

9 Ethisches Gestalten

Bisher wurde in dieser Arbeit eine deontologische Risikoethik entwickelt, mit deren Hilfe die Zulässigkeit von Risikoauferlegung bewertet werden kann. Daraufhin wurden die beiden Risiken, die in Datenökosystemen entstehen, genau betrachtet, um herauszufinden, unter welchen Bedingungen diese Risiken entstehen und welche Pflichten daraus erwachsen. Daraus ließen sich administrative Anforderungen ableiten, welche erfüllt werden sollten, um vor den Risiken schützen zu können und die Auferlegung der Risiken ethisch zulässig zu machen. Es wurde außerdem klar, dass Datengenerierende und Datenverarbeitende verantwortlich sind für die ethische Gestaltung von Datenökosystemen, aber es stellt sich die Frage, wie sich die bisherigen ethischen Überlegungen in die Praxis umsetzen lassen. Angenommen ein Unternehmen möchte ein Datenökosystem entwickeln, wie kann es die ethischen Überlegungen aus dieser Arbeit in die Entwicklung einfließen lassen?

Da die Entwicklung eines Datenökosystems sehr umfangreich ist, wird diese im Folgenden in ihre verschiedenen Verarbeitungsschritte zerlegt betrachtet. Diese verschiedenen Schritte eines Datenökosystems sind das Sammeln, Speichern und Verarbeiten von Daten, sowie das Bereitstellen ihrer Ergebnisse. Jeder dieser Schritte bedarf ein Zusammenspiel zwischen sozialen Akteuren und der von ihnen genutzten Technologie. Die Technologie ist es, die es den menschlichen Akteuren ermöglicht, Daten zu sammeln, zu speichern, zu verarbeiten und Ergebnisse bereitzustellen. Auf Grund potenzieller Gefahren durch den Einsatz von datenverarbeitenden Technologien sowie auf Grund von deren Auswirkungen auf moralische Handlungen und Entscheidungen haben Technologien und ihre Verwendung eine moralische Signifikanz.¹¹⁶³ Mit Hilfe der Ethik konnten in dieser Arbeit Antworten gefunden werden auf die Frage danach, unter welchen

1163 Kiran et al. 2015:10

Umständen es zulässig ist, diese verschiedenen Schritte durchzuführen.

So gibt die in dieser Arbeit entwickelte deontologische Risikoethik eine Antwort auf die Frage, wie Datenökosysteme gestaltet werden sollten – nämlich als gerechtes soziales System der Risikobereitschaft von dem alle profitieren und in dem niemand als bloßes Mittel gebraucht wird. Was genau das in der Praxis bedeutet, kann die Ethik allein nicht beantworten. Die KI-Ethikerin Anaïs Rességuier und die Rechtswissenschaftlerin Rowena Rodrigues stimmen zu: „The objective of ethics itself is not to impose particular behaviours and to ensure these are complied with.“¹¹⁶⁴ Die Umsetzung ethischer Überlegungen ist an empirische Gegebenheiten gebunden, welche die Ethik selbst nicht antizipieren kann. Deshalb reicht es nicht, allein eine ethische Antwort auf Gestaltungsfragen zu haben, sondern es bedarf einer Methode zur Umsetzung, damit diese ethische Antwort in der Praxis Anklang findet.

Um zu verstehen, in welchem organisatorischen Kontext eine solche Umsetzungsmethode zur Anwendung kommt, wird im Folgenden beleuchtet, wie technologische Gestaltungsprozesse in Organisationen funktionieren. Das heißt, es wird beleuchtet wie die Akteure mit Kontrolle über das Vorgehen im Datenökosystem – Datengenerierende und Datenverarbeitende – im Allgemeinen organisatorisch strukturiert sind. Der deutsche Ethiker und Anthropologe Arne Manzeschke definiert Organisationen als gesellschaftliche Intermediäre: „In ihnen wird ein Großteil der Entscheidungen verfertigt, die das Leben des Einzelnen betreffen und das Zusammenleben in der Gesellschaft strukturieren.“¹¹⁶⁵ In einer Organisation gibt es jedoch auch Entscheidungsprobleme, weil viele verschiedene Interessen in Einklang gebracht werden müssen. Manzeschke identifiziert hier mehrere Konfliktlinien: „Die Interessen und Einschätzungen der verschiedenen fachlichen Professionen, die fachlichen und ökonomischen Prioritäten und eng damit verkoppelt die technischen Möglichkeiten, das Interesse des Individuums auf Berücksichtigung als besonderer „Einzelfall“ und das Interesse der Organisation, möglichst reibungsfrei die verschiedenen Fallgruppen nach einem Stan-

1164 Rességuier, Rodrigues 2020:2

1165 Manzeschke 2021:220

dard abzuarbeiten.“¹¹⁶⁶ Diese Konfliktlinien erschweren auch schon ohne ethische Überlegungen Entscheidungen in einer Organisation. Nach Ansicht des Soziologen Dirk Baecker erzeugt eine getroffene Entscheidung dann aber Gewissheit und Bestimmtheit, obwohl viele Unsicherheiten und Widersprüche bestehen.¹¹⁶⁷

Laut Manzeschke ist es Aufgabe der von ihm so genannten „Organisationsethik“, mit Hilfe ethischer Reflexion Moral in Entscheidungsprozesse von Organisationen einzubinden.¹¹⁶⁸ Ethik in einer Organisation habe die Aufgabe, die vielfältigen moralischen Auffassungen und die unterschiedlichen Arten der Entscheidungsfindung der Akteure zusammenzuführen und sie „(...) in einer komplexitätsadäquaten Theorie [zu] konzeptualisieren.“¹¹⁶⁹ Manzeschke bezeichnet dabei „Entscheidungen als Zielpunkt ethischer Bemühungen (...)“.¹¹⁷⁰ Weil ethische Fragen nie ohne Widerspruch abschließend geklärt werden können, kann eine organisatorische Entscheidung eine gewisse Bestimmtheit bieten.

An dieser Stelle stellt sich die Frage, auf welcher Organisationsebene die ethischen Fragen verhandelt werden sollten. Dem Verhaltensökonom und Ethiker Jan Gogoll und seinen Kolleg:innen zufolge sind viele Probleme, die scheinbar auf die Software zurückzuführen sind, in Wirklichkeit das Ergebnis von Geschäftsmodellen oder politischen, rechtlichen und kulturellen Bedingungen. Deshalb sollten ethische Herausforderung nicht nur auf der Ebene der Entwickler:innen behandelt werden. Sondern es müsse schon im Bereich der Unternehmensethik die ethische Frage geklärt werden, ob ein bestimmtes technisches Produkt überhaupt entwickelt werden sollte.¹¹⁷¹

Abbildung 4 veranschaulicht die verschiedenen Einflüsse und Ebenen der ethischen Entscheidungsfindung bei der Entwicklung einer Software. Laut Gogoll et al. ist jede Organisation eingebettet in ein Geflecht aus gesellschaftlichen Erwartungen, rechtlichen Anforderungen und kulturellen Normen. Sobald feststeht, dass ein ge-

1166 Manzeschke 2021:228

1167 Baecker 1999:146

1168 Manzeschke 2021:220

1169 Ebd.:228f.

1170 Ebd.:224

1171 Ebd.:1086

plantes Produkt legal ist und es einen gesellschaftlichen Konsens bezüglich des Produktes gibt, kann die Geschäftsführung eines Unternehmens als Strategisches Management beschließen, das Produkt herzustellen. Als Nächstes wird ein Projektteam innerhalb des Unternehmens über die genauen Spezifikationen des Produkts entscheiden – dies ist die Phase der Konzeptentwicklung. In dieser Phase werden die Anforderungen an das Produkt definiert. Dabei werden auch ethische Fragen besprochen und entschieden. Dann erst beginnt der Entwicklungsprozess. Hier nehmen die Teams von Entwickler:innen Einfluss auf die Gestaltung des Produkts, aber nur innerhalb der engen Grenzen der Spezifikationen, die in der Phase der Konzeptentwicklung festgelegt wurden: „Naturally, these parameters will never be completely determined. Therefore, the development team has some leeway in the development of the product.”¹¹⁷² Selbst scheinbar harmlose Designentscheidungen können große Auswirkungen haben. So kommt es zum Beispiel vor, dass Familiennamen von Software als ungültig eingestuft werden, weil sie einen nicht vorgesehenen Bindestrich enthalten oder weil der Name nur aus einem Buchstaben besteht, aber mindestens zwei erwartet werden. Durch solche scheinbar unwichtigen Vorgaben definieren Programmierende, was als Name anerkannt wird und was nicht.

Gogoll et al. beschreiben diese verschiedenen Stufen der ethischen Entscheidungsfindung anhand der Entwicklung eines Pflegeroboters. Zuerst stellt die Gesellschaft in einem politischen Rahmen die Frage, ob ältere Menschen durch Roboter gepflegt werden sollten. Die Antwort auf die Frage wird in Form eines rechtlichen Rahmens oder in Form gesellschaftlicher und kultureller Bedingungen festgehalten. Als nächstes fragt ein konkretes Unternehmen im Bereich des Strategischen Managements, ob sie diejenigen sein wollen, die so einen Pflegeroboter bauen und mit welchem Geschäftsmodell. Eine positive Antwort resultiert in einem Projekt, das gestartet wird. Dann stellt das Projektteam im Rahmen der Konzeptionalisierung des Produktes die Frage danach, welche Fähigkeiten der Roboter haben sollte und welche ethischen Risiken bei der Gestaltung vermieden werden sollten und antwortet mit konkreten Anforderungen an das Produkt. Daraufhin stellt das Team von Entwickler:innen im Entwicklungsprozess die Frage, wie unter Berücksichtigung der

1172 Gogoll et al. 2021:1087

Anforderungen des Projektteams konkrete Funktionen implementiert werden können. Die schwierige Aufgabe der technischen Umsetzung aller gesellschaftlicher und ethischer Anforderungen fällt den Entwickler:innen zu. Außerdem müssen sie alle verbleibenden ethischen Fragen klären bzw. an das Management oder das Projektteam zurückspielen.¹¹⁷³

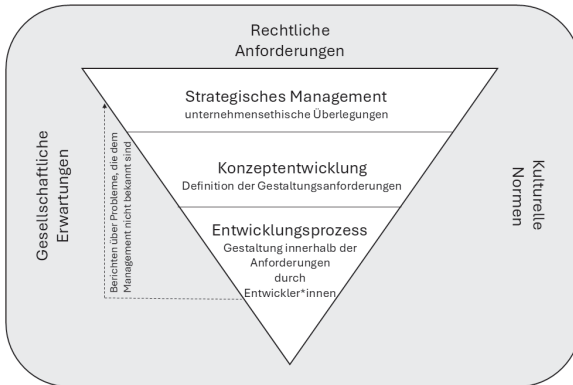


Abbildung 4 Entscheidungsebenen bei der Software-Entwicklung¹¹⁷⁴

Den medizinischen Informatiker:innen Maurice Mulvenna, Jennifer Boger und Raymond Bond zufolge ist es essentiell, dass alle Beteiligten die ethischen Aspekte einer Technologieentwicklung identifizieren, diskutieren und verstehen, da hier Funktionen implementiert werden können, welche für die Nutzenden der Technologie ethisch problematisch sind.¹¹⁷⁵ Der Einsatz von ethischer Reflexion in einer Organisation kann aber laut Manzeschke auch Probleme zur Folge haben: „Wenn Organisationen sich auf eine Bearbeitung von moralischen Fragen einlassen, dann stehen sie in der Gefahr, die Kontrolle über die darin verhandelten Themen zu verlieren.“¹¹⁷⁶ So könnte zum Beispiel bei dem Entwicklungsprozess eines Pflegeroboters auf jeder der Entscheidungsebenen Uneinigkeit bestehen über morali-

1173 Gogoll et al. 2021:1091

1174 Leicht veränderte Darstellung nach Gogoll et al. 2021:1089

1175 Mulvenna et al. 2017:51; 53

1176 Manzeschke 2021:229

sche Fragen. Wenn es auf der Ebene des Strategischen Managements um die Frage geht, ob der Pflegeroboter gebaut werden sollte, dann halten das die Utilitarist:innen innerhalb des Managements wahrscheinlich für moralisch zulässig, weil dadurch das Pflegepersonal entlastet und die Patient:innen versorgt werden. Die Tugendethiker:innen innerhalb des Managements könnten hingegen argumentieren, dass die empathische Beziehung zwischen Pflegepersonal und Patient:in essenziell ist und es nicht zulässig wäre, diese durch einen Roboter ohne Einfühlungsvermögen zu ersetzen. Gleichermäßen könnten sich die Teammitglieder bei der Konzeptentwicklung uneinig sein, ob die Möglichkeit, dass der Pflegeroboter für lebendig gehalten wird, ein ethisches Risiko darstellt oder ob genau diese Täuschung eine hilfreiche Funktion des Roboters erfüllt. Selbst die Entwickler:innen im Entwicklungsprozess, die genaue Anforderungen erfüllen müssen, haben noch einen Spielraum, in welchem ethische Fragen zu Uneinigkeit führen können: „Developers can, for example, choose a technology that protects the user’s privacy but still ensures the achievement of established business objectives.“¹¹⁷⁷ Fragen danach, ob über die Anforderungen aus der Konzeptentwicklung hinaus noch Schutzvorkehrungen getroffen werden sollten, können auch auf dieser Entscheidungsebene noch zu Konflikten führen.

