



# Fuzzy Ethizität: Radar für ethische Künstliche Intelligenz

Michael Kaufmann  · Andreas Meier 

Eingegangen: 5. Oktober 2021 / Angenommen: 28. Februar 2022 / Online publiziert: 29. März 2022  
© Der/die Autor(en) 2022

**Zusammenfassung** Eine Informations- und Wissensgesellschaft betreibt ihre Wertschöpfungskette primär durch Aufbau, Verarbeitung und Weitergabe von Informationen (digitalen Produkten und Dienstleistungen) respektive von Wissen (intelligenten Produkten und Wissensdienstleistungen). Mit Künstlicher Intelligenz angereicherte Produkte und Dienstleistungen erhöhen deren Verfügbarkeit und Sicherheit, indem Entscheidungen datenbasiert automatisiert werden. Allerdings müssen Ethikregeln berücksichtigt werden, um die Vertrauensbasis digitaler Gesellschaften zu stärken und nachhaltige Entwicklung zu fördern. Im Folgenden wird ein Radarsystem für die Bewertung der Ethizität im Digitalzeitalter vorgeschlagen mit dem Ziel, eine ethische Entwicklung von intelligenten Informationssystemen zum Wohl der Gemeinschaft zu überprüfen. Die sechs Dimensionen dieses Systems basieren auf der Grundlage von Menschenrechten und einer umfassenden Analyse bestehender Ethik-Rahmenwerke für Künstliche Intelligenz: Gleichheit & Gerechtigkeit, Schutz & Hilfe, Privatsphäre & Datenschutz, Freiheit & Autonomie, Solidarität & Nachhaltigkeit und Verantwortlichkeit & Transparenz. Der Beitrag diskutiert diese sechs Dimensionen digitaler Ethik und schlägt ein Radarsystem zur visuellen Analyse der Ethizität von intelligenten Informationssystemen vor. Danach wird das Radarsystem mithilfe der Fallstudie Privacy Setting Framework auf Nützlichkeit evaluiert. Ein Fazit des Beitrags summiert die Erkenntnisse.

**Schlüsselwörter** Digitale Ethik · Ethizität · Künstliche Intelligenz · Radarsystem · Gradual Computational Ethics · Fuzzy Logic

---

Michael Kaufmann (✉)  
Hochschule Luzern, Rotkreuz, Schweiz  
E-Mail: [m.kaufmann@hslu.ch](mailto:m.kaufmann@hslu.ch)

Andreas Meier  
Universität Fribourg, Fribourg, Schweiz  
E-Mail: [andreas.meier@unifr.ch](mailto:andreas.meier@unifr.ch)

**Abstract** An information and knowledge society operates its value chain primarily by building, processing, and passing on information (digital products and services) and knowledge (intelligent products and knowledge services). Artificial intelligence-enriched products and services increase their availability and security by automating data-based decisions. However, ethics rules need to be considered to strengthen the trust basis of digital societies and promote sustainable development. In the following, a radar system for the evaluation of ethicity in the Digital Age is proposed with the aim of checking ethical development of intelligent information systems for the benefit of the community. The six dimensions of this system are based on human rights and a comprehensive analysis of existing ethics frameworks for Artificial Intelligence: equality & justice, protection & assistance, privacy & data protection, freedom & autonomy, solidarity & sustainability, and accountability & transparency. The paper discusses these six dimensions of digital ethics and proposes a Radar System for visually analyzing the ethics of intelligent information systems. Afterwards, the radar system is evaluated for usefulness using the Privacy Setting Framework case study. A conclusion of the paper sums up the findings.

**Keywords** Digital ethics · Ethicity · Artificial intelligence · Radar system · Gradual computational ethics · Fuzzy logic

## 1 Motivation

Wirtschaft und Gesellschaft befinden sich hinsichtlich der Digitalisierung von Wertschöpfungsketten und der dadurch folgenden Automation von Entscheidungen in einem Wandel. Unter dem Stichwort digitale Transformation werden Geschäftsmodelle auf den Kopf gestellt, Prozesse von Grund auf neu konzipiert, intelligente Monitore für die Steuerung komplexer Informationslandschaften eingesetzt oder digitale Automaten für diverse Service-Dienstleistungen aller Anspruchsgruppen erprobt. Digitale Daten beziehungsweise daraus extrahierte Informationen und entstandenes Wissen sind der Treibstoff dieser Entwicklung. Analog sind Methoden der Künstlichen Intelligenz (KI) der Motor. Mit der explosionsartigen Entwicklung der weltweiten Datensammlung des World Wide Web (WWW) wurden Datenmengen verfügbar, welche KI-Systeme praxistauglich machen. Beispiele dafür sind Spamfilter, Navigationssysteme, teilautomatisierte Autos, Sprachsteuerung, Gesichtserkennung und Empfehlungssysteme. Damit wird unser Leben mehr und mehr durch intelligente Systeme und automatisierte Entscheidungen bestimmt und es stellt sich die Grundsatfrage, wie eine Ethik im Digitalzeitalter gestaltet und umgesetzt werden könnte (Grimm et al. 2021; Spiekermann 2019).

Unter Ethik versteht die Internet Encyclopedia of Philosophy die Systematisierung, Verteidigung und Empfehlung von Konzepten für richtiges und falsches Verhalten (Fieser 2021). So verlangt beispielsweise Kant ethische Normen, die nicht aus der Erfahrung abgeleitet sind, sondern allgemeine Gültigkeit beanspruchen und für alle Menschen gleich verbindlich sind. Eine vernünftige, am allgemeinen Gesetz orientierte Handlung braucht deswegen noch nicht moralisch gut zu sein. Sie ist es nach Kant erst dann, wenn das innerlich zustimmende Wollen hinzukommt. So lau-

tet Kants kategorischer Imperativ: Handle nur nach denjenigen Regeln, von denen du zugleich möchtest, dass sie ein allgemeines Gesetz werden.

Kant's kategorischer Imperativ soll nach wie vor gelten, doch wie soll er im Global Village umgesetzt werden? Heute bestimmen ausgeklügelte Algorithmen, oft angereichert mit Künstlicher Intelligenz unseren Alltag. Empfehlungssysteme begleiten uns auf Schritt und Tritt, sei es bei der Suche nach einem Restaurant, einem Krimi, einem Kunstraum oder eines Lebenspartners oder -partnerin. Hier stellt sich die Herausforderung, nach welchen Grundprinzipien diese Algorithmen entwickelt, evaluiert und genutzt werden sollen. Wie kann ethisches Handeln bei der Delegation von Entscheidungen an Maschinen gewährleistet werden?

In Abschnitt 2 werden wir zuerst die Entwicklungsstufen der Maschinenethik diskutieren. Zudem schlagen wir vor, anstelle zweiwertiger Logik mit den Wahrheitswerten wahr und falsch stattdessen Logiken mit unbegrenzten Wahrheitswerten (Fuzzy Logic) anzuwenden (Abschn. 3). Diese ermöglichen Differenzierungen, um Einschätzungen und Entscheidungen mit Augenmaß zu fällen. Anstelle von schwarz und weiss (entweder-oder) werden Grautöne miteinbezogen, die das Prinzip sowohl als auch unterstützen. Als Werkzeug für die Bewertung der Moralität von intelligenten Informationssystemen schlagen wir ein Radarsystem vor, das sechs Dimensionen aufweist (Abschn. 4). Wir zeigen auf, wie dieses Radarsystem auf der Basis der unscharfen Logik ermöglicht, auf einen Blick die ethischen Auswirkungen von KI-Anwendungen zu erfassen (Abschn. 5). Mithilfe der unscharfen Logik sowie dem Ethik-Radar wird das Privacy Setting Framework in Abschnitt 6 vertieft: Hier wird ersichtlich, wie politische Partizipation gelebt und politische Empfehlungssysteme unter Schutz der Privatsphäre im Zeitalter der Maschinenethik realisiert werden könnten. Ein Fazit mit den Lessons Learned und einem Ausblick (Abschn. 7) runden den Beitrag ab.

