Seitz.
Tugendhat, Ernst (1984): *Probleme der Ethik.* Stuttgart: Reclam.

Unesco (2022): Recommendations on the Ethics of Artificial Intelligence. Adopted on 23 November 2021. https://unesdoc.unesco.org/ark:/48223/pf0000381137/PDF/381137eng.pdf.multi [Stand vom 28-07-2024].

United Nations (2024): General Assembly adopts landmark resolution on artificial intelligence, 21 March 2024 https://news.un.org/en/story/2024/03/1147831 [Stand vom 28-07-2024].

Williams, Bernard (1986): *Der Begriff der Moral.* Stuttgart: Reclam.

Winner, Langdon (1980): Do Artifacts Have Politics? *Daedalus* 109 (1), S. 121–136.