Die Überforderung durch die Behandlung von moralischen Fragen stellt laut Manzeschke eine Bedrohung für die Handlungsfähigkeit der Organisation dar: „Moral, sofern sie in diesen Entscheidungsprozessen tatsächlich artikuliert wird, vervielfacht die Differenzen und Einschätzungen in Bezug auf eine Entscheidung, weshalb es eben auch starke Tendenzen gibt, der Moral bzw. den Moralen möglichst wenig Raum zu geben, um als Organisation handlungsfähig zu bleiben und nicht (zu) viel Zeit (=Geld) auf die Bearbeitung dieser Differenzen zu verwenden.“¹¹⁷⁸ Laut Manzeschke induziert Moral Differenzen und Krisen in Organisationen und wird deshalb, insbesondere in Unternehmen, aus Entscheidungsprozessen ausgeschlossen.¹¹⁷⁹

Ein weiterer Grund für den fehlenden Einsatz von Ethik kann dem KI-Ethiker Thilo Hagendorff zufolge ihre Wahrnehmung als

1177 Gogoll et al. 2021:1089

1178 Manzeschke 2021:228

1179 Ebd.:220

eine Behinderung für die Prozesse in der Organisation sein: „In business contexts, speed is everything in many cases and skipping ethical considerations is equivalent to the path of least resistance.“¹¹⁸⁰ Ethik wird oft nur als Hilfsmittel verstanden, das aber dem Gewinn des Unternehmens untergeordnet ist.¹¹⁸¹ Außerdem herrscht oft das Vorurteil, dass Ethik Aktivitäten stoppen und Handlungen verbieten oder vorschreiben will.¹¹⁸²

Genau dieses Behindern von Aktivität sollte jedoch laut Hagedorff nicht das Ziel der Ethik sein: „It should not be the objective of ethics to stifle activity, but to do the exact opposite, i.e. broadening the scope of action, uncovering blind spots, promoting autonomy and freedom, and fostering self-responsibility.“¹¹⁸³ Ethik kann im Gegenteil dazu beitragen, Aktivitäten zu fördern oder zu verbessern.¹¹⁸⁴

Es wird deutlich, dass innerhalb von Organisationen der Ethik gegenüber Skepsis herrscht. Um erfolgreiche Methoden der Technologieethik einzuführen, dürfen dem Technologieethiker Matt A. Murphy keine erheblichen Änderungen der bestehenden Prozesse gefordert werden: „They [the methods] must strike the balance of being rigorous while still remaining easy and relatively seamless to implement.“¹¹⁸⁵ Im Folgenden soll erörtert werden, wie sich diese schwierige Aufgabe in der Praxis umsetzen lässt. Da die Ethik allein keine praktischen Handlungsanweisungen bietet, soll nun eine Umsetzungsmethode gefunden werden, um die ethischen Erkenntnisse wirksam in die Praxis überführen zu können.

Es gibt verschiedenen Umsetzungsmethoden, die in der Literatur diskutiert und in der Praxis angewandt werden. Sie lassen sich grob in drei Kategorien einteilen: Ethische Leitlinien, die bei der Gestaltung der Technologie abgehakt werden können, einen aktiven Gestaltungsprozess, der durch ethische Werte geprägt ist oder Risikoanalysen, die vor der Entwicklung durchgeführt werden. Aus der bisherigen Arbeit lassen sich drei allgemeine Kriterien ableiten, die eine solche Umsetzungsmethode erfüllen sollte, damit sie bei der

1180 Hagedorff 2020:108

1181 Murphy 2023:4

1182 Boddington 2017:8

1183 Hagedorff 2020:112f.

1184 Boddington 2017:8

1185 Murphy 2023:4

Übersetzung meiner ethischen Erkenntnisse in die Praxis hilfreich sein kann:

Allgemeine Kriterien für die Umsetzungsmethode zur ethischen Technologieentwicklung

- i) Die Umsetzungsmethode sollte konkrete praktische Anleitung bieten.
- ii) Die Umsetzungsmethode sollte den Einfluss der Technologie auf alle Betroffenen berücksichtigen.
- iii) Im Rahmen der Umsetzungsmethode sollte die Verantwortung für die Umsetzung der ethischen Überlegungen bestimmten menschlichen Akteuren zugeordnet werden.

Dadurch, dass die Umsetzungsmethode konkrete praktische Anleitung bietet für die Ausrichtung technischer Entwicklungen an ethischen Maßstäben (i), wird es für eine Organisation möglich, ethische Überlegungen in die Praxis zu übersetzen. Außerdem bleibt die Orientierung an ethischen Überlegungen auf diese Weise kein Lippenbekenntnis, denn es fällt auf, wenn die konkrete praktische Anleitung nicht befolgt wird und somit bezüglich der ethischen Überlegungen keine Durchsetzungsmaßnahmen ergriffen werden. Damit der Einfluss der Technologie auf alle Betroffenen berücksichtigt werden kann (ii), muss der Vorteil aller betroffenen Personen das Ziel sein. Das heißt, die Risiken für eine Gruppe der betroffenen Personen können nicht einfach abgewogen werden gegen die Vorteile von anderen Personen – dann wäre die Umsetzungsmethode nicht geeignet für eine Technologieentwicklung an ethischen Maßstäben. Die Verantwortung für die Umsetzung der ethischen Überlegungen in die Praxis sollte bestimmten menschlichen Akteuren zugeordnet werden (iii), da auf diese Weise ein Akteur sowohl vorausschauend das Wohl aller Menschen im Blick hat als auch im Schadensfall für negative Folgen geradesteht.¹¹⁸⁶ Bei den Verantwortungsträgern kann es sich allerdings nicht um Technologien handeln. In Kapitel 8 wurde deutlich, dass es für einen ethischen Umgang mit Technologien wichtig ist, dass Personen oder Kollektive von Personen die Verantwortung tragen für potenzielle Schäden im Nachhinein und für die

1186 Hart 1968:212; Young 2011:104; Danaher 2016:299; Kiener 2022:584; Nyholm 2024:192

durchdachte Gestaltung im Vorhinein – die Technologie kann diese Verantwortung nicht übernehmen.

Diese drei Kriterien können als notwendige Bedingungen für die ethische Gestaltung eines jeden technischen Teils eines Datenökosystems verstanden werden. Denn die drei allgemeinen Kriterien für die gesuchte Umsetzungsmethode müssen alle erfüllt sein, bevor überhaupt in Betracht gezogen werden kann, dass die betrachtete Methode zur Ausrichtung technischer Entwicklungen an ethischen Maßstäben auch für gesamte Datenökosysteme geeignet ist. Dabei gelten dann zusätzlich noch drei Kriterien, die sich aus der bisherigen Arbeit ableiten lassen:

Spezifische Kriterien für die Umsetzungsmethode zur ethischen Gestaltung eines Datenökosystems

- I. Die Umsetzungsmethode sollte es ermöglichen, die Auferlegung von Risiken zu regulieren.
- II. Die Umsetzungsmethode sollte es ermöglichen, die administrativen Anforderungen zur Umsetzung von Privatsphäre für die Datenquellen zu erfüllen.
- III. Die Umsetzungsmethode sollte es ermöglichen, die administrativen Anforderungen zur Umsetzung von Transparenz für die Datennutzenden zu erfüllen.

Bei der ethischen Gestaltung eines Datenökosystems ist es wichtig, die Risikoauferlegung zu problematisieren, welcher die Datenquellen durch Datengenerierenden und die Datennutzenden durch Datenverarbeitende ausgesetzt sind. Die Auferlegung von Risiken kann reguliert werden (I), wenn es möglich ist, das Risiko zu identifizieren und wenn klar ist, unter welchen Umständen das Risiko zulässig ist. Des Weiteren sollten die administrativen Anforderungen zur Umsetzung von Privatsphäre für die Datenquellen erfüllt werden (II), da die Auferlegung des Risikos einer verletzten Privatsphäre der Datenquellen nur so Teil eines gerechten sozialen Systems der Risikobereitschaft sein kann. Es muss sichergestellt werden, dass beim Sammeln der Daten das Einverständnis der Datenquellen eingeholt wird, sie dabei vor unfreiwilliger Einwilligung geschützt werden, indem sie über jeden Verwendungszweck der Daten informiert werden und ihnen die Möglichkeit gegeben wird, ihre Einwilligung zu verweigern. Zusätzlich müssen alle Daten vor Zweckentfremdung

geschützt werden. Damit für die Datennutzenden das Risiko der fehlenden Transparenz Teil eines gerechten sozialen Systems der Risikobereitschaft sein kann, müssen ebenfalls die administrativen Anforderungen zur Umsetzung von Transparenz für die Datennutzenden erfüllt werden (III). Hier muss es darum gehen, algorithmische Ergebnisse im zum Verständnis notwendigen Kontext darzustellen, sodass klar wird, wie die Ergebnisse entstanden sind und welche Aussagen sie machen können.

Im Folgenden werden verschiedene Umsetzungsmethoden mit Hilfe der aus meiner Arbeit abgeleiteten Kriterien auf ihre Eignung als Grundlage zur ethischen Gestaltung eines Datenökosystems geprüft. Die Umsetzungsmethoden gliedern sich in die drei Kategorien Ethische Leitlinien, aktiver Gestaltungsprozess, und Risikoanalysen.

9.1 Ethische Leitlinien

Eine Liste von Anforderungen wird in der Regel in Form von Ethischen Leitlinien oder von einem Verhaltenskodex erstellt. Die Technologieethiker:innen Kelly Laas, Michael Davis, und Elisabeth Hildt definieren einen Ethikkodex folgendermaßen: „A code of ethics is an authoritative formulation of the (morally permissible) standards governing the conduct of members of a group, just because they are members of that group.“¹¹⁸⁷ Die Verbindlichkeit eines Ethikkodexes begründet sich immer in der Moral.¹¹⁸⁸ Ein Ethikkodex hat laut Laas et al. mehrere Funktionen. So kann er besondere Verhaltensnormen dokumentieren, erklären oder festlegen, insbesondere wenn Traditionen oder Gewohnheiten nicht mehr ausreichen. Ein solcher Kodex hilft Personen, die in der Praxis neu sind, zu lernen, wie sie sich verhalten sollen und erinnert Personen mit viel Erfahrung an das, was sie sonst vergessen würden. Zudem kann im Fall eines Streits mit Hilfe eines Ethikkodexes geschlichtet werden. Außerdem kann ein Ethikkodex dazu beitragen, dass Außenstehende verstehen, was von den Mitgliedern der Gruppe erwartet werden kann. Schließlich

1187 Laas et al. 2022:2

1188 Ebd.

kann auf Grund von einem Ethikkodex auch eine rechtliche Haftung gerechtfertigt werden.¹¹⁸⁹

In der akademischen Literatur wurden in den letzten Jahren viele verschiedene Ethikkodizes und Leitlinien für die Nutzung von Daten oder KI aufgestellt.¹¹⁹⁰ Im Jahr 2019 wurden in einer Studie 84 verschiedene ethische Richtlinien für KI allein in der akademischen normativen Ethikliteratur gefunden – Dokumente aus dem öffentlichen und privaten Sektor sind dabei noch nicht berücksichtigt.¹¹⁹¹ Aber auch Non-Profit-Organisationen, Unternehmen und politische Institutionen geben ihre eigenen digitalen Ethikrichtlinien heraus. So schreibt der Sozialwissenschaftler und Historiker Markus Schmitz: „Viele Institutionen wie die EU, die OECD und das Future of Life Institute sowie einige größere Technologieunternehmen haben bereits ethische Grundsätze zum Einsatz von KI und zur Nutzung von Daten veröffentlicht.“¹¹⁹² Als Orientierungshilfe für die ethische Gestaltung der Datennutzung werden ethische Grundsätze genannt, die berücksichtigt werden sollten, damit alle von den Technologien profitieren und gleichzeitig die Risiken minimiert werden können.¹¹⁹³ In den akademischen Leitlinien zur KI sind die wiederkehrenden Grundsätze: "(...) transparency, justice and fairness, non-maleficence, responsibility and privacy."¹¹⁹⁴

Als erstes fällt auf, dass die Herkunft dieser Grundsätze unklar ist. Laas et al. argumentieren, dass Ethische Leitlinien abhängig davon, wer sie verfasst, unterschiedlich ausfallen, da jede Gruppe von Autor:innen ihre eigene Sicht der Dinge hat. Für die Fachleute in dem betreffenden Bereich können ganz andere Aspekte einer neuen Technologie wichtig sein als für die Öffentlichkeit oder für Ethiker:innen.¹¹⁹⁵ Um den Einfluss der Technologie auf alle Betroffenen zu berücksichtigen (ii), erscheint es sinnvoll, alle Perspektiven in den Ethischen Leitlinien und Kodizes zu repräsentieren. Allerdings gibt es im Kontext dieser Umsetzungsmethode keine Anleitung dafür, wie entschieden wird, welche Grundsätze eine Rolle spielen sollen.