## 2 Ethizität Künstlicher Intelligenz

Laut C. Misselhorn beschäftigt sich die Maschinenethik mit der Möglichkeit, Maschinen mit der Fähigkeit zu moralischem Entscheiden und Handeln auszustatten (Misselhorn 2019, S. 70). Nun stellt sich die Frage, ob Maschinen (Hardware und Software) überhaupt moralische Akteure sein können.

Professor J. H. Moor war Ethiker am Department of Philosophy des Dartmouth College in Hanover, New Hampshire, USA. In seinem Forschungspapier „The Nature, Importance, and Difficulty of Machine Ethics“ (Moor 2006) unterscheidet er drei Entwicklungsstufen zur Maschinenethik:

- **Implicit ethical agents:** Solch implizit ethische Agenten berücksichtigen bestimmte Wertvorstellungen unserer Ethik und können Sicherheitsabwägungen vornehmen. Beispiele solcher Agenten sind Warnsysteme in Flugzeugen, die Alarm schlagen, sobald sich ein Flugzeug auf Kollisionskurs mit einem anderen Flugobjekt befindet oder falls Flughöhen oder Sicherheitsabstände unterschritten werden.

- **Explicit ethical agents:** Hier kann ein Computersystem moralisch relevante Informationen erkennen, verarbeiten und entsprechende Entscheidungen selbst treffen. James Moor schreibt dazu: Diese Agenten verwenden zum Beispiel einen „learning algorithm to adjust judgments of duty by taking into account both prima facie duties (Pflichten auf den ersten Blick) and past intuitions about similar or dissimilar cases involving those duties“. Dabei denkt Moor u. a. an autonome Roboter, die in Krisengebieten zum Einsatz kommen (z. B. zur Entschärfung von Minen) oder zu militärischen Zwecken verwendet werden (intelligente Drohnen).
- **Full ethical agents:** Hier wird ein hoher Anspruch an Agenten gestellt, denn „a full ethical agent can make explicit ethical judgments and generally is competent to reasonably justify them“. Menschen haben Bewusstsein und freien Willen und Moor fragt sich zurecht, wieweit eine Maschine als voll ethischer Agent wirken kann.

Bei der Unterscheidung dieser drei Entwicklungsstufen für ethische Agenten können wir heute festhalten, dass implizite bereits in verschiedenen Anwendungsgebieten im Einsatz sind und dass explizite ethische Agenten für Teilgebiete realisiert werden können.

Allerdings bleibt die Frage offen, ob Maschinen in Zukunft den Reifegrad von voll ethischen Agenten erreichen. Professor P. G. Kirchschräger (2019) von der Universität Luzern verneint dies. Für Kirchschräger ist der Begriff der „moralischen Technologien“ irreführend, da technologische Systeme zwar Normen befolgen und umsetzen können, aber nicht die moralische Fähigkeit besitzen, gemäß Kant willentlich moralische Normen zu entwerfen, die verallgemeinert werden können. Es fehlt den Maschinen also an Autonomie, Freiheit und Gewissen, um moralisch zu handeln. Daher spricht Kirchschräger technologischen Systemen moralische Fähigkeit ab, selbst wenn sie so entscheiden und handeln, als ob sie moralische Akteure wären. Damit bleibt die moralische Verantwortung bei den Menschen, welche die Maschinen und ihre Funktion in irgendeiner Form mitgestalten, sei es als Entwicklerin, Designer, Managerin, Investor, Politikerin oder Forscher. Die moralische Verantwortung liegt bei den *Architekten* der Künstlichen Intelligenz.

Wir gehen in diesem Zusammenhang von der Annahme aus, dass menschliche Intelligenz nicht künstlich konstruierbar ist. Schwache KI oder komputationale Intelligenz bedeutet, dass Maschinen entwickelt werden, die sich so verhalten, als ob sie intelligent wären; aber sie verfügen über kein Bewusstsein, Empfindungsvermögen, gesunden Menschenverstand oder menschliche Intelligenz. Daher werden sie nicht fähig sein, ethisch zu handeln. Doch selbst wenn diese Hypothese in Zukunft durch humanoide oder biologische Rechner aufgeweicht würde, dürfen die Auswirkungen der KI auf die Gesellschaft nicht unterschätzt werden.

In diesem Zusammenhang kann es helfen, zwischen Ethik und Ethizität zu unterscheiden. Systeme mit Künstlicher Intelligenz sind mehr als nur Automation. Seit der Industrialisierung wurden immer wieder Prozesse automatisiert. Der springende Punkt der Künstlichen Intelligenz ist die *Automatisierung von Entscheidungen*, welche bisher der Mensch getroffen hat. Diese Entscheidungen werden durch die automatisierte Auswertung von Daten und den Vergleich verschiedener Alternativen anhand von Zielgrößen erarbeitet. Damit stellen sich ethische Fragen, sobald diese







Entscheidungen negative Auswirkungen auf Menschen haben. Wer ist verantwortlich? Obwohl nach Kirchschräger entscheidende Systeme nicht moralisch handeln können, also keine Ethik aufweisen, ist es möglich, die *Ethizität* ihrer Auswirkungen zu bewerten, inwiefern diese ethisch vertretbar sind. Unter dem Begriff der *Ethizität* versteht man also die Eigenschaft, als ethisch bewertet werden zu können.

### 3 Gradual Computational Ethics

Nach welchen Kriterien soll die Ethizität Künstlicher Intelligenz bewertet werden? Aufgrund der Meinungsfreiheit gibt es unterschiedlichste Vorstellungen von gutem Verhalten. Kirchschräger (2019) setzt auf die Menschenrechte, welche einen moralischen Relativismus aufnehmen und die Freiheit weit möglichst schützen. Die Menschenrechte sind universell akzeptiert, sie bewahren die Vielfalt der Meinungen und Kulturen und sind daher besonders geeignet, in einer globalisierten Welt als ethische Standards zu wirken. Kirchschräger bleibt bei seinen Betrachtungen allerdings auf der philosophischen Ebene haften. Er zeigt, wie viele Forschende auf dem Gebiet der digitalen Ethik, keine implementierbaren Mechanismen für sein Ethikmodell auf. Die Architekten der KI brauchen jedoch konkrete Anleitungen zur Implementierung der Ethizität intelligenter Systeme.

Computational Ethics ist ein Projekt, Ethik berechenbar zu machen, schreibt S. T. Segun (2021) von der Forschungsgruppe für Data and Computational Ethics der Philosophischen Fakultät an der Universität Johannesburg. Für Segun geht das über die theoretischen Grenzen der Maschinenethik hinaus und soll aufzeigen, wie ethische Intelligenzsysteme effektiv *umgesetzt und genutzt* werden könnten. Der Stand der Technik zur Computational Ethics wurde in diesem Schwerpunktthema in einem

**Tab. 1** Unterschied zwischen klassischer Ethizität und Ethicity by Differentiation, eigene Darstellung

	Ethicity by Black and White	Ethicity by Differentiation
		
Definition	Charakteristische Funktion $\chi_A(x): X \rightarrow \{0, 1\}$ mit $A \subset X$	Zugehörigkeitsfunktion $\mu_A(x): X \rightarrow [0, 1]$ mit $A \subset X$
Charakterisierung	 Sprunghaftigkeit	 Stetigkeit
Rechenvorgang	Hard Computing with DIGITS	Soft Computing with WORDS
Regelsystem	 Approximation	 Intuition
Ethizität	$\varepsilon_A(x)$ entspricht charakteristischer Funktion $\chi_A(x)$	$\varepsilon_A(x)$ entspricht einer Aggregation über Dimensionen I bis VI mit den Teilfunktionen $\varepsilon_A^+(x)$ und $\varepsilon_A^-(x)$