1189 Laas et al. 2022:2

1190 Laas et al. 2022:3; Jobin et al. 2019:391

1191 Jobin et al. 2019:391

1192 Schmitz 2020:368

1193 Ebd.

1194 Jobin et al. 2019:391

1195 Laas et al. 2022:4f.

Die Grundsätze sind in allen Sektoren ähnlich, aber die Häufigkeit einzelner Grundsätze variiert je nachdem, wer die Leitlinien verfasst hat. So haben die KI-Forscher Yi Zeng, Enmeng Lu und Cunqing Huangfu herausgefunden, dass der Grundsatz „Privatsphäre“ in den von Regierungen erstellten Leitlinien sehr präsent ist, in denen von Wissenschaftler:innen dagegen kommt er statistisch gesehen deutlich seltener vor und im Unternehmenssektor sogar noch seltener.¹¹⁹⁶ Aber selbst wenn die gleichen Grundsätze hervorgehoben werden, unterscheiden sich ihre Auslegung und die Ideen für ihre Umsetzung erheblich.¹¹⁹⁷ Der Gedanke findet sich auch bei dem KI-Experten Sebastian Hallensleben und seinen Kolleg:innen: „Values that form the basis for AI ethics are open to different interpretations and may vary depending on an AI system’s application context.“¹¹⁹⁸ Abhängig vom Kontext, in dem ein KI-System eingesetzt wird, unterscheiden sich die Ansprüche an seine ethischen Werte. So ist zum Beispiel der Anspruch an Privatsphäre und Fairness höher im Justizbereich als in der industriellen Produktion.¹¹⁹⁹

Durch die unterschiedlichen Ansprüche an die ethischen Werte in verschiedenen Kontexten besteht in der praktischen Umsetzung Unsicherheit darüber, welche ethischen Grundsätze Vorrang haben und wie Konflikte zwischen ethischen Grundsätzen gelöst werden sollten.¹²⁰⁰ Ein solcher Konflikt könnte zum Beispiel dann entstehen, wenn durch das Sammeln möglichst vieler personenbezogener Daten einerseits die algorithmischen Ergebnisse weniger verzerrt und diskriminierend würden, aber andererseits die einzelnen Personen durch das Sammeln der Daten in ihrer Privatsphäre beschränkt würden.¹²⁰¹ Es ist nicht klar geregelt, ob das Entdiskriminieren der Ergebnisse oder der Privatsphäreschutz der Datenquellen Vorrang hat und wie dieser Konflikt in der Praxis aufgelöst werden soll. Ethische Prinzipien werden in der Literatur für den Kontext der Praxis als nutzlos, ineffektiv und zahnlos bezeichnet.¹²⁰² Dem Informatiker

1196 Zeng et al. 2019:3, Figure 3; Gogoll et al. 2021:1094

1197 Jobin et al. 2019:391

1198 Hallensleben et al. 2020:10

1199 Ebd.

1200 Jobin et al. 2019:396

1201 Jobin et al. 2019:396

1202 Rességuier, Rodrigues 2020:1; Munn 2023:869, 871; Hagendorff 2020:99; Mulvenna et al. 2017:51

Luke Munn zufolge ist es deshalb leicht die ethischen Prinzipien zu ignorieren: „The failure of AI ethical principles is not spectacular but silent, resulting in the desired outcome: business as usual.“¹²⁰³ Dass trotz der Anerkennung ethischer Prinzipien diese in der Praxis ignoriert werden deutet darauf hin, dass zwischen den Prinzipien und ihrer praktischen Umsetzung eine Diskrepanz besteht. Selbst wenn man sich auf Prinzipien geeinigt hat, bewirkt das noch keine tatsächliche Veränderung in der Gestaltung algorithmischer Systeme.¹²⁰⁴

Die Softwareentwickler Andrew McNamara, Justin Smith and Emerson Murphy-Hill fanden in einer Studie mit sowohl Studierenden der Softwaretechnik als auch professionellen Softwareentwickler:innen heraus, dass die ausdrückliche Anweisung, einen ethischen Kodex zu berücksichtigen, keinen erkennbaren Unterschied in der Arbeit gemacht hat – die Studienteilnehmer:innen haben ihre bewährten Arbeitsmethoden nicht auf Grund des ethischen Kodex geändert.¹²⁰⁵ Die bloße Einführung von Ethikgrundsätzen in einer Organisation ändert also das Verhalten der Entwickler:innen nicht.¹²⁰⁶

Es gibt Gründe für diese beobachtbare Ineffizienz der ethischen Prinzipien. Einer der Gründe ist, dass ein Abweichen von einem Ethik-Kodex keine Konsequenzen nach sich zieht und es auch sonst keinen Mechanismus für die Durchsetzung von Leitlinien gibt.¹²⁰⁷ Dieses Problem könnte laut Schmitz zum Beispiel durch den Einsatz eines Ethikbeirates gelöst werden, welcher die Ethischen Leitlinien des betreffenden Unternehmens ausarbeitet und dann ihre Einhaltung unterstützt und kontrolliert.¹²⁰⁸

Laut Gogoll et al. ist ein weiterer Grund für die fehlende Umsetzung ethischer Prinzipien in die Praxis die Unterbestimmtheit („underdetermination“) von ethischen Werten. Sie fungierten als allgemeine Aussagen, die alleine keine konkrete oder praktische Orientierung bieten können.¹²⁰⁹ Mittelstadt stimmt zu: „AI Ethics initiatives have thus far largely produced vague, high-level principles and value statements which promise to be action-guiding, but in practice

1203 Munn 2023:872

1204 Morley et al. 2020:2147

1205 McNamara et al. 2018:732

1206 Murphy 2023:3

1207 Hagendorff 2020:99, 113; Schmitz 2020:369f.; Munn 2023:871

1208 Schmitz 2020:371

1209 Gogoll et al. 2021:1095

provide few specific recommendations (...).¹²¹⁰ Ohne ausreichende Konkretisierung und Kontextualisierung sind Software-Ingenieure mit der Aufgabe, moralische Werte zu beachten, auf sich allein gestellt. Eine Schwierigkeit daran ist, dass Werte nicht prozessorientiert sind und somit auch die Mittel, mit denen sie erreicht werden können, nicht logisch erschlossen werden können. Deshalb werden laut Gogoll et al. Werte oft erst im Nachhinein betrachtet, anstatt das Handeln von vorneherein danach auszurichten. Hinzu kommt, dass es sich Menschen gerne leicht machen und am liebsten automatische Lösungen finden durch Anwendung von Dispositionen, Konventionen oder moralischen Regeln – das ist aber bei Widersprüchen nicht mehr zielführend.¹²¹¹ Das größte Problem bei Ethischen Leitlinien ist also das Fehlen konkreter praktischer Anleitung (Kriterium i).

Ein weiterer Grund dafür, dass ethische Prinzipien nicht in die Praxis umgesetzt werden, könnte sein, dass manche Unternehmen kein ethisches Fundament haben. Denn ohne eine breitere moralische und ethische Grundlage können Prinzipien in der Praxis missachtet werden.¹²¹² Nur weil ein Unternehmen ethische Richtlinien und Checklisten einführt, bedeutet das noch nicht, dass dadurch ethische Produkte entwickelt werden – im Gegenteil: Unternehmen können den Anschein von Ethik genießen, ohne dass diese Substanz hat.¹²¹³ So können mit Hilfe der Einführung ethischer Prinzipien kritische Stimmen beruhigt werden, während sich aber an den kritisierten Praktiken nichts ändert.¹²¹⁴

Diese Beibehaltung der bestehenden Praxis ist manchmal genau das, was durch das Aufstellen von ethischen Prinzipien erzielt werden soll, denn dadurch wird laut Floridi der Regulierung durch Gesetze zuvorgekommen¹²¹⁵: „(...) ethics may be exploited as if it were an alternative to legislation (...)“¹²¹⁶ Ethische Leitlinien werden manchmal als nichtlegislative politische Instrumente oder auch „Soft Law“ bezeichnet.¹²¹⁷ Diese ethischen Prinzipien sind aber kein Er-

1210 Mittelstadt 2019:1

1211 Gogoll et al. 2021:1098

1212 Lauer 2021:24

1213 Munn 2023:872

1214 Hagendorff 2020:100

1215 Ressayguier, Rodrigues 2020:2; Floridi 2019:188

1216 Floridi 2019:188

1217 Jobin et al. 2019:389

satz für gesetzliche Regulierung, da sie keine konkreten Verhaltensregeln vorschreibt¹²¹⁸ und, wie bereits erläutert, oft nur schwer in die Praxis übersetzbar ist. Die Ethischen Leitlinien sollten als Methode zur Umsetzung der Ethik bei dieser Übersetzung behilflich sein. Doch der Interpretationsspielraum, den Ethische Leitlinien zulassen, ermöglicht die Ausnutzung der Ethik durch zum Beispiel Ethics Shopping und Ethics Washing.¹²¹⁹

Mit Ethics Shopping bezeichnet man die Maßnahme, bestimmte ethische Grundsätze aus einer Vielzahl möglicher Grundsätze auszuwählen, um bereits bestehende Verhaltensweisen im Nachhinein zu rechtfertigen.¹²²⁰ Gogoll et al. stellen fest, dass keine kohärente Ethik verwirklicht werden kann, wenn Verhaltenskodizes als Ansammlungen von Werten verstanden werden, aus denen wir mehr oder weniger willkürlich Werte auswählen können.¹²²¹ Weil dieses Ethics Shopping möglich ist, sind die Autor:innen der Meinung, dass Ethische Leitlinien und Verhaltenskodizes keine normative Orientierung bieten. Diese Orientierungsfunktion werde außerdem durch eine Indifferenz verhindert. Denn ein bestimmter Verhaltenskodex könne zur Rechtfertigung unterschiedlicher oder sogar widersprüchlicher Handlungen herangezogen werden.¹²²²

Ethics Washing bezeichnet ein Vorgehen, bei dem man nur den Anschein erweckt, sich ethisch zu verhalten, aber tatsächlich Fehlinformationen über ethische Werte verbreitet oder oberflächliche Maßnahmen zugunsten dieser Werte ergreift.¹²²³ So können zum Beispiel ethische Prinzipien wie Gerechtigkeit passend zu den Produktmerkmalen definiert werden, sodass man behaupten kann, man hielte ethische Grundsätze ein, ohne sich aber auf die nötige Reflexion oder Umgestaltung einzulassen.¹²²⁴

Es stellt sich also die Frage, wie ethische Prinzipien so konkretisiert werden können, dass ihre Umsetzung ermöglicht wird und dass gleichzeitig Ethics Shopping und Ethics Washing verhindert werden

1218 Rességuier, Rodrigues 2020:2

1219 Ebd.; Floridi 2019:188

1220 Floridi 2019:186

1221 Gogoll et al. 2021:1097

1222 Ebd.:1097

1223 Floridi 2019:187

1224 Munn 2023:870

können. Murphy weist auf folgenden Lösungsansatz hin: „One popular way to try and concretize ethical principles is to create ethical checklists to be filled out alongside development.“¹²²⁵ Den Technologiephilosoph:innen Asle H. Kiran, Nelly Oudshoorn und Peter-Paul Verbeek zufolge vernachlässigt eine statische Checkliste allerdings, dass ethische Grundsätze durch neue technische Entwicklungen beeinflusst werden können und auch dass in verschiedenen kulturellen Umfeldern unterschiedliche ethische Ansprüche herrschen.¹²²⁶ Alternativ können laut Murphy ethische Prinzipien auch konkretisiert werden, indem diese Prinzipien als ethische Fragen formuliert werden.¹²²⁷ Hier erkennt allerdings Murphy selbst ein Problem: „The more open-ended or unstructured questions are in methods like this, the more they invite unstructured and unsystematic answers.“¹²²⁸ Hagendorff ist der Meinung, dass Ethische Leitlinien wirksamer sein könnten, wenn sie durch detailliertere technische Erläuterungen ergänzt würden: „The fact that not a single prominent ethical guideline goes into greater technical detail shows how deep the gap is between concrete contexts of research, development, and application on the one side, and ethical thinking on the other.“¹²²⁹ Die Ergänzung der Ethischen Leitlinien um technische Details erfordert jedoch eine Übersetzung der Theorie in die Praxis.

Das VCIO (Values, Criteria, Indicators, Observables) Modell ist eine Methode zur Spezifizierung und Operationalisierung von ethischen Werten, die von Hallensleben et al. entwickelt wurde. Diese Methode soll ethische Grundsätze praktikabel, vergleichbar und messbar machen.¹²³⁰ Denn ethische Werte sollten menschliches Handeln leiten. Sie enthalten jedoch keine Standards zur Bewertung ihrer Umsetzung. Deshalb sollen nach dem VCIO-Modell andere Komponenten diese Aufgaben erfüllen – und zwar Kriterien, Indikatoren und beobachtbare Größen. Die Kriterien legen fest, wann Werte erfüllt sind oder verletzt werden. Um wiederum feststellen zu können, ob ein Kriterium erfüllt ist, werden Indikatoren benötigt.

1225 Murphy 2023:2

1226 Kiran et al. 2015:6

1227 Murphy 2023:3

1228 Ebd.:3

1229 Hagendorff 2020:111

1230 Hallensleben et al. 2020:8

Diese prüfen, ob die Kriterien erfüllt sind. Schließlich muss es noch beobachtbare Größen geben. Diese quantifizieren oder qualifizieren inwieweit die Indikatoren erfüllt sind. Die Kriterien bilden auf einer abstrakten Ebene notwendige Bedingungen für die Erfüllung von Werten und die Indikatoren stellen notwendige Bedingungen für die Erfüllung der Kriterien dar. Sie werden in deliberativen Prozessen aufgestellt und festgelegt.¹²³¹

Das sogenannte „AI Ethics Label“ soll der Übersichtlichkeit dienen. Dieses Label ist inspiriert von der Energieverbrauchskennzeichnung. Mit Hilfe von Buchstaben und Farben soll auf einen Blick verständlich sein, inwiefern ein bestimmter Wert erfüllt wird. Dafür werden die verschiedenen beobachtbaren Größen für einen bestimmten Wert in einer Bewertung zusammengefasst. Wenn zum Beispiel der Wert „Transparenz“ mit Hilfe des „AI Ethics Labels“ dargestellt werden soll, dann werden die zur Erfüllung notwendigen Bedingungen in Form von Kriterien und ihren Indikatoren betrachtet. Die beobachtbaren Größen quantifizieren oder qualifizieren, inwieweit die Indikatoren erfüllt sind und können dann mit A bis G bewertet werden. Der Durchschnitt dieser Bewertungen wird dann auf dem Label angezeigt.¹²³²

Bei der Bewertung eines Algorithmus ist ein Kriterium des Wertes „Transparenz“ die „Zugänglichkeit“. Indikatoren für dieses Kriterium sind dann die Frage danach, wer Zugang zu dem Algorithmus hat und ob es für den Konfliktfall eine Schlichtungsstelle gibt. Beobachtbare Größen sind dann die Antworten auf diese Fragen. So könnte jede Person gleichermaßen Zugang zu dem Algorithmus haben – das wäre mit einer guten Bewertung (A) zu versehen – oder es hat niemand Zugang zu dem Algorithmus – das wäre mit einer schlechten Bewertung (G) zu versehen. Ebenfalls ist beobachtbar, ob es für den Konfliktfall eine (kompetente) Schlichtungsstelle gibt (A) oder nicht (G). Aus dem Durchschnitt der Bewertungen ergibt sich dann die Gesamtbewertung, welche anzeigt, inwiefern der zu bewertende Algorithmus das Kriterium „Zugänglichkeit“ des Wertes „Transparenz“ erfüllt.¹²³³

1231 Hallensleben et al. 2020:15ff.

1232 Ebd.:31f.

1233 Ebd.:21

Die Anforderungen an einen Algorithmus sind unterschiedlich, je nachdem in welchem Kontext er eingesetzt wird. Deshalb haben die Sozioinformatiker:innen Tobias D. Krafft und Katharina Zweig eine Risikomatrix aufgestellt, welche das Risiko für den potenziellen Schaden eines KI-Systems in Abhängigkeit von seiner Nutzung aufzeigt. Dabei wird einerseits die Intensität des potenziellen Schadens eines KI-Systems bewertet, in dem die Auswirkungen auf individuelle Rechte und die Gesellschaft als Ganzes betrachtet werden. Andererseits wird die Abhängigkeit der potenziell betroffenen Parteien von der algorithmischen Entscheidung bewertet, in dem die Möglichkeiten zur Vermeidung des potenziellen Schadens betrachtet werden.¹²³⁴

Krafft und Zweig leiten fünf verschiedene Regulierungsklassen ab, die unterschiedlich hohe Anforderungen an Transparenz und Nachvollziehbarkeit stellen. Während für KI-Systeme in Klasse 0 – bei zum Beispiel der personalisierten Auswahl von Werbung für Kleidung – keine Transparenzpflichten gefordert werden, sollen KI-Systeme in Klasse 4 – bei zum Beispiel der personalisierten Auswahl von Werbung für medizinische Dienstleistungen – nur dann eingesetzt werden, wenn dem Risiko nachweisbar ein hoher Gesamtnutzen entgegensteht. In Klasse 1 bis 3 steigen die Anforderungen an Transparenzpflichten mit dem Risiko, welches das KI-System mit sich bringt.¹²³⁵

Mit dem VCIO-Modell gelingt es, ethische Prinzipien so zu konkretisieren, dass ethischen Leitlinien oder Kodexe konkrete praktische Anleitung bieten können und somit Kriterium i) erfüllen. Auch Kriterium iii) wird erfüllt, da die Verantwortung für die ethische Gestaltung der betreffenden Technologie im VCIO-Modell bestimmten menschlichen Akteuren zugeordnet wird.¹²³⁶ Insbesondere durch die Risikomatrix von Krafft und Zweig Betrachtung wird deutlich, dass die Auswirkungen der Technologie auf alle potenziell betroffenen Parteien berücksichtigt wird und also auch Kriterium ii) erfüllt ist.

Auch wenn die ethischen Prinzipien mit Hilfe des VCIO-Modells konkretisiert werden und praktische Anleitung bieten können, bleibt

1234 Krafft, Zweig 2019:19ff.; Hallensleben et al. 2020:37f.

1235 Hallensleben et al. 2020:38ff.; Krafft, Zweig 2019:31

1236 Hallensleben et al. 2020:27

die Herkunft dieser Prinzipien weiter unklar. Denn Hallensleben et al. schreiben: „For the proposed AI Ethics Label, we carefully suggest six values (...) based on contemporary discourse and operability.”¹²³⁷ Im Kontext des VCIO-Modells wurden die ethischen Werte, die spezifiziert werden sollten, auf der Grundlage des heutigen Diskurses und ihrer Operationalisierbarkeit ausgewählt. Es handelt sich bei den sechs Werten nur um einen Vorschlag, aber trotzdem birgt die fehlende Begründung der Auswahl der Werte die Gefahr von Ethics Shopping. Die Relevanz im zeitgenössischen Diskurs ist keine gute Begründung, denn der Diskurs kann manche Werte vernachlässigen und kann außerdem seine Schwerpunkte verändern. Zudem sind Werte nicht deshalb relevant, weil sie operationalisierbar sind – diese Perspektive ist pragmatisch, aber erfüllt nicht den ursprünglichen Zweck der Umsetzung ethischer Werte in die Praxis.

Im VCIO-Modell ist die Wahl der ethischen Werte nicht gut begründet, aber die in dieser Arbeit betonten ethischen Werte wurden aus der epistemischen Asymmetrie, Kontrollungleichheit und dadurch ermöglichten Risikoauferlegung im Datenökosystem hergeleitet. Mit Hilfe der Umsetzungsmethode der Ethischen Leitlinien könnte also die von mir begründeten ethischen Prinzipien der Privatsphäre und Transparenz spezifiziert werden. Da sich die administrativen Anforderungen zur Umsetzung von Privatsphäre und Transparenz mit Sicherheit in den Indikatoren für Privatsphäre und Transparenz wiederfinden würden, wären dadurch auch Kriterium (II) und (III) erfüllt. Auf diese Weise ließe sich zumindest ein Teil meiner ethischen Überlegungen in die Praxis übersetzen. Aber Kriterium (I) erfüllt die Umsetzungsmethode der Ethischen Leitlinien nicht. Denn sie ermöglicht es nicht, die Risikoauferlegung zu problematisieren und kann dadurch auch nicht die Auferlegung von Risiken in der Praxis regulieren. Da aber das Kriterium (I) als notwendige Bedingung für die ethische Gestaltung eines Datenökosystems verstanden werden kann, eignen sich Ethische Leitlinien nicht zur Umsetzung meiner ethischen Überlegungen in die Praxis.

1237 Hallensleben et al. 2020:7

9.2 Gestaltungsprozess

Anstatt der Anwendung Ethischer Leitlinien bei der Entwicklung von Technologien, gibt es auch die Möglichkeit, ethische Überlegungen direkt in den Gestaltungsprozess bei der Technologieentwicklung zu integrieren. Dadurch werden die ethischen Überlegungen lebendig gehalten, anstatt sie in feste Leitlinien zu gießen. Das ist laut Rességuier, Rodrigues notwendig: „AI ethics is in a deep and vital need of the continuously renewed process of questioning the world and the lenses through which we see it – consciously, constantly and iteratively.“¹²³⁸ Diese KI- oder Technologieethik kann als ein reflektierender Prozess zur Entwicklung möglichst ethisch gestalteter Technologien verstanden werden, der unter anderem dabei hilft, Voreingenommenheit unter den am Gestaltungsprozess Beteiligten aufzudecken.¹²³⁹ Die Technologiephilosoph:innen und Digitaletiker:innen Jessica Morley, Luciano Floridi und Libby Kinsey sowie die Psychologin Anat Elhalal sind der Meinung, dass ein solcher Wechsel von ethischen Grundsätzen hin zu einem ethischen Gestaltungsprozess die Ethik in der Praxis nachvollziehbarer macht.¹²⁴⁰

Der österreichische Designer Victor Papanek definiert einen Gestaltungsprozess als „Das Planen und Konzipieren von Handlungen auf ein erwünschtes, absehbares Ziel hin (...)“.¹²⁴¹ Ein Prozess kann also gestaltet werden, indem man sich bewusst um eine sinnvolle Ordnung bemüht.¹²⁴² Ob eine solche Ordnung sinnvoll oder richtig ist, hängt laut Papanek davon ab, welche Bedeutung wir ihr zumessen.¹²⁴³ Er schreibt: „Design muss sinnvoll sein.“¹²⁴⁴ Aber das Wort „sinnvoll“ könne durch alle möglichen Begriffe ersetzt werden, welche dann den Zweck des Designs, und damit seine Funktion, definieren. Kann die Funktion eines Designs zum Beispiel sein, schön oder abstrakt zu wirken, so kann sie auch sein, ethisch durchdacht zu sein.