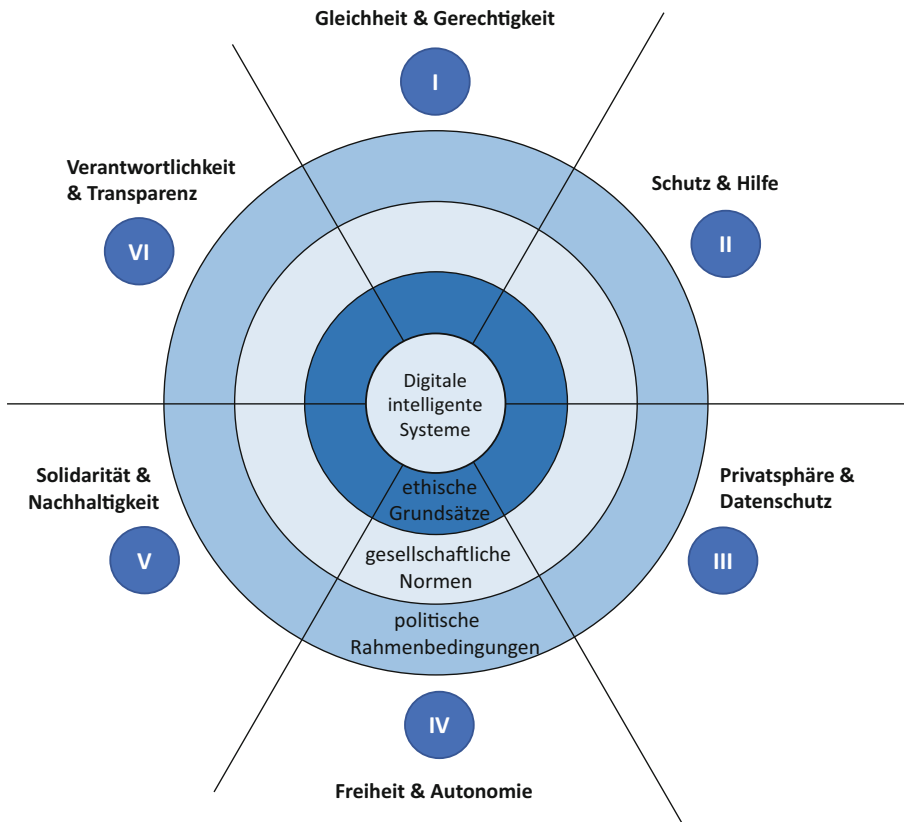
Überblicksartikel vorgestellt. Im Folgenden zeigen wir als Operationalisierung, wie ein Radarsystem die Bewertung der Ethizität berechenbar machen kann.

Mit welchen Methoden und Techniken lässt sich die Ethizität von Technologien vorantreiben? Wir sind der Meinung, dass die klassische Logik mit den beiden Wahrheitswerten richtig und falsch zu kurz greift und dass Logiken mit Differenzierungen notwendig sind (Bergmann 2008). Eine binäre Einteilung in ethische und unethische Technologien ist nicht zielführend. Die Bewertung der Ethizität von intelligenten Systemen sollte graduell und differenziert erfolgen, wie das Modell der Gradual Computational Ethics aus Abschn. 5, das auf dem sechsteiligen Radarsystem (Abschn. 4) basiert.

Der Lösungsansatz besteht darin, die klassische Logik mit den begrenzten Wahrheitswerten wahr und falsch durch differenziertere Logiken zu ersetzen. In Tab. 1 sind die wichtigsten Unterschiede zwischen Ethizität in Schwarz und Weiss sowie Ethizität in Grautönen zusammengefasst:

- **Definition – Differentiation matters:** Anstelle der charakteristischen Funktion  $\chi$ , die nur die Extremwerte richtig oder falsch zulässt, tritt die unscharfe Logik. Diese umfasst die Mengenzugehörigkeitsfunktion  $\mu$ , die für jedes Element der Menge den Grad der Zugehörigkeit zur Menge misst.
- **Charakterisierung – Analog statt digital:** Die Sprunghaftigkeit digitaler Ansätze lässt sich mit stetigen Verfahren vermeiden. Anstelle von Treppenfunktionen werden stetige Funktionen für die Steuerung verwendet, um rechnergesteuerte Geräte dem menschlichen Empfinden und Verhalten anzupassen (Fuzzy Control).
- **Rechenvorgang – Symbiose zwischen Hard und Soft Computing:** Hard Computing basiert auf exakten Fakten und statistischen Analysen. Soft Computing zielt darauf ab, menschliche Fähigkeiten und menschliches Verhalten nachzubilden. Ein Paradigmenwechsel steht an, Algorithmen des Hard Data Mining durch solche des Soft Data Mining zu ergänzen, falls Informationen lückenhaft, unpräzise oder widersprüchlich vorliegen (Meier 2021).
- **Regelsystem – Fließende intuitive Entscheidungen:** Anstelle herkömmlicher Entscheidungssysteme bieten sich regelbasierte Maschinen und Algorithmen an, die zum Soft Data Mining zählen wie Fuzzy Clustering, Evolutionäre Algorithmen, Künstliche Neuronale Netze oder Probabilistisches Schließen.
- **Ethizität – Explizite ethische Agenten:** Für kritische Anwendungen, die Fragen der Ethik und Moral betreffen, sollten explizite ethische Agenten implementiert werden, unter Nutzung des Modells für Gradual Computational Ethics (Abschn. 5).

Im Folgenden wird zuerst das Radarsystem ethischen Handelns mit sechs Dimensionen eingeführt (Abschn. 4), bevor eine Anleitung zur Implementierung des Radars für intelligente Informationssysteme anhand des Modells Gradual Computational Ethics in Abschn. 5 erfolgt.



**Abb. 1** Dimensionen ethischen Handelns im Radarsystem. (Angelehnt an Kirchschräger (2019) und Jobin et al. (2019))

#### 4 Radar ethischen Handelns

Radar ist die Abkürzung für Radio Detection and Ranging oder frei übersetzt mit Ortung und Abstandsmessung im Radiofrequenzbereich. Radargeräte sind Frühwarnsysteme zur Überwachung von Fahrzeugen, Schiffen oder Flugobjekten und dienen der Aufklärung. Im übertragenen Sinn ist ein Radarsystem für ethisches Handeln ein Orientierungs- und Controlling-Instrument für digitale Gesellschaften. Die Radarmethode (Abb. 1) vermittelt Kriterien zur multidimensionalen Bewertung der Ethizität eines digitalen intelligenten Systems. Für jedes Kriterium kann ein Eintrag gemacht werden, wie gut ein System oder Teilsystem es erfüllt (vgl. Abschn. 5). So kann das Radarbild visuell aufzeigen, in welchen Ethik-Dimensionen eine Lösung Stärken und Schwächen aufweist.

Welche Ethik-Kriterien sollen für digitale intelligente Systeme maßgebend sein? Kirchschräger (2019) suggeriert, Kriterien für ethische Digitalisierung und Künstliche Intelligenz auf der Grundlage der allgemeinen Erklärung der Menschenrechte zu entwickeln. Nach Kirchschräger sind die Menschenrechte ein geeigneter Stan-

**Tab. 2** Ableitung der Dimensionen des Ethik-Radars aufgrund der allgemeinen Erklärung der Menschenrechte nach Kirchschräger (2019) und den Kategorien der Ethik der Künstlichen Intelligenz nach Jobin et al. (2019)

Dimensionen	Ethische Prinzipien für die KI (Jobin et al. 2019)	Menschenrechte im Digitalzeitalter (Kirchschräger 2019; Ohchr.org 2021)
<i>I Gleichheit &amp; Gerechtigkeit</i>	Justice & Fairness	1. We are all Born Free and Equal 2. Freedom from Discrimination 6. Right to Recognition Before the Law 7. Right to Equality Before the Law
<i>II Schutz &amp; Hilfe</i>	Beneficence Non-Maleficence Dignity	3. Right to Life 4. Freedom from Slavery 5. Freedom from Torture 8. Right to Remedy 9. Freedom from Arbitrary Detention 10. Right to a Fair Trial 11. Presumption of Innocence 14. Right to Asylum 29. Duty to Your Community 30. Rights are Inalienable
<i>III Privatsphäre &amp; Datenschutz</i>	Privacy	12. Right to Privacy
<i>IV Freiheit &amp; Autonomie</i>	Freedom & Autonomy	1. We are all Born Free and Equal 13. Freedom to Movement 15. Right to Nationality 16. Rights to Marry and Found a Family 17. Right to Own Property 18. Freedom of Religion or Belief 19. Freedom of Opinion and Expression 20. Freedom of Assembly and Association 21. Democracy 28. Right to a Free and Fair World
<i>V Solidarität &amp; Nachhaltigkeit</i>	Sustainability Solidarity	22. Right to Social Security 23. Right to Work 24. Right to Rest and Leisure 25. Right to Adequate Standard of Living 26. Right to Education 27. Right to Cultural, Artistic and Scientific Life
<i>VI Verantwortlichkeit &amp; Transparenz</i>	Responsibility Trust Transparency	No corresponding human rights exist

pard, der Raum bietet für kulturelle Unterschiede. Weiter haben Jobin et al. (2019) bestehende Ethik-Frameworks für Künstliche Intelligenz analysiert, codiert und eingeordnet. Die Autoren kommen so auf 11 Kategorien. Wir haben die 30 Forderungen der allgemeinen Menschenrechte von Kirchschräger mit den Kategorien von Jobin et al. abgeglichen und auf sechs Dimensionen zusammengefasst (vgl. Tab. 2).



Die sechs Dimensionen sind die folgenden:

- **I Gleichheit & Gerechtigkeit:** Das Gleichstellungsrecht verlangt, dass Regeln für alle Menschen gleich gelten und alle Menschen vor Diskriminierung geschützt werden, sei es aufgrund von körperlichen, geistigen oder psychischen Einschränkungen, Religion, Politik, Geschlecht, Sexualität, Hautfarbe usw. Die Anwendung von Algorithmen zum KI-Profilung (Einstufung von Webnutzern) beispielsweise kann zu widersprüchlichen Entscheidungen oder zur Ungleichbehandlung von Nutzern führen. Auch ist es möglich, dass gewisse Vorurteile (Bias) durch Maschinenlernen aus den Daten übernommen werden. Wenn ein digitales intelligentes System Entscheidungen trifft, die Menschen Vor- oder Nachteile ermöglichen, dann muss sichergestellt werden, dass die Differenzierung aufgrund legitimer Attribute (z. B. Leistung) und nicht aufgrund politischer oder anderweitig diskriminierender Eigenschaften geschieht. Zudem können Differenzierungen graduell im Sinne der unscharfen Logik geschehen, was zum Beispiel im Kundenbeziehungsmanagement zu mehr Gleichbehandlung führt (Meier und Werro 2018).
- **II Schutz & Hilfe:** Die Schutzrechte verhindern, dass Menschen ausgebeutet oder verletzt werden. Sie beinhalten das Recht auf Leben und den Schutz vor Folter, Willkür und Sklaverei. Im Bereich der autonomen Waffensysteme stellen sich heute schon ethische Fragen, welche diese Schutzrechte tangieren. Auch vor autonomen Fahrzeugen müssen Menschen in Schutz genommen werden. Wenn ein digitales intelligentes System Entscheidungen trifft, welche Menschen gefährden, sind genaue ethische Abklärungen und Vorsichtsmaßnahmen geboten.
- **III Privatsphäre & Datenschutz:** Das Recht auf den Schutz der Privatsphäre ist ein Menschenrecht, das im digitalen Zeitalter zunehmend an Bedeutung gewinnt. Mithilfe von Big Data können umfangreiche Datenbestände aus unterschiedlichen Quellen kombiniert und persönliche Profile mit Algorithmen des Data Mining, des Soft Computing (vgl. Abschn. 3) respektive der Künstlichen Intelligenz erarbeitet werden. Datenschutzgesetze westlicher Demokratien verbieten es, ohne Zustimmung der Betroffenen Datenspuren über die Zeit anzulegen und auszuwerten. Wenn digitale intelligente Systeme Daten über Personen sammeln und verarbeiten, ist nicht nur aus ethischer Sicht, sondern auch aus rechtlicher Perspektive Vorsicht angebracht. Personenbezogene Daten dürfen nur mit ausdrücklicher Zustimmung und nur für den festgelegten Zweck bearbeitet werden.
- **IV Freiheit & Autonomie:** Zu Freiheit und Autonomie gehören die Meinungs-, Religions- und Bewegungsfreiheiten und die Eigentumsrechte. Beim automatisierten Fact-Checking und bei der digitalen Zensur muss die Meinungsfreiheit sichergestellt werden. Wenn Softwaresysteme Inhalte filtern, dürfen Webnutzer nicht aufgrund ihrer politischen, religiösen oder gesellschaftlichen Einstellung diskriminiert werden. Weiterhin ist das Schützen der Eigentumsrechte wie Copyright im Digitalzeitalter eine besondere Herausforderung. Digitale Daten und Produkte unterscheiden sich von physischen Dokumenten dadurch, dass sie einfach und schnell kopiert und verteilt werden können. Neben der Nutzung digitaler Wasserzeichen helfen kryptografische Verfahren oder Smart Contracts, Eigentumsverhältnisse weltweit eindeutig sicherzustellen.

- **V Solidarität & Nachhaltigkeit:** Die Solidarität bedeutet gegenseitige Hilfe und Rücksichtnahme. Wenn intelligente Systeme menschliche Arbeit automatisieren oder abwerten, braucht es entsprechende Abfederungen. Die Scheinselbstständigkeit im Sinne der Plattformökonomie ist nicht zulässig. Digitalsteuer, Transaktionssteuer oder Grundeinkommen könnten Möglichkeiten sein. Neben dem Recht auf Arbeit gibt es auch ein Recht auf Erholung und Ferien. Die ständige Erreichbarkeit im digitalen Zeitalter – zeit- wie standortbezogen – darf nicht dazu führen, dass die Lebensqualität zurückgeht. Es sollte möglich sein, dass sich der Einzelne jederzeit aus dem Cyberspace ausklinken kann. Zur Solidarität gehört aber auch der Umweltschutz. Unter dem Stichwort Green IT sollen sowohl für Informations- und Kommunikationstechnologien wie für Anwendungen der Umweltschutz sowie Aspekte der Ressourceneffizienz berücksichtigt werden.
- **VI Verantwortlichkeit & Transparenz:** Unter Verantwortung wird die Pflicht von Individuen oder Institutionen verstanden, für die möglichen Folgen einer Entscheidung oder einer Handlung einzustehen und Rechenschaft abzulegen. Im Digitalzeitalter steht im Fokus, Bedeutung und Wirkung digitaler Systeme und Künstlicher Intelligenz abzuschätzen und den Nutzern die Risiken für die Gesellschaft aufzeigen zu können. Dazu zählen Transparenz sowie das Recht auf Information und auf Wissen. Die Entscheidungen, welche digitale intelligente Systeme autonom treffen, müssen transparent und nachvollziehbar sein. Wenn diese Entscheidungen Menschen betreffen, sind erklärbare Auswertungsmethoden (eXplainable Artificial Intelligence) wie z.B. die unscharfe Logik ethisch angebracht.

In den letzten Jahren sind unzählige Frameworks für digitale Ethik aus Wissenschaft und Praxis vorgeschlagen worden. Diverse Regierungsstellen haben zudem mit Experten aus Informatik, Ethik, Soziologie oder Governance zusammengearbeitet und nationale Ethik-Leitfaden für das Digitalzeitalter erarbeitet und publiziert. Auch die Europäische Kommission diskutiert im Rahmen der Initiative European Data Protection Supervisor Fragen der digitalen Ethik. Zudem haben namhafte Wissenschaftlerinnen und Experten aus Belgien, Deutschland, England, Frankreich, Holland, Italien, Schweiz und aus den USA unter dem Lead des Digital Ethics Lab der Oxford Universität den Report „AI4People’s Ethical Framework for a Good AI Society: Opportunities, Risks, Principles, and Recommendations“ (Floridi et al. 2018) herausgegeben, u. a. unterstützt vom Atomium-European Institute for Science, Media and Democracy. Bei der Sichtung dieser nationalen, europäischen und internationalen Vorschläge für digitale Ethik fällt auf, dass diese sich neben Fragen des Datenschutzes und der Datensicherheit immer wieder auf die Erklärung der Menschenrechte berufen. Aus diesem Grunde haben wir uns auf die breit abgestützte Untersuchung von Jobin et al. (2019) mit den fünf ethischen Grundsätzen Transparenz, Gerechtigkeit und Fairness, Wohltätigkeit resp. Nicht-Schadensprinzip, Verantwortung und Datenschutz konzentriert und die Erkenntnisse mit den allgemeinen Menschenrechten für digitale Ethik nach Kirchschräger (2019) kombiniert.