1238 Rességuier, Rodrigues 2020:4

1239 Morley et al. 2021:246f.

1240 Ebd.:252

1241 Papanek 2009:20

1242 Ebd.:20

1243 Papanek 2009:21

1244 Ebd.:22

Der erste Schritt in einem Gestaltungsprozess ist das Formulieren von Anforderungen an den zu gestaltenden Gegenstand. Dabei spielen viele Akteure eine Rolle – Personen, die ein Interesse an dem zu gestaltenden Gegenstand haben oder Personen, die von seinen Auswirkungen betroffen sein könnten, wie zum Beispiel Auftraggebende und Kund:innen.¹²⁴⁵ Die Anforderungen werden nicht einmalig zu Beginn definiert, sondern sie werden während des gesamten Gestaltungsprozesses weiter formuliert.¹²⁴⁶ In einem Gestaltungsprozess wird nicht nur die optimale Lösung für bestimmte Ziele und Anforderungen gesucht, sondern es werden auch Entscheidungen getroffen, die über die ursprünglich formulierten Anforderungen und Ziele hinausgehen.¹²⁴⁷

Manchmal entsteht aber auch das Problem, dass nicht alle Anforderungen an die Gestaltung gleichzeitig erfüllt werden können. Eine mögliche Lösung im Engineering Bereich ist das Formulieren von Schwellenwerten für bestimmte Anforderungen, bei deren Überschreitung der Entwurf als inakzeptabel angesehen wird – diese Methode nennt sich „satisficing“. Die Sicherheitsethikerin Anke van Gorp und der Technologieethiker Ibo van de Poel definieren sie folgendermaßen: „Satisficing means that a designer is satisfied with any solution that is “good enough” given the design requirements and will not look for optimal solutions.“¹²⁴⁸ Dieser Ansatz minimiert die Komplexität der Gestaltung, sie führt jedoch nicht zu den ethisch wünschenswertesten Lösungen. Ethisch problematisch wird es, wenn zum Beispiel Sicherheitsanforderungen überdacht oder verwässert werden, weil sie mit anderen Anforderungen kollidieren.¹²⁴⁹

Hier besteht laut der Technologieforscherin Barbara Ribeiro und ihren Kolleg:innen ein „(...) issue of societal alignment (...)“¹²⁵⁰ – eine Problematik der Anpassung an die Gesellschaft. Diese beschreibt die Notwendigkeit, eine bessere Abstimmung zwischen den Zielen von Innovationen und den Zielen der Öffentlichkeit zu erreichen. Das heißt, bei der Gestaltung von Innovationen soll sicherge-

1245 Van Gorp, van de Poel 2001:17

1246 Ebd.:18

1247 Ebd.:18

1248 Ebd.:19

1249 Ebd.:21

1250 Ribeiro et al. 2018:318

stellt werden, dass die Innovationsergebnisse mit den Werten und Bedürfnissen der Öffentlichkeit in Einklang gebracht werden. Die Herausforderung besteht dabei darin, die Öffentlichkeit einzubeziehen, gesellschaftliche Bedürfnisse zu formulieren und die Ziele der Innovationen auf die Erfüllung dieser Bedürfnisse abzustimmen.¹²⁵¹

Gestaltungsprozessmethoden zur Umsetzung von ethischen Überlegungen in die Praxis sind in der Literatur in sieben verschiedene Ausgestaltungen prävalent: Embedded Values, Embedded Ethics, Ethics by Design, Human-centered Design, Privacy by Design, Responsible Innovation und Value-Sensitive Design. Im Folgenden werden die ersten vier nur kurz vorgestellt, da sie Kriterium i) nicht erfüllen und keine konkrete praktische Anleitung bieten. Sie begründen aber, warum die ethische Abwägung bezüglich der praktischen Anwendung von Technologien zur Datenverarbeitung schon bei ihrer Gestaltung essentiell ist. Die letzten drei Methoden werden etwas ausführlicher dargestellt, um zeigen zu können, dass keine von ihnen alle Kriterien erfüllt.

9.2.1 Vier Ansätze

Embedded Values

Der Embedded Values Ansatz geht von dem Verständnis aus, dass die Werte der Designer in die Technologie einfließen. In Bezug auf eine ethische Produktentwicklung ist der Anspruch, dass die gestalteten Technologien wünschenswerte Werte fördern oder unterstützen und dadurch wünschenswerte Praktiken aufrechterhalten.¹²⁵² In der Literatur gibt es drei verschiedene Positionen dazu, wie Werte in technologische Entwürfe einfließen¹²⁵³: *verkörpert*, *exogen* und *interaktionell*. Die *verkörperte Position* besagt, dass die Designer

1251 Ribeiro et al. 2018:318, 324

1252 Zuber et al. 2022:3–5

1253 Dem Technikforscher Patrick Feng zufolge haben in den 1970er Jahren die Sozialwissenschaftler*innen Technologie noch als wertneutral verstanden. Im Jahr 2000 jedoch schreibt er, dass die Wissenschaftler*innen sich mittlerweile einig sind, dass Technologien Werte verkörpern, die sich in ihrer Gestaltung widerspiegeln (Feng 2000:212).

ihre eigenen Absichten und Werte in die Technologie übertragen.¹²⁵⁴ Die *exogene Position* geht davon aus, dass gesellschaftliche Kräfte maßgeblich beeinflussen, wie eine Technologie genutzt wird. Technologische Systeme seien deshalb keinesfalls wertneutral, sondern begünstigen die Interessen von Menschen mit wirtschaftlicher und politischer Macht.¹²⁵⁵ Die *interaktionale Position* besagt, dass die Merkmale oder Eigenschaften, die Menschen in Technologien einfließen lassen, bestimmte Werte eher unterstützen und andere behindern, während die tatsächliche Nutzung der Technologie von den Zielen der Menschen abhängt, die mit ihr interagieren.¹²⁵⁶

Der Embedded Values Ansatz macht die Designer als bestimmten kollektiven Akteur dafür verantwortlich, dass sie ihre ethischen Werte im Rahmen des Gestaltungsprozesses in die betreffende Technologie einfließen lassen (iii). Das widerspricht im Prinzip nicht dem Ziel, dass der Einfluss der Technologie auf alle Betroffenen berücksichtigt werden soll (ii), aber da es sich bei den Designern um einen sehr kleinen Teil der Gesellschaft handelt, besteht die Gefahr, dass sie Nachteile für andere gesellschaftliche Gruppen übersehen. Außerdem fehlt die konkrete praktische Anleitung aus Kriterium i), da unklar ist, welche Schritte man gehen müsste, um ethische Werte in die Technologie einzubetten. Da schon die allgemeinen Kriterien nicht vollständig erfüllt sind, erübrigt es sich die spezifischen Kriterien zu betrachten. Somit eignet sich der Embedded Values Ansatz nicht zur Umsetzung von ethischen Überlegungen in die Praxis.

Embedded Ethics

Der Ansatz der Embedded Ethics hat laut dem Medizinethiker Stuart McLennan und seinen Kolleg:innen ein übergeordnetes Ziel: „The overarching objective of an embedded ethics approach is to help develop AI technologies that are ethically and socially responsible, and that benefit and do not harm individuals and communities.”¹²⁵⁷ Es gibt allerdings unterschiedliche Wege, auf denen dieses Ziel erreicht werden kann. So wird der Begriff Embedded Ethics ver-

1254 Friedman, Kahn 2008:1242f.

1255 Ebd.:1243

1256 Ebd.:1243f.

1257 McLennan et al. 2020:488

wendet für die Verknüpfung von Ethik und automatisiertem Denken in autonomen Maschinen¹²⁵⁸, aber auch für integrierte Unterrichtseinheiten zum Thema ethisches Denken im Informatik-Lehrplan¹²⁵⁹ und schließlich für das Integrieren einer Ethikerin oder eines Ethikers in den Entwicklungsprozess einer Technologie¹²⁶⁰.

Der McLennan et al. vertreten die Ansicht, dass durch die integrierten Ethiker:innen erklärt und begründet werden sollte, warum der verwendete ethische Ansatz geeignet ist, um die spezifischen Ziele des Projekts zu erreichen.¹²⁶¹ Die Verantwortung tragen also die einzelnen integrierten Ethiker:innen (iii). Da es ihre Aufgabe ist, die Entwicklung von Technologien zu begleiten, die dem Einzelnen und der Gemeinschaft nutzen und nicht schaden¹²⁶², ist das Kriterium, dass der Einfluss der Technologie auf alle Betroffenen berücksichtigt wird, erfüllt (ii). Jedoch ist auch bei dem Ansatz der Embedded Ethics das Fehlen der konkreten praktischen Anleitung für seine Umsetzung ein Problem (i). Denn wie genau Maschinen programmiert werden müssten, um ethische Empfehlungen geben zu können, was genau Informatikstudierende oder -auszubildende im Unterricht lernen müssten, und was genau die Aufgaben der integrierten Ethiker:innen sind, über das Begründen des verwendeten ethischen Ansatzes hinaus, bleibt unklar. Weil Kriterium i) nicht erfüllt ist, wird auch der Embedded Ethics Ansatz in dieser Arbeit nicht weiterverfolgt.

Ethics by Design

Der Ethics by Design Ansatz umfasst die proaktive Identifizierung und Behandlung potenzieller ethischer Fragen während des Designprozesses.¹²⁶³ Bei der Entwicklung von Technologien, die deutliche Auswirkungen auf unsere Gesellschaft haben könnten, stellt sich die Frage, wie sie mit den Werten in Einklang gebracht werden können, auf denen die Menschenrechte und das Wohlergehen beru-

1258 Bonnemains et al. 2018

1259 Grosz et al. 2019

1260 McLennan et al. 2020:488; McLennan et al. 2022:3

1261 Ebd.:5

1262 Ebd.:488

1263 Kellmeyer 2024:251

hen.¹²⁶⁴ Zur Beantwortung dieser Frage gibt es aber dem auf Technologierecht spezialisierten Lachlan Urquhart und dem Technologieforscher Peter J. Craigon zufolge nicht nur einen Ansatz: „(...) there are many approaches to doing ‘ethics by design’, often using different labels (...).“¹²⁶⁵ Es gibt nicht nur verschiedene Ansätze zur ethischen Gestaltung von Technologien mit vielen verschiedenen Namen, sondern es werden sogar die gleichen Bezeichnungen für unterschiedliche Herangehensweisen verwendet. So bedeutet Ethics by Design laut der Informatikerin Virginia Dignum und ihren Kolleg:innen, dass Maschinen so programmiert werden können, dass sie sich nach ethischen Grundsätzen verhalten.¹²⁶⁶ Laut Mulvenna et al. erfordert der Ethics by Design Ansatz jedoch eine Zusammenarbeit zwischen Technik und Ethik während des gesamten Gestaltungsprozesses.¹²⁶⁷ Deshalb schlagen sie vor, dass eine ständige internationale Arbeitsgruppe eingerichtet wird, die auf der Basis von 12 Prinzipien Grundsätze ableitet.¹²⁶⁸

Ich halte es für gefährlich, Kriterium iii) nicht zu erfüllen, stattdessen die Verantwortung der Technik zu übertragen und zu behaupten, dass es möglich ist, Ethik einzuprogrammieren.¹²⁶⁹ Da eine Ethik komplexer ist als ein heruntergebrochener Verhaltenskodex, würde diese Programmierung oft einer ethischen Entscheidung nicht gerecht. Wenn man sich trotzdem darauf verlässt, können unethische Entscheidungen unbemerkt getroffen werden, welche die Nutzenden womöglich stark beeinflussen.

Leider bleibt bei Mulvenna et al. unklar, woher die 12 Prinzipien stammen und wie sie sich in einem Gestaltungsprozess umsetzen lassen, sodass hier Kriterium i) nicht erfüllt ist. Da jeweils mindestens eins der allgemeinen Kriterien nicht erfüllt wird, fehlt auch dem Ethics by Design Ansatz etwas Wesentliches bei der Umsetzung von ethischen Überlegungen in die Praxis.