Mit dem Radarsystem für Gradual Computational Ethics haben wir eine Methode zur Hand, mit der wir Schritt für Schritt die Ethizität intelligenter Systeme anhand des folgenden Modells bewerten können.

## 5 Modellierung einer Fuzzy Ethizität

In der klassischen Mengenlehre wird eine Menge  $A$  als eine Sammlung von wohlunterscheidbaren Elementen aus einer Universalmenge  $X$  (universe of discourse) definiert. Die Mengenzugehörigkeitsfunktion, welche die Elemente der Menge  $A$  bewertet, ist eine charakteristische Funktion  $\chi$ : Entweder gehört ein Element  $x$  zur Menge  $A$  oder es gehört nicht dazu. Etwas präziser formuliert: Falls  $X$  die Universalmenge ist und  $A$  eine Teilmenge von  $X$  ist, d.h.  $A \subset X$ , dann gibt es eine charakteristische Funktion  $\chi_A(x): X \rightarrow \{0, 1\}$  für alle Elemente  $x$  aus  $X$ . Mit anderen Worten: Jedes Element  $x$  aus  $X$  kriegt mithilfe der charakteristischen Funktion  $\chi$  den Wahrheitswert 1, falls  $x$  in der Menge  $A$  liegt, andernfalls 0. Anschaulich gesprochen stellt die klassische Logik eine Schwarzweiss-Landschaft dar mit den beiden Wahrheitswerten 0 für schwarz und 1 für weiss.

L. A. Zadeh (1965) charakterisiert die unscharfen Mengen (Fuzzy Sets) wie folgt: Ist  $X$  die Universalmenge und  $A$  eine Teilmenge davon, d.h.  $A \subset X$ , dann wird die unscharfe Menge  $\mathbf{A}$  als  $\mathbf{A} = \{(x, \mu_A(x)) | x \in X\}$  definiert. Die Mengenzugehörigkeitsfunktion  $\mu$  nimmt Werte auf dem Einheitsintervall an, d.h.  $\mu_A(x): X \rightarrow [0, 1]$  für alle  $x$  aus  $X$ . Eine unscharfe Menge  $\mathbf{A}$  ist demnach eine Menge von Paaren der Form  $(x, \mu_A(x))$ , wobei zu jedem Element  $x$  auch sein Zugehörigkeitsgrad  $\mu_A(x)$  zur Menge  $\mathbf{A}$  mitgeliefert wird.

Das Zugehörigkeitsmaß  $\mu$  schliesst 1 für wahr und 0 für falsch mit ein, lässt aber zusätzlich alle Wahrheitswerte zwischen 0 und 1 zu. Der Grad  $\mu_A(x) = 1$  bedeutet volle Mengenzugehörigkeit; das heisst,  $x$  gehört zu 100 % zur Menge  $\mathbf{A}$ . Der Wert 0 drückt aus, dass das Element  $x$  nicht zur Menge  $\mathbf{A}$  gehört. Ein Wert 0,8 bedeutet, dass das entsprechende Element  $x$  zu 80 % in der Menge  $\mathbf{A}$  liegt. Mit anderen Worten wird die Schwarzweiss-Landschaft um unendlich viele Grauwerte zwischen schwarz und weiss erweitert.

Damit bei der Nutzung eines Informationssystems die Anwender sich nicht mit mathematischen Formeln oder unscharfen Operatoren herumschlagen müssen, werden unscharfe Konzepte mit linguistischen Variablen und Termen aus dem Sprachgebrauch der Anwender benutzt. Variablen, deren Ausprägungen sprachliche Grössen betreffen, werden als linguistische Variablen bezeichnet. Hinter linguistischen Variablen und Termen stehen unscharfe Mengen, die die linguistischen Konzepte im Detail definieren. So kann als Beispiel die Körpergrösse als linguistische Variable mit den Termen „klein“, „mittelgross“ und „gross“ aufgefasst und mit Mengenzugehörigkeitsfunktionen hinterlegt werden.

Allgemein eignet sich die unscharfe Logik für das Rechnen mit Worten (Computing with Words). Neben der Verwendung von linguistischen Variablen und Termen können Informationssysteme genutzt werden, bei denen die Anwender in natürlicher Sprache mit der Maschine interagieren. Verschiedene Wissenschaftler auf dem Gebiet des Soft Computing haben das Potenzial zum Rechnen mit Worten in einem Beitrag zusammengefasst (Mendel et al. 2010).

Dementsprechend können wir die ethischen Kriterien aus Abschnitt 4 als linguistische Variablen modellieren. Das Modell der Ethizität besteht allgemein aus den sechs unscharfen Dimensionen (linguistischen Variablen)  $D_1$  bis  $D_{VI}$  des ethischen Radarsystems aus Abb. 1, den entsprechenden Zugehörigkeitsfunktionen  $\mu_1$  bis  $\mu_{VI}$

und einer Aggregationsfunktion  $\bigwedge$  (beispielsweise dem  $\gamma$ -Operator<sup>1</sup>) der unscharfen Logik, um die Zugehörigkeitsgrade zu kombinieren. Die Ethizität  $\varepsilon$  wird als Aggregation aller Mengenzugehörigkeitswerte der sechs Radar-Dimensionen zwischen 0 und 1 berechnet, wobei 0 als völlig unethisch, 1 als ethisch optimal und 0,5 als neutral gilt:

$$\varepsilon(x) = \bigwedge_{i \in \{I, II, III, IV, V, VI\}} \mu_i(x)$$

Wie soll diese Kombination berechnet werden? Die klassische Konjunktion der unscharfen Logik, das Fuzzy AND, ist der Minimum-Operator. Eine solche Kombination würde zwar sensitiv auf einzelne als unethisch bewertete Dimensionen reagieren. Allerdings würden damit alle anderen Bemühungen nicht berücksichtigt, denn es spielt bei der Minimum-Aggregation keine Rolle, wie gut andere Dimensionen bewertet würden.

Eine weitere Möglichkeit wäre eine Durchschnittbildung. Hier kann es aber vorkommen, dass sich gut bewertete Dimensionen mit den schlecht bewerteten Dimensionen gegenseitig auslöschen. Dasselbe gilt für andere kompensatorische Aggregationen wie den  $\gamma$ -Operator. Daher denken wir, es wäre sinnvoll, sowohl die positive als auch die negative Ethizität zu modellieren, und zwar in zwei separaten Kennzahlen  $\varepsilon^+$  bzw.  $\varepsilon^-$ .

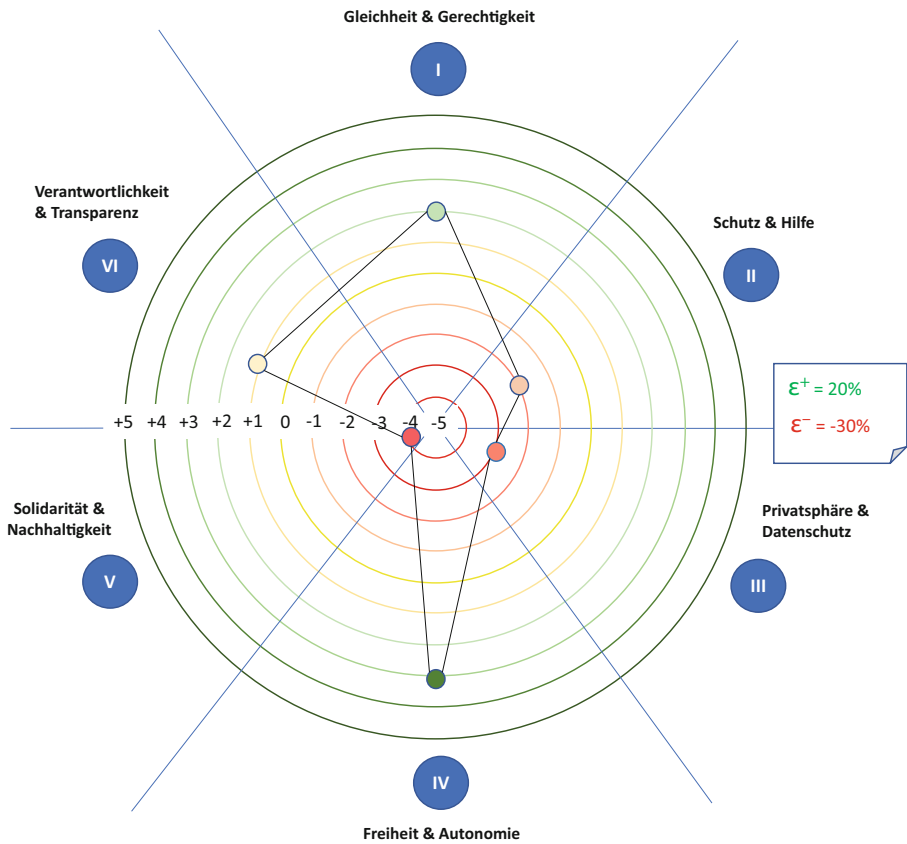
Die beiden Ethizitäten  $\varepsilon^+$  bzw.  $\varepsilon^-$  sind wie folgt definiert:  $\varepsilon^+$  ist der Durchschnitt der ethisch positiv bewerteten Kriterien ( $\mu > 0,5$ ), und  $\varepsilon^-$  ist der Durchschnitt der ethisch negativ ( $\mu < 0,5$ ) bewerteten Kriterien.

$$\varepsilon^+(x) = \frac{1}{6} \sum_{i | \mu_i(x) > 0.5} 2(\mu_i(x) - 0.5) \quad \varepsilon^-(x) = -\frac{1}{6} \sum_{k | \mu_k(x) < 0.5} 2\mu_k(x)$$

Um nicht zu tief in die Berechnungsvielfalt der unscharfen Logik einzutauchen, verwenden wir in diesem Beitrag für den Anwender im Radar anstelle des Kontinuums ein vereinfachtes Scoringmodell mit 6 Dimensionen samt Bewertungsraster von -5 bis +5. Die einzelnen Kreissegmente (siehe Abb. 2) bilden in den äußeren Ringen den grünen Bereich (+5, +4, +3, +2 und +1), der mittlere gelbe Kreis ist neutral gesetzt mit dem Wert 0, die innersten roten Bereiche der Ampel haben die Werte (-1, -2, -3, -4 und -5).

Betrachten wir nun das vereinfachende Scoringmodell aus Abb. 2 zur Berechnung der Ethizität anhand eines fiktiven Beispiels. Die drei Dimensionen I (Gleichheit & Gerechtigkeit), IV (Freiheit & Autonomie) und VI (Verantwortlichkeit & Transparenz) liegen bei der Bewertung im grünen Bereich. Damit ergibt sich eine Ethizität von 20%, da  $\varepsilon^+ = 1/6 * (3/5 + 2/5 + 1/5) = 6/30 = 0,2$ .

<sup>1</sup> Der sogenannte  $\gamma$ -Operator der unscharfen Logik entspricht einem kompensatorischen UND und wurde empirisch getestet als sinnvoll erachtet. Er berechnet ein ausgewogenes Mittelmaß zwischen den unterschiedlichen Bewertungsachsen (hier Dimensionen I bis VI des Radars), wobei die Ausgewogenheit mit der Wahl des  $\gamma$ -Wertes zwischen 0 und 1 eingestellt werden kann. Für Interessierte verweisen wir auf die weiterführenden Literaturquellen, insbesondere Zimmermann und Zysno (1980).



**Abb. 2** Modell zur Berechnung der Ethizität im Radarsystem, eigene Darstellung

Entsprechend berechnet sich der rote Bereich der Dimensionen II, III und V auf  $\varepsilon^- = 1/6 * ((-4/5) + (-3/5) + (-2/5)) = -9/30 = -0,3$ .

Ein konkretes Anwendungsbeispiel zur Ethizität eines Privacy Setting Framework für politische Partizipation wird im nächsten Abschnitt ausführlicher diskutiert.

## 6 Fallbeispiel Privacy Setting Framework für politische Partizipation

Nun soll das ethische Radarsystems anhand eines Anwendungsbeispiels vertieft werden, nämlich für den Schutz der Privatsphäre bei politischer Partizipation.

### 6.1 Privacy Setting Framework

Beim Betreiben eines eGovernment-Portals mit elektronischen Diensten für Bürgerinnen und Bürgern kommt dem Schutz der Privatsphäre ein besonderes Augenmerk zu. Wichtig ist, dass der Citizen differenziert seine Schutzbedürfnisse einstellen und jederzeit ändern kann. Dazu wird ein fuzzy-based Privacy Setting Framework vor-

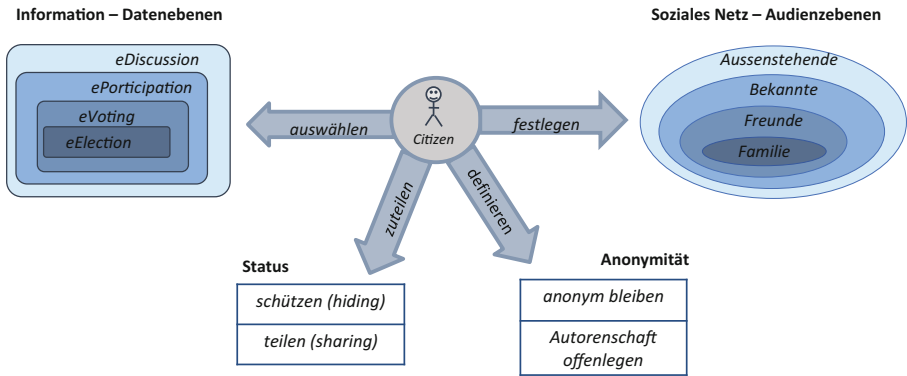


Abb. 3 Privacy Setting Framework für politische Partizipation. (Angelehnt an Kaskina und Meier 2016)

geschlagen, das an der Universität Fribourg entwickelt und erprobt wurde (Kaskina and Meier 2016; Kaskina 2022).

Das Privacy Setting Framework aus Abb. 3 besteht aus Teilbereichen, die vom Citizen festgelegt werden müssen:

- **Information:** Der Teilbereich Information unterscheidet vier Datenebenen. Unter eDiscussion werden alle Diskussionsbeiträge des jeweiligen Citizen zusammengefasst. Die Datenebene eParticipation umfasst alle Handlungen, die der Bürger für die Allgemeinheit erbracht hat. Im Gefäß eVoting werden die persönlichen Abstimmungen des Citizen aufbewahrt und eElection sammelt seine Stimmabgaben bei politischen Wahlen.
- **Audienz:** Neben dem Teilbereich Information werden im Teilbereich Audienz die vier möglichen Anspruchsgruppen des sozialen Netzes festgelegt, mit denen der Citizen einen Informationsaustausch pflegen möchte. Familie bedeutet, dass der Citizen mit ausgewählten Familienmitgliedern seine politische Partizipation teilen möchte. Entsprechend werden Anspruchsgruppen für Freunde, Bekannte und Aussenstehende festgelegt.
- **Status und Anonymität:** Zwei wichtige Entscheidungsboxen ergänzen das Privacy Setting Framework: Mit der Statusbox legt der Citizen fest, ob er eine bestimmte Information für sich behalten (hiding) oder teilen (sharing) möchte. Bei der Anonymitätsbox geht es darum, ob der Citizen die Information anonym oder mit seiner Autorenschaft weitergeben möchte.

Kurz zusammengefasst könnte ein möglicher Vektor für den Schutz der Privatsphäre eines Benutzers A wie folgt lauten (Datenebene: eElection, Audienzebene: Freunde, Status: Sharing, Anonymität: mit Autorenschaft). Dieser Beispielvektor sagt aus, dass der Benutzer A seine Mandatswahl ausgewählten Freunden (und damit auch ausgewählten Familienmitgliedern) mit seiner Autorenschaft weitergeben möchte.