1264 Dignum et al. 2018:64

1265 Urquhart, Craigon 2021:96

1266 Dignum et al. 2018:65

1267 Mulvenna et al. 2017:51, 53

1268 Ebd.:53f.

1269 Dignum et al. 2018:65

Human-centered Design

Human-centered Design bedeutet, dass der Mensch nicht nur als Teil des Systems betrachtet wird, sondern in jedem Aspekt des Designs eine zentrale Rolle spielt. Dabei fungiert der Designer als Übersetzer von gesellschaftlichen Bedürfnissen in das Design von KI-Systemen.¹²⁷⁰ Der KI-Ethiker Philipp Kellmeyer ist davon überzeugt, dass durch eine Einbeziehung aller Stakeholder in den Entwicklungsprozess (Stakeholder-Befragung) die daraus resultierenden Technologien technisch solide sind und wirklich auf die Bedürfnisse und Werte derer eingehen, denen sie helfen sollen.¹²⁷¹ Die Umsetzung dieser Einbindung der Stakeholder in den Gestaltungsprozess erfordere jedoch eine engagierte Gemeinschaft von Stakeholdern, eine inklusive Integration aller Stakeholder und den Ausbau ihrer Fähigkeiten durch iteratives Lernen.¹²⁷²

Der Human-centered Design Ansatz hat die Funktion, auf die Rolle der Menschen im Gestaltungsprozess aufmerksam zu machen. Dadurch sollen ethische Anforderungen an zu gestaltende Technologien identifiziert werden. Das Problem bei der Methode der Stakeholder-Befragung ist jedoch ihre fehlende Skalierbarkeit. So können kontextspezifische Lösungen gefunden werden für die Gruppe der befragten Stakeholder, jedoch lassen sich diese Lösungen nur schwer auf größere Populationen übertragen.¹²⁷³ Es ist zwar das Ziel des Human-centered Designs, nur Technologien zu entwickeln, die den Einfluss der Technologie auf alle Betroffenen berücksichtigen (ii) und somit zum Vorteil für alle von der Technologie betroffenen Personen sind, es ist aber unklar, ob dieses Ziel mit Hilfe der Stakeholder-Befragung erreicht werden kann. Außerdem ist die Identifikation ethischer Anforderungen durch die Stakeholder Beteiligung zwar ein wichtiger erster Schritt zur ethischen Gestaltung von Technologien, aber hier fehlen konkrete Umsetzungsvorschläge (i). Und es bleibt unklar, wer die Verantwortung für die Umsetzung eines der Human-centered Design Ansätze tragen soll (iii). Da die allgemeinen Kriterien nicht (mit Sicherheit) erfüllt werden, reicht auch

1270 Auernhammer 2020:1320

1271 Kellmeyer 2024:252

1272 Ebd.:254

1273 Ebd.:257f.

der Human-centered Design Ansatz nicht aus zur Umsetzung von ethischen Überlegungen in die Praxis.

9.2.2 Privacy by Design

Ein konkreter Umsetzungsvorschlag findet sich in der Privacy by Design (PbD) Methode. Hier wird die Privatsphäre als ethische Anforderung an technische Systeme identifiziert. Die Methode stammt von Ann Cavoukian, der Informations- und Datenschutzbeauftragten der kanadischen Stadt Ontario: „Privacy by Design (PbD), the framework to proactively embed privacy directly into information technology, business practices, physical design, and networked infrastructures – making it the default.“¹²⁷⁴ Bei der Privacy by Design Methode sollen Datenschutzprinzipien proaktiv in ein System integriert werden, durch zum Beispiel Standardeinstellungen zum Schutz der Privatsphäre sowie Sicherheitsmaßnahmen. Zudem soll durch die Methode der Schutz der Privatsphäre in den Risikomanagementprozessen der Unternehmen anerkannt werden.¹²⁷⁵

Beim Schutz der Privatsphäre geht es darum, dass so wenige personenbezogene Daten wie möglich erhoben werden und der Einzelne die größtmögliche Kontrolle über seine personenbezogenen Daten hat. Laut Cavoukian muss eine Organisation alle Aspekte des Schutzes der Privatsphäre und der Informationssicherheit unter Kontrolle haben, um von sich behaupten zu können, dass sie die Fähigkeit hat, Rechenschaft ablegen zu können.¹²⁷⁶ Idealerweise sollten technische Systeme laut der Wirtschaftsinformatikerin Sarah Spiekermann so aufgebaut werden, dass die Kontrolle der Nutzenden maximiert und die Beteiligung von Dienstanbietern minimiert wird. Wenn dieses Ideal nicht umsetzbar ist, sollte der Schutz der Privatsphäre unterstützt werden durch technisch durchsetzbare Standardrichtlinien („Opt-out“-Einstellungen) oder Datenknappheitsrichtlinien, Datenübertragbarkeit und Benutzerzugriffs- und Löschrechte.¹²⁷⁷

1274 Cavoukian 2012:18

1275 Spiekermann 2012:38

1276 Cavoukian et al. 2010:409

1277 Spiekermann 2012:40

Die Ziele des „Privacy by Design“ können durch die Einhaltung von sieben Grundprinzipien erreicht werden: Die Technologie soll erstens proaktiv und präventiv anstatt reaktiv gestaltet werden, so dass Verletzungen der Privatsphäre von vorneherein verhindert werden können. Zweitens gilt der Schutz der Privatsphäre als Standard, das heißt, die Nutzenden müssen nicht aktiv ihre Privatsphäre schützen, denn der Schutz der Privatsphäre ist drittens bereits in das Design eingebettet. Dabei wird aber viertens die Funktionalität des Systems nicht beeinträchtigt, da andere Geschäftsinteressen mit dem Schutz der Privatsphäre koexistieren. Fünftens sind starke Sicherheitsmaßnahmen als Grundlage für den Schutz der Privatsphäre unerlässlich. Des Weiteren sind sechstens Sichtbarkeit und Transparenz unerlässlich, um Verantwortlichkeit und Vertrauen zu schaffen. Das siebte Grundprinzip von Privacy by Design ist der Respekt für die Privatsphäre der Nutzenden, denn die Gestaltung von Technologie sollte bewusst auf die Interessen und Bedürfnisse der einzelnen Nutzenden abgestimmt sein, sodass diese ihre eigenen personenbezogenen Daten verwalten können.¹²⁷⁸

Es gibt laut Cavoukian unendlich viele Möglichkeiten, wie Unternehmen den Schutz der Privatsphäre kreativ in ihre Abläufe und Produkte integrieren können¹²⁷⁹: „(...) the task for privacy-aware engineers and systems architects is to translate the PbD conceptual framework into a set of specific, and operationally feasible, tools.”¹²⁸⁰ Die bereits erwähnten Marie Caroline Oetzel und Sarah Spiekermann haben sich dieser Aufgabe angenommen und das Privacy Impact Assessment (PIA) entwickelt. Der PIA-Rahmen schlägt konkrete Ziele für den Schutz der Privatsphäre vor und beschreibt sieben Schritte, um diese zu erreichen.¹²⁸¹ Bei der Ausführung von PIAs werden Bedrohungen der Privatsphäre untersucht und dazu angeregt, schon in einem frühen Stadium des Gestaltungsprozesses Schutzmaßnahmen einzuplanen und diese fortlaufend zu überprüfen. So können Probleme beim Schutz der Privatsphäre während der

1278 Cavoukian et al. 2010:409f.

1279 Ebd.:413

1280 Cavoukian 2012:19

1281 Spiekermann 2012:39

Produktentwicklung erkannt und proaktiv Maßnahmen zur Verbesserung des Schutzes in Systeme eingebaut werden.¹²⁸²

Dank der konkreten Grundprinzipien von Cavoukian und des Privacy Impact Assessments von Oetzel und Spiekermann, bietet die Privacy by Design Methode eine praktische Anleitung zur Umsetzung des Wertes der Privatsphäre in die Gestaltung von Technologien (i). Der Einfluss der Technologie auf jede Person, deren Privatsphäre durch die Technologie betroffen ist, wird berücksichtigt (ii) und die Zuschreibung von Verantwortung (iii) für die einzelnen Schritte fällt dem die PIA durchführenden Unternehmen zu.¹²⁸³ Es wird also durch die Privacy by Design Methode nicht klar, wem die Verantwortung für die Gestaltung zugeordnet wird, sondern nur wer sie zuordnen soll. Selbst wenn man das durchgehen ließe, und damit alle allgemeinen Kriterien erfüllt wären, scheitert die Privacy by Design Methode an mindestens einem der spezifischen Kriterien für die ethische Gestaltung eines Datenökosystems. Denn die Privacy by Design Methode betrachtet zwar ausführlich die Umsetzung der Privatsphäre (II), jedoch spielen andere im Datenökosystem relevante Werte wie die Transparenz keine Rolle. Deshalb ermöglicht es die Privacy by Design Methode nicht, die administrativen Anforderungen zur Umsetzung von Transparenz für die Datennutzenden zu erfüllen, sodass Kriterium III nicht erfüllt werden kann. Aus dem Grund wird auch die Privacy by Design Methode nicht weiterverfolgt als Lösung zur Umsetzung von ethischen Überlegungen in die Praxis.

9.2.3 Responsible Innovation

Ein weiterer konkreter Umsetzungsvorschlag dazu, wie sich technische Entwicklungen an ethischen Maßstäben ausrichten lassen, stellt die Responsible Innovation Methode dar (auch als „Responsible Research and Innovation“ bezeichnet). Ursprünglich bezieht sich der Begriff der Innovation auf Neuerungen aller Art.¹²⁸⁴ Mittlerweile

1282 Oetzel, Spiekermann 2014:126ff.

1283 Ebd.:142

1284 Van den Hoven 2013:80

wird der Innovationsbegriff jedoch den bereits erwähnten Technologiephilosophen Vincent Blok und Pieter Lemmens zufolge nur noch bei der Erforschung neuer Technologien und der kommerziellen Nutzung dieser neuen Technologien eingesetzt.¹²⁸⁵ Bei verantwortungsvoller Innovation (Responsible Innovation) geht es darum, soziale und ethische Aspekte der Erforschung und Nutzung von Technologien ausdrücklich zu berücksichtigen und sie mit wirtschaftlichen, soziokulturellen und ökologischen Aspekten in ein Gleichgewicht zu bringen.¹²⁸⁶ Die Berücksichtigung von ethischen und gesellschaftlichen Bedenken soll dadurch umgesetzt werden, dass die von dem Innovationsprozess direkt Betroffenen (Stakeholder) und Vertreter der Öffentlichkeit in den Innovationsprozess integriert werden.

Verantwortung wird im Rahmen der Responsible Innovation Methode als proaktiv und nicht als reaktiv betrachtet, das heißt, es geht nicht um Rechenschaftspflicht für potenziell unerwünschte Ergebnisse, sondern um aktive Gestaltung von Innovationsprozessen für gesellschaftlich wünschenswerte Ziele.¹²⁸⁷ Laut dem bereits erwähnten Technologieforscher Owen et al. ist Responsible Innovation philosophisch verankert in den vorausschauenden Dimensionen der Verantwortung: Sorgfalt und Reaktionsfähigkeit (care und responsiveness). Diese Dimensionen betonen den Technologieforscher:innen Mirjam Burget, Emanuele Bardone und Margus Pedaste zufolge, dass es nicht um die Ergebnisse geht, sondern um den Prozess selber: „(...) *taking care* describes more an attitude rather than a result.“¹²⁸⁸ Reaktionsfähigkeit soll bedeuten, dass Reflexion und Deliberation mit Handlungen verknüpft werden, die einen wesentlichen Einfluss auf die Innovation haben.¹²⁸⁹

Owen et al. definieren im Jahr 2013 Responsible Innovation folgendermaßen: „Responsible innovation is a collective commitment of care for the future through responsive stewardship of science and innovation in the present.“¹²⁹⁰ Diese kollektive Verpflichtung

1285 Blok, Lemmens 2015:28

1286 Jakobsen et al. 2019:2339; Blok, Lemmens 2015:20

1287 Stahl et al. 2021:177

1288 Burget et al. 2017:14

1289 Owen et al. 2013:29

1290 Ebd.:36

zur Sorge für die Zukunft lässt sich bei der Entwicklung neuer Technologien umsetzen, indem der Innovationsprozess antizipativ, reflexiv, deliberativ und responsiv gestaltet wird.¹²⁹¹ Der Gestaltungsprozess ist antizipativ, wenn die Folgen des zu gestaltenden Produktes analysiert werden. Er ist reflexiv, wenn ethische Implikationen des Produktes analysiert werden. Der Gestaltungsprozess ist deliberativ, wenn die Perspektiven, Visionen, Fragen und Probleme der Stakeholder in Bezug auf das Produkt berücksichtigt werden. Und der Prozess ist responsiv, wenn ethische Probleme zu Veränderungen im Prozess und im Produkt führen.¹²⁹² Dadurch, dass der verantwortungsvolle Innovationsprozess und insbesondere die Eigenschaft der Deliberation inklusiv ist, wird zumindest im Prinzip allen Menschen die Teilhabe ermöglicht.¹²⁹³

Um wirksam zu sein, darf es sich bei den vier Kernanforderungen (Antizipation, Reflexivität, Deliberation und Responsivität) nicht nur um Grundsätze handeln, sie müssen institutionell eingebettet sein in die Innovation und in den Innovationsprozess.¹²⁹⁴ Eine konkrete Umsetzung kann jedoch flexibel an den Anwendungskontext angepasst werden. So könnte zum Beispiel das Team, das den Innovationsprozess voranbringt, zum verantwortungsvollen Innovieren gebracht werden, in dem in Entscheidungsprozessen immer wieder die passenden Fragen gestellt werden: Woran arbeiten Sie gerade? (Reflexiv) Warum arbeiten Sie daran? (Reflexiv, Deliberativ) Wie könnten Sie es anders angehen? (Responsiv) Wer könnte in Zukunft betroffen sein? (Antizipativ).¹²⁹⁵

Die Antworten des Teams könnten dann wiederum von einem unabhängigen Gremium bewertet werden anhand von fünf Kriterien, mit denen die Kernanforderungen der Responsible Innovation geprüft werden. Erstens müssen die Hauptrisiken ermittelt, verwaltet und für akzeptabel befunden werden (Reflexiv), zweitens muss das Vorgehen den einschlägigen Vorschriften entsprechen (Reflexiv), drittens muss allen relevanten Parteien klar mitgeteilt werden, welchen Zweck die Innovation erfüllt, um die Stakeholder zu informie-

1291 Owen et al. 2013:44

1292 Niehaves, Heger 2015:172; Owen et al. 2013:38

1293 Burget et al. 2017:14

1294 Owen et al. 2013:39

1295 Ebd.:42

ren und eine ausgewogene Diskussion zu fördern (Reflexiv, Deliberativ), viertens müssen künftige Anwendungen und Auswirkungen beschrieben und bei neuen Informationen überprüft werden (Antizipativ, Reflexiv), fünftens müssen Mechanismen ermittelt werden, um die Ansichten der Öffentlichkeit und der Interessengruppen zu den möglichen Anwendungen und Auswirkungen zu verstehen (Deliberativ, Reflexiv).¹²⁹⁶

Die Wirtschaftswissenschaftler Rob Lubberink, Johan van Ophem und Onno Omta sowie der Technologiephilosoph Vincent Blok entwickeln aus einer Literaturanalyse eine andere Strategie zur Operationalisierung von Responsible Innovation. Sie übersetzen die vier Kernanforderungen (zuzüglich Inklusion und Wissensmanagement) in eine Liste von wichtigen Aktivitäten und dazugehörigen Strategien.¹²⁹⁷ Die Kernanforderung der Antizipation umfasst dabei das frühzeitige Erkennen erwünschter und unerwünschter Auswirkungen von Innovationen durch die Beobachtung gesellschaftlicher und ökologischer Bedürfnisse, um daraus konkrete Innovationsanforderungen abzuleiten.¹²⁹⁸ Reflexivität bedeutet, dass Unternehmen kritisch über ihre eigenen Werte, Motive, Handlungen und Wissenslücken nachdenken und deren Einfluss auf den Innovationsprozess analysieren.¹²⁹⁹ Deliberation wird durch den offenen, strukturierten Austausch zwischen Stakeholdern auf Basis von genauen und transparenten Informationen unter Berücksichtigung der Interessen aller Stakeholder erreicht.¹³⁰⁰ Reaktionsfähigkeit wird erreicht, indem Organisationen durch den Abbau bürokratischer Hürden und der Neuausrichtung interner Rollen die Innovationsprozesse flexibel an veränderte Bedingungen und Bedürfnisse anpassen.¹³⁰¹

Lubberink et al. fügen den vier Kernanforderungen der Responsible Innovation außerdem die Anforderungen der Inklusion und des Wissensmanagements hinzu: Inklusion bezieht sich auf die Einbindung vielfältiger Stakeholder in den Innovationsprozess, wobei eine faire Verteilung von Aufgaben und Nutzen sowie die Stärkung von

1296 Stilgoe et al. 2013:1575; Owen et al. 2013:43

1297 Lubberink et al. 2017:11–18

1298 Ebd.:11

1299 Ebd.:13

1300 Ebd.:15

1301 Lubberink et al. 2017:16f.

Beziehungen im Fokus stehen.¹³⁰² Wissensmanagement umfasst die gezielte Gewinnung und Integration relevanten Wissens in den Innovationsprozess, um Innovationen im Einklang mit den Bedürfnissen der Stakeholder zu entwickeln – dies wird unterstützt durch Austauschplattformen und flache Hierarchien.¹³⁰³ Lubberink et al. argumentieren, dass das Übersetzen der Kernanforderungen von Responsible Innovation in die wichtigsten Aktivitäten und die dazugehörigen Strategien dabei helfen kann, alle von der Innovation betroffenen Akteure einzubeziehen und dadurch den Innovationsprozess so zu steuern, dass letztlich nachhaltige, gesellschaftlich wünschenswerte und ethisch akzeptable Lösungen erzielt werden.¹³⁰⁴

Es ist deutlich erkennbar, dass die Responsible Innovation Methode konkrete praktische Anleitung bietet (i), weil auf Grundlage ihrer Kernanforderungen bestimmte Leitfragen und Umsetzungsstrategien entwickelt wurden. Das Ziel des Berücksichtigens des Einflusses der Technologie auf alle Betroffenen soll durch das Einbeziehen der Stakeholder erreicht werden (ii) – hier gibt es jedoch Umsetzungsschwierigkeiten in der Praxis. Die Betriebswirtschaftlerinnen Luciana Maines da Silva und Claudia Cristina Bitencourt stellen zusammen mit Kadѓgia Faccin und Tatiana Iakovleva aus dem Innovationsmanagement bei einer Literaturanalyse von empirischen Studien zur Umsetzung von Responsible Innovation fest, dass die Stakeholder häufig erst einbezogen werden, wenn die Innovation bereits auf dem Markt ist.¹³⁰⁵ Auch dem Technologiephilosophen und -ethiker Bernd Carsten Stahl und seinen Kolleg:innen fällt beim Erproben der Umsetzung dieser Methode auf, dass Responsible Innovation trotz der ursprünglichen Intention, Stakeholder von Anfang an einzubeziehen, in der Praxis oft erst dann eingesetzt wird, wenn die wichtigsten Entscheidungen bereits getroffen sind.¹³⁰⁶ Problematisch ist das, weil auf diese Weise das Feedback von Stakeholdern nur eingeschränkt berücksichtigt werden kann und womöglich nicht alle

1302 Lubberink et al. 2017:14

1303 Ebd.:17f.

1304 Ebd.:23

1305 Silva et al. 2019:16

1306 Stahl et al. 2021:189

Risiken antizipiert werden können.¹³⁰⁷ Die Umsetzung bleibt also hinter dem Anspruch zurück bei der Entwicklung der Technologie ihren Einfluss auf alle Betroffenen zu berücksichtigen und dadurch einen Vorteil für alle von der Technologie betroffenen Personen zu bewirken (ii). Insbesondere auch, weil es in der Praxis schwierig ist nachzuvollziehen, ob die Deliberation mit den Stakeholdern direkte Auswirkungen auf die Innovation hat und ob die Ergebnisse des Innovationsprozesses den Wünschen der Stakeholder entsprechen.¹³⁰⁸

Außerdem ist auch die Zuordnung der Verantwortung bei der Umsetzung der Responsible Innovation Methode unklar (iii). Denn laut Lubberink et al. führt die Beratung mit den Stakeholdern zu einer kollektiven Verantwortung.¹³⁰⁹ Ich denke jedoch, Stakeholder sollten nicht verantwortlich gemacht werden, solange nicht nachvollziehbar ist, ob ihre Deliberation wirklich Auswirkungen hat auf die daraufhin entwickelte Technologie.¹³¹⁰ Owen et al. schlagen hingegen eine Co-Verantwortung vor: Bei der Umsetzung von Responsible Innovation sollen nicht nur die Innovator:innen eine wichtige Rolle spielen, sondern auch die Institutionen, in denen Innovationen gestaltet werden, sowie diejenigen, die Innovation fördern. Insbesondere die Fördernden haben laut Owen et al. eine einzigartige Position, um die Responsible Innovation institutionell zu verankern.¹³¹¹ So fällt die Verantwortung zur Umsetzung der Responsible Innovation allen am Entwicklungsprozess Beteiligten zu – das lässt aber auch befürchten, dass niemand sich konkret verantwortlich fühlt. Aus meiner Sicht müsste klar sein, welcher bestimmte menschliche Akteur dafür zuständig ist, die vier Kernanforderungen institutionell einzubetten, damit Kriterium iii) erfüllt wäre. Da die allgemeinen Kriterien ii) und iii) nicht erfüllt werden, erübrigt es sich die spezifischen Kriterien zu betrachten. So weist auch die Responsible Innovation Methode Mängel auf bei der Umsetzung ethischer Überlegungen in die Praxis.

1307 Silva et al. 2019:16

1308 Stahl et al. 2021:180f.

1309 Lubberink et al. 2017:2

1310 Stahl et al. 2021:180f.

1311 Owen et al. 2013:46

9.2.4 Value-Sensitive Design

Die Informationswissenschaftlerin Batya Friedman hat als erste die Methode des Value-Sensitive Design (VSD) vorgebracht und meint damit, dass man sich beim Gestalten und Implementieren von Computer Technologien auf menschliche Werte fokussieren sollte. Sie ruft schon 1997 dazu auf, dass Value-Sensitive Design als Teil der Kultur der Informatik begriffen werden soll und dass es als Kriterium dienen sollte, nach dem Systeme als mangelhaft und Designer als nachlässig beurteilt werden können.¹³¹² Werte verstehen sie und ihre Kollegen im Jahr 2008 als etwas, das eine Person oder eine Gruppe von Menschen im Leben für wichtig hält. Somit hängen Werte wesentlich von den Interessen und Wünschen der Menschen je nach kulturellem Milieu ab.¹³¹³

Bei dieser Umsetzungsmethode orientiert sich der Gestaltungsprozess an als wichtig und erstrebenswert empfundenen Werten mit Hilfe einer integrativen und iterativen, dreiteiligen Methodik, die aus *konzeptionellen*, *empirischen* und *technischen* Untersuchungen besteht. So wird *konzeptionell* untersucht, welche Stakeholder es gibt und welche Werte wichtig sein könnten. Dabei werden sorgfältige Überlegungen darüber angestellt, wie sich die eigenen technologischen Entwürfe auf die Stakeholder auswirken könnten. Die als wichtig identifizierten Werte müssen dann sorgfältig konzeptualisiert werden, da ein Begriff sehr unterschiedlich verstanden werden kann. Eine gut ausgearbeitete Definition der betreffenden Werte klärt dann schon die grundlegenden Fragen, die das jeweilige Projekt aufwirft, und bietet eine Grundlage für den Vergleich der Ergebnisse verschiedener Forschungsteams. *Empirisch* wird untersucht, welche Werte die Stakeholder für wichtig halten und durch welche Gestaltungsentscheidungen welche Werte realisiert werden können: Friedman beschreibt das mit dem bereits erwähnten Psychologen Peter H. Kahn Jr. und dem Informatiker Alan Borning so: „Empirical investigations encompass any human activity that can be observed, measured, or documented.“¹³¹⁴ Auf diese Weise kann auch der Erfolg und die Nutzbarkeit einer bestimmten Gestaltung bewertet

1312 Friedman 1997:12f.

1313 Friedman et al. 2008:70f.

1314 Friedman et al. 2002:3

werden. *Technisch* wird überprüft, wie bestimmte Eigenschaften von Technologien menschliche Werte unterstützen oder behindern, und es werden technische Lösungen gesucht, um die Werte umsetzen zu können. Die Annahme dabei ist, dass sich bestimmte Technologien für die Umsetzung bestimmter Werte eignen – diese Eignung folgt aus den Eigenschaften der Technologien.¹³¹⁵

Die drei Teile der Methodik sind voneinander abhängig und beeinflussen sich gegenseitig: Es gibt keine vorgeschriebene Reihenfolge für die drei Teile. Jedoch halten es die Wirtschaftsinformatiker:innen Till Winkler und Sarah Spiekermann für ratsam, erst konzeptionell zu arbeiten und dann empirisch zu analysieren, bevor technisch untersucht wird – insbesondere wenn für Wiederholungen des Prozesses die Zeit fehlt.¹³¹⁶

Value-Sensitive Design befasst sich in erster Linie mit Werten, die sich auf menschliches Wohlbefinden, Würde, Gerechtigkeit, Wohlfahrt und Rechte beziehen. Solche Werte haben laut dieser Methode einen moralisch-epistemischen Stellenwert, unabhängig davon, ob eine bestimmte Person oder Gruppe diese Werte vertritt – ihre Ausgestaltung kann aber in unterschiedlichen Kulturen oder zu verschiedenen Zeitpunkten variieren.¹³¹⁷ Friedman und Kahn haben die folgenden zwölf Aspekte entwickelt, welchen sie als spezifische menschliche Werte mit ethischer Bedeutung einen ausgeprägten Anspruch auf Ressourcen im Designprozess zuordnen: Menschliches Wohlergehen, Eigentum und Besitz, Privatsphäre, Unvoreingenommenheit, universelle Nutzbarkeit, Vertrauen, Autonomie, informierte Zustimmung, Verantwortlichkeit, Identität, Gelassenheit und ökologische Nachhaltigkeit.¹³¹⁸