Das hier vorgestellte Framework ist mit Algorithmen der unscharfen Logik (fuzzy c-means) implementiert worden, um mehrdimensionale Cluster aus Präferenzen von

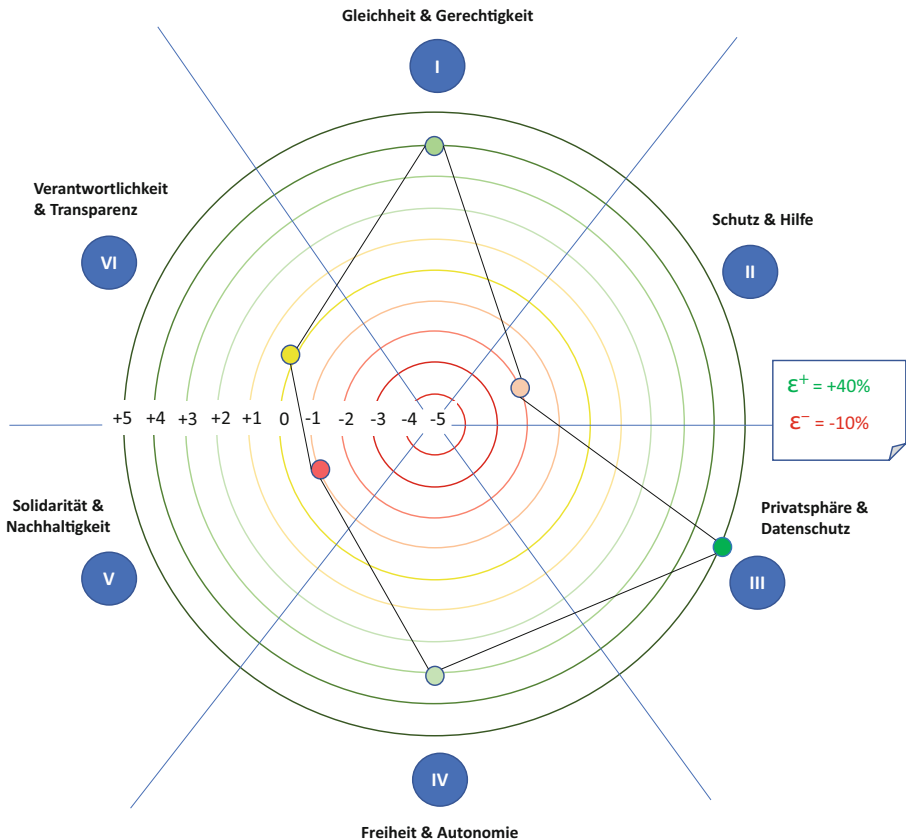
Citizen bilden zu können (Kaskina 2022). Darauf aufbauend lassen sich Empfehlungssysteme für die Wahl politischer Mandatsträger realisieren.

Das Privacy Setting Framework der Abb. 3 samt dem darauf entwickelten Empfehlungssystem wurde für die Präsidentschaftswahl in Ecuador (erste Wahl am 19. Februar 2017 und Stichwahl am 2. April 2017) verwendet und ausgewertet. Weiterführende Diskussionen finden sich in der Monografie von Kaskina (2022).

## 6.2 Gradual Computational Ethics für Privacy Setting Framework

Das Privacy Setting Framework für politische Partizipation wird nun einem Ethiktest unterworfen, unter Anwendung des Radarsystems ethischen Handelns mit den sechs Dimensionen. Der Einfachheit halber werden die Bewertungen der Dimensionen jeweils auf den ganzzahligen Kreisen gesetzt, damit der Berechnungsmechanismus für die einzelnen Indexkomponenten der Ethizität nachvollzogen werden kann.

Die Dimension I (Gleichheit & Gerechtigkeit) liegt im grünen Bereich (Bewertung +4), da der Schutz der Privatsphäre die Gleichberechtigung fördert. Auch die



**Abb. 4** Fallbeispiel Gradual Computational Ethics für Privacy Setting Framework politischer Partizipation

Dimension III Privatsphäre & Datenschutz wird mit Bewertung +5 als äusserst positiv angesehen, da das Framework sowohl den Datenschutz erhöht und die Privatsphäre nach Wunsch der Anwender schützt. Die Dimension IV (Freiheit & Autonomie) wird ebenfalls positiv mit +3 bewertet, da das Framework die Meinungsfreiheit und Vielfalt unterstützt. Damit resultiert eine positive Ethizität für die Dimensionen I, III und IV von 40% resp.  $\varepsilon^+ = 1/6 * (4/5 + 5/5 + 3/5) = 12/30 = 0,4$ .

Bei der Verantwortlichkeit und Transparenz der Dimension VI wird eine neutrale Einschätzung vorgenommen, d. h. die Ethizität dieser Dimension kriegt den Wert 0.

Die beiden übrigen Dimensionen II (Schutz & Hilfe) und V (Solidarität & Nachhaltigkeit) liegen beide im roten Bereich. Ein Grund liegt darin, dass die zehn in der Tab. 2 aufgelisteten Schutzbedürfnisse wie Recht auf Leben, keine Sklaverei oder Recht auf Asyl vom Privacy Setting Framework für politische Partizipation zwar nicht betroffen sind, aber im Sinne der Informationssicherheit bei Daten-Leaks oder Cyberangriffen betroffen sein könnten, wenn die Informationen in falsche Hände geraten. Insgesamt resultiert für die Dimension II ein Wert von -2. Der Wert bei der Solidarität und Nachhaltigkeit (Dimension 5) liegt mit -1 ebenfalls im negativen Bereich, da die Lösung nicht CO<sub>2</sub>-neutrale Energie verbraucht. Damit resultiert eine negative Ethizität für die Dimensionen II und V von -10% resp.  $\varepsilon^- = 1/6 * ((-2/5) + (-1/5)) = -3/30 = -0,1$  Die resultierende Visualisierung (Abb. 4) stellt diese Bewertungen für den Anwender leicht erfassbar dar und ermöglicht entsprechende Massnahmen.

## 7 Fazit

Es wurde bereits viel über Vorschriften diskutiert, um KI zu regulieren (Ryan und Stahl 2021). Wir sind der Meinung, dass ethische Digitalisierung nach dem Bottom-Up-Prinzip der Softwareentwicklung vorangetrieben werden sollte. Dies bedeutet, dass abgrenzbare Teilprobleme vorerst gelöst werden, mit deren Hilfe umfassendere Probleme angegangen werden. Die einzelnen Teillösungen werden von unten nach oben zusammengesetzt, bis das Gesamtproblem zufriedenstellend gelöst ist.

Das vorgestellte Radarsystem für Gradual Computational Ethics unterstützt den Bottom-Up-Approach auf ideale Art und Weise: Teilkomponenten eines Systems lassen sich mithilfe einer Radarbewertung verifizieren. Bei kritischen Ausschlägen bezüglich Ethizität müssten Nachbesserungen in der Software resp. bei den Algorithmen vorgenommen werden, um sukzessive die gewünschte Qualität betreffend Ethizität fürs Gesamtsystem zu erreichen. Mit anderen Worten: Das Radarsystem für Gradual Computational Ethics ist ein bedeutender und flexibler Baukasten fürs Software Engineering betreffend Ethizität.

Mit dem vorgestellten unscharfen Raster zur Ethizität lassen sich intelligente Systeme, Entscheidungsalgorithmen, Geschäftsprozesse mit entsprechenden Dienstleistungen, Empfehlungssysteme, Roboteranwendungen etc. beurteilen. Einzelne Dimensionen des Rasters legen dar, wo die Stärken bezüglich Ethizität ( $\varepsilon^+$ ) und wo die Schwächen ( $\varepsilon^-$ ) zu verzeichnen sind. Mit diesen Kennzahlen können einzelne Anwendungen oder Systemkomponenten untereinander verglichen werden. Zudem: Bei der Evaluation von Softwareprodukten kann der Radar für Gradual Compu-



tational Ethics Hinweise geben, welche der Produkte die Ansprüche an Ethizität künftiger Anwender oder Anspruchsgruppen am besten abdeckt.