Laut Friedman und Kahn sind die ersten neun Aspekte traditionell und beruhen auf deontologischen und konsequentialistischen moralischen Orientierungen. Problematisch ist, dass diese beiden Orientierungen ganz unterschiedliche Werte hervorbringen – in Entscheidungssituationen beziehen sie sich sogar oft auf widersprüchliche Werte. Außerdem gibt es in keinem der beiden einen

1315 Friedman et al. 2008:71–73; Friedman et al. 2002:2f.; Friedman, Kahn 2008:1251; Niehaves, Heger 2015:173f.; Winkler, Spiekermann 2018:17

1316 Winkler, Spiekermann 2018:19

1317 Friedman, Kahn 2008:1250f.

1318 Ebd.:1251

abgeschlossenen Wertekanon, sodass sehr viele Werte als traditionell bezeichnet werden können. Die letzten drei Werte seien nicht traditionell innerhalb der Gemeinschaft von Designern und Forschern, die sich mit der Interaktion von Menschen und Computern auseinandersetzen.¹³¹⁹ Hier besteht das Problem, dass es innerhalb der Gemeinschaft von Designern und Forschern, die sich mit der Interaktion von Menschen und Computern auseinandersetzen, noch viel mehr Werte gibt, die als nicht traditionell gelten können. Außerdem behaupte ich, dass informierte Zustimmung gar kein Wert ist – es handelt sich vielmehr um eine Maßnahme, die Werte wie Privatsphäre oder Autonomie schützen kann.

Wie schon bei den Ethischen Leitlinien, könnten die von mir aus der epistemischen Asymmetrie eines Datenökosystems hergeleiteten Prinzipien Privatsphäre und Transparenz als Werte gelten, an denen sich ein ethischer Gestaltungsprozess ausrichten sollte. Das heißt, die konzeptionelle Untersuchung der wichtigen Werte könnte sich auf die in dieser Arbeit identifizierten Konzepte stützen. Denn dank der dreiteiligen Methodik bietet die Value-Sensitive Design Methode konkrete praktische Anleitung (i) und es wird der Einfluss der zu entwickelnden Technologie auf alle Stakeholder sowohl konzeptionell als auch empirisch berücksichtigt (ii). Die empirische Berücksichtigung der Werte der Stakeholder bereitet der Value-Sensitive Design Methode jedoch ähnliche Schwierigkeiten, wie der Responsible Innovation Methode. So haben Winkler und Spiekermann bei der Analyse von 17 Studien, welche die Anwendung der Value-Sensitive Design Methode abbilden, festgestellt, dass oft kein guter methodischer Ansatz zur Identifizierung der Stakeholder, sondern nur grob definierte Stakeholder-Gruppen verwendet wurden. Hierdurch können Stakeholdern und ihre Werte unabsichtlich ausgeschlossen werden.¹³²⁰ Das gefährdet die Erfüllung von Kriterium ii).

Außerdem ist auch Kriterium iii) nicht eindeutig erfüllt, denn die Angaben dazu, wer die Verantwortung für die Umsetzung der ethischen Überlegungen tragen soll, variieren. Einerseits sollen Programmierende durch Value-Sensitive Design Sozialinformatik prak-

1319 Friedman, Kahn 2008:1251

1320 Winkler, Spiekermann 2018:19

tizieren und anwenden können¹³²¹, andererseits haben in den Anwendungsbeispielen scheinbar Forscher:innen die Value-Sensitive Design Methode durchgeführt¹³²². An wieder anderer Stelle wird Fachleuten für Mensch-Computer-Interaktion (human-computer interaction professionals) auf Grund ihrer beruflichen Stellung eine besondere ethische Verantwortung zugeschrieben.¹³²³ Da also die allgemeinen Kriterien für die Umsetzungsmethode zur ethischen Technologieentwicklung vom Value-Sensitive Design nicht alle erfüllt werden, erübrigt es sich die spezifischen Kriterien zur ethischen Gestaltung eines Datenökosystems zu betrachten. Deshalb wird auch die Value-Sensitive Design Methode im Folgenden nicht genutzt, um die Umsetzung von ethischen Überlegungen in die Praxis anzuleiten.

9.3 Risikoanalyse

Eine dritte mögliche Kategorie von Umsetzungsmethoden zur Ausrichtung technischer Entwicklungen an ethischen Maßstäben ist die Betrachtung der Risiken, die durch den Einsatz der Technik entstehen können. Der Informatiker Peter G. Neumann meint, dass technische Risiken durch langfristige Planung vermieden werden können. Es sollten also nicht Risiken für hinreichend unwahrscheinlich erklärt werden oder nur Versicherungsschutz angestrebt werden, sondern es sollte proaktiv gehandelt werden, indem die Systeme mit Bedacht entwickelt werden.¹³²⁴ Im Folgenden werden drei verschiedene Methoden diskutiert, welche auf einer Risikoanalyse basieren: die Folgenabschätzung, das interne Audit und die ethische Risikoanalyse.

9.3.1 Folgenabschätzung

Der bereits erwähnte Technologierechtler Lachlan Urquhart und der Technologieforscher Peter J. Craigon haben ein praktisches

1321 Friedman et al. 2002:7

1322 Ebd.:4f.

1323 Friedman, Kahn 2008:1242, 1259

1324 Neumann 2009:35–38

Instrument und eine Methodik entwickelt, um Entwicklungsteams bei der Reflexion über normative Aspekte der Technologieentwicklung zu unterstützen. Die Autoren bezeichnen ihre Methode als ein Werkzeug zur Umsetzung von Ethics by Design, aber ich ordne ihre Methode unter Risikoanalysen ein, da seine Grundlage eine Folgenabschätzung ist. Folgenabschätzungen (Impact Assessments) für neue Technologien sind laut Urquhart und Craigon ein beliebtes Instrument zur Vorhersage von Problemen und zur strukturierten Entwicklung von Handlungsstrategien. Die Autoren schlagen zur Durchführung einer solchen Folgenabschätzung die Nutzung eines Spielbretts vor, das sie Moral-IT Impact Assessment Board nennen. Für die Folgenabschätzung wurden vier relevante Phasen identifiziert: Erstens die Ermittlung möglicher Risiken; zweitens die Bewertung der Bedeutung jedes der Risiken und der Wahrscheinlichkeit seines Eintretens; drittens die Festlegung geeigneter Schutzmaßnahmen für diese Risiken; und viertens die Untersuchung der praktischen Umsetzungsprobleme.¹³²⁵ Diese vier Phasen sollen anhand der zu entwickelnden Technologie durchgespielt werden, sodass geeignete Antworten und Aktionspläne ausgearbeitet werden können.¹³²⁶

Dabei helfen soll das Moral-IT Kartendeck. Hierbei handelt es sich um einen Satz physischer Karten mit Fragen zu der Technologie, welche die Werte und Praktiken hinterfragen, die in technologische Systeme eingebettet sind. Der Text auf den Karten soll provozieren, um einen Ausgangspunkt für Diskussionen bilden zu können. Die Karten stellen ein Reflexionsinstrument dar, welches Entwicklungsteams dabei unterstützen soll, sich ihrer ethischen Verantwortung bewusst zu werden. Die Value-Sensitive Design Methode hat auf der Suche nach Umsetzungswegen die Gestaltung der Karten beeinflusst, und unterstützt damit Entwicklungsteams beim Nachdenken über Werte im Design.¹³²⁷ Das Moral-IT Kartendeck und das Moral-IT Spielbrett zur Folgenabschätzung unterstützen auch die Methode der Responsible Innovation. Denn in einem deliberativen Prozess (deliberation) werden ethische Fragen antizipiert (anticipation), zum Nachdenken über deren Auswirkungen angeregt (reflection) und zum Antworten und Pläneformulieren aufgefordert

1325 Urquhart, Craigon 2021:101

1326 Ebd.:95

1327 Ebd.:94–104

(responsiveness).¹³²⁸ Das Neue an ihrer Methode ist laut Urquhart und Craigon die Kombination der Folgenabschätzung mit einem Reflexionsinstrument wie dem Moral-IT Kartendeck. Hierdurch soll es möglich sein, während der Planung konkurrierende Interessen auszugleichen und schwierige moralische Entscheidungen zu treffen.¹³²⁹

Das Moral-IT Kartendeck und das Spielbrett für die Folgenabschätzung bieten konkrete praktische Anleitung (i). Laut Urquhart und Craigon sind es die Entwicklungsteams, die sich ihrer ethischen Verantwortung bewusst werden und mit Hilfe der Folgenabschätzung über Werte in der Gestaltung von Technologien nachdenken sollten (iii). Aber es bleibt unklar, ob der Einfluss der Technologie auf alle Betroffenen berücksichtigt wird (ii), denn die Folgenabschätzung kann kritisiert werden als Methode, die sich ausschließlich auf die Risikobewertung stützt und die potenziellen Auswirkungen auf die Gesellschaft oder die Umwelt kaum berücksichtigt.¹³³⁰ Laut Hermansson und Hansson liefert die klassische Risikoanalyse nicht alle notwendigen Informationen, denn sie fokussiert sich auf Wahrscheinlichkeiten und Schweregrade von Ergebnissen und missachtet ethische Aspekte des Handelns und der zwischenmenschlichen Beziehungen: „In order to appraise an action of risk-taking or risk imposition from a moral point of view, it is not sufficient to know the values and probabilities of its possible outcomes.”¹³³¹ Bei der Folgenabschätzung fehlt die Information, wer die Handlung durchführt und mit welchen Absichten.¹³³² Denn wenn Risiko und Nutzen nur anhand der Wahrscheinlichkeit des Eintretens und der Schwere der Konsequenzen abgewogen werden, ist es möglich, dass einzelne Menschen bei der Vorkehrung von Risiken als finanzielle Bürde bewertet werden und nicht der Wert eines jeden betroffenen Menschen betrachtet wird.¹³³³ Da somit das allgemeine Kriterium ii) nicht erfüllt wird, ist es obsolet die spezifischen Kriterien zu betrachten und somit weist auch die Folgenabschätzung gemäß meiner Kriterien Mängel auf.

1328 Owen et al. 2013:44; Urquhart, Craigon 2021:97

1329 Urquhart, Craigon 2021:95, 97

1330 Owen et al. 2013:41

1331 Hermansson, Hansson 2007:130

1332 Hansson 2018:1820f.; Hermansson, Hansson 2007:130

1333 Lincke 2016:78

9.3.2 Internes Audit

Audits sind Instrumente zur Überprüfung komplexer Prozesse. Hierdurch kann festgestellt werden, ob Prozesse mit den Unternehmensrichtlinien, Industrienormen oder Vorschriften übereinstimmen. Die Informatikerin und Aktivistin Inioluwa Deborah Raji und ihre Kolleg:innen schlagen vor, interne Algorithmus-Audits zu nutzen, um zu überprüfen, ob die Prozesse der Entwicklung von KI-Systemen den erklärten ethischen Erwartungen und Standards der Organisation entsprechen. Mit Hilfe eines solchen Audits können potenzielle negative Folgen im Vorhinein antizipiert werden. Anschließend bieten die Auditergebnisse Entscheidungshilfen für die Entwicklung von Abhilfemaßnahmen oder auch für den Abbruch der Technologieentwicklung, falls die Risiken den Nutzen überwiegen. Interne Audits werden während der Produktentwicklung und vor der Markteinführung durchgeführt. Die Durchführung wird angeleitet durch ein Audit-Team, aber auch das Produktteam, die Geschäftsleitung und andere Stakeholder sind an dem Audit beteiligt.¹³³⁴

Das interne Audit dient der Bewertung und Optimierung im Hinblick auf den gesellschaftlichen Nutzen und erklärte ethische Erwartungen und Standards der Organisation. Diese können zum Beispiel KI-Grundsätze sein, denen sich das betreffende Unternehmen verschrieben hat.¹³³⁵ Interne Audits überprüfen die Übereinstimmung einer Technologie mit den erklärten KI-Grundsätzen noch vor ihrem Einsatz. Das funktioniert