In diesem Artikel haben wir Kirchschrägers Vorschlag, die Menschenrechte als Basis für eine digitale Ethik mit einer Radar-Methode für Anwender operationalisiert. Dabei haben wir Lücken festgestellt. Für so zentrale Themen wie die Nachhaltigkeit oder Transparenz gibt es keine Menschenrechte. Daher decken die Menschenrechte die Anforderungen an eine Ethik der KI aus heutiger Sicht (vgl. Jobin et al. 2019) nur teilweise ab.

Wir konnten argumentieren, dass die unscharfe Logik einen wesentlichen Beitrag zu ethischeren Systemen leisten kann. Erstens führt eine graduelle Bewertung beispielsweise in der Dimension Gleichheit & Gerechtigkeit zu weniger willkürlichen Diskriminierungen. Denn Gleiches wird mit etwas Toleranz tendenziell gleicher behandelt als mit einer Schwarz-Weiss-Bewertung. Zweitens sind in der Dimension Verantwortlichkeit & Transparenz die Algorithmen der unscharfen Logik oft erklärbarer als Black-Box-Verfahren, da die Zugehörigkeitsfunktionen visuell die Bedeutung einer Bewertung darstellbar und interpretierbar machen. Drittens ist die Bewertung der Ethizität selbst graduell. Hier liegt denn auch der springende Punkt unseres Beitrags, denn Gradual Computational Ethics vereint zwei wichtige Fragen der digitalen Ethik: fließende Bewertungen sowie berechenbare und implementierbare Bewertungssysteme.

Bei der Definition der vorgeschlagenen Radar-Methode stießen wir auf weiterführende Forschungsfragen. Wie beeinflusst die Wahl und die Gestaltung der Berechnungsmethoden die Ethizität der Resultate, die von der Methode selbst erzeugt werden? Beispielsweise hat die Wahl der Aggregationsfunktion einen grossen Einfluss auf die Bewertung. Können unethische Eigenschaften mit ethischeren Eigenschaften kompensiert werden? Da wir keine Ethiker sind, wäre es wohl sinnvoll, diese Forschungsfragen mit Ethikwissenschaftlerinnen zu vertiefen. Es wäre wünschenswert, weitere unscharfe Aggregationsfunktionen für unser Modell heranzuziehen und in einem Experiment die Eignung derselben zu prüfen. Zudem haben wir das Radar-Modell aufgrund eigener Annahmen als ersten Entwurf gestaltet. Wir haben aber noch keinen Feedback von Endanwendern eingeholt. Dies wäre sicher der nächste Schritt zur Ausgestaltung eines praxistauglichen Bewertungssystems.

**Danksagung** Gerne möchten wir uns bei Kollegen bedanken, die unseren Vorschlag über Ethizität intelligenter Systeme kritisch kommentiert haben: Michael Mörike von der integrata Stiftung in Tübingen und Fabrice Tschudi von der APP Unternehmensberatung AG in Bern. Wir erhielten kreative Impulse von Edy Portmann und Benjamin Emmenegger vom Institut HUMAN\_IST der Universität Fribourg betreffend digitaler Ethik. Zudem haben wir Gutachten erhalten und bedanken uns für die detaillierten und differenzierten Verbesserungsvorschläge, die wir weitgehend berücksichtigen konnten.

**Funding** Open access funding provided by Lucerne University of Applied Sciences and Arts

**Open Access** Dieser Artikel wird unter der Creative Commons Namensnennung 4.0 International Lizenz veröffentlicht, welche die Nutzung, Vervielfältigung, Bearbeitung, Verbreitung und Wiedergabe in jeglichem Medium und Format erlaubt, sofern Sie den/die ursprünglichen Autor(en) und die Quelle ordnungsgemäß nennen, einen Link zur Creative Commons Lizenz beifügen und angeben, ob Änderungen vorgenommen wurden.

Die in diesem Artikel enthaltenen Bilder und sonstiges Drittmaterial unterliegen ebenfalls der genannten Creative Commons Lizenz, sofern sich aus der Abbildungslegende nichts anderes ergibt. Sofern das betref-

fende Material nicht unter der genannten Creative Commons Lizenz steht und die betreffende Handlung nicht nach gesetzlichen Vorschriften erlaubt ist, ist für die oben aufgeführten Weiterverwendungen des Materials die Einwilligung des jeweiligen Rechteinhabers einzuholen.

Weitere Details zur Lizenz entnehmen Sie bitte der Lizenzinformation auf <http://creativecommons.org/licenses/by/4.0/deed.de>.

## Literatur

- Bergmann M (2008) An introduction to many-valued and fuzzy logic—semantics, algebras, and derivation systems. Cambridge University Press, New York
- Fieser J (2021) Ethics. Internet encyclopedia of philosophy
- Florida L et al (2018) AI4People's ethical framework for a good AI society: opportunities, risks, principles, and recommendations. [https://www.eismd.eu/wp-content/uploads/2019/11/AI4People's-Ethical-Framework-for-a-Good-AI-Society\\_compressed.pdf](https://www.eismd.eu/wp-content/uploads/2019/11/AI4People's-Ethical-Framework-for-a-Good-AI-Society_compressed.pdf). Zugegriffen: 9. Dez. 2021
- Grimm P, Keber TO, Zöllner O (Hrsg) (2021) Digitale Ethik: Leben in vernetzten Welten. Reclam, Ditzingen
- Jobin A, Ienca M, Vayena E (2019) Artificial Intelligence: the global landscape of ethics guidelines. *Nat Mach Intell* 1:389–399. <https://doi.org/10.1038/s42256-019-0088-2>
- Kaskina A (2022) Citizen privacy framework—case of a fuzzy-based recommender system for political participation. Springer, Heidelberg
- Kaskina A, Meier A (2016) Integrating privacy and trust in voting advice applications. Third International Conference on eDemocracy eGovernment (ICEDEG), Universidad de las Fuerzas Armadas ESPE, Sangolqui, Ecuador, 30.03.2016-01.04.2016, S 20–25
- Kirchschläger PG (2019) Digital transformation of society and economy—ethical considerations from a human rights perspective. *IJHRCS* 6:301. <https://doi.org/10.1504/IJHRCS.2019.102483>
- Meier A (2021) Rundgang big data Analytics—hard & soft data mining. In: D'Onofrio S, Meier A (Hrsg) Big Data Analytics – Grundlagen, Fallbeispiele und Nutzungspotenziale. Edition HMD. Springer, Heidelberg, S 3–23
- Meier A, Werro N (2018) Unschärfes Portfolio Management – Nutzenpotenziale. In: Meier A, Seising R (Hrsg) Vague Information Processing. Praxis der Wirtschaftsinformatik. HMD Band 55, Heft 3, S 528–539 <https://doi.org/10.1365/s40702-018-0394-x>
- Mendel JM et al (2010) What computing with words means to me. *IEEE Comput Intell Mag*. <https://doi.org/10.1109/MCI.2009.934561>
- Misselhorn C (2019) Grundfragen der Maschinenethik. Reclam, Ditzingen
- Moor JH (2006) The nature, importance, and difficulty of machine ethics. *IEEE Intell Syst* 21:18–21. <https://doi.org/10.1109/MIS.2006.80>
- Ryan M, Stahl BC (2021) Artificial intelligence ethics guidelines for developers and users: clarifying their content and normative implications. *J Inf Commun Ethics Soc* 19(1):61–86. <https://doi.org/10.1108/JICES-12-2019-0138>
- Segun ST (2021) From machine ethics to computational ethics. *AI Soc* 36:263–276. <https://doi.org/10.1007/s00146-020-01010-1>
- Spiekermann S (2019) Digitale Ethik – Ein Wertesystem für das 21. Jahrhundert. Droemer, München
- United Nations Human Rights Office of the High Commissioner OHCHR (2021) 30 articles on the 30 Articles of the Universal Declaration of Human Rights. <https://www.ohchr.org/EN/NewsEvents/Pages/DisplayNews.aspx?NewsID=23871>. Zugegriffen: 31. Aug. 2021
- Zadeh LA (1965) Fuzzy sets. *Inf Control* 8:338–353
- Zimmermann H-J, Zysno P (1980) Latent connectives in human decision making. *Fuzzy Sets Syst* 4:37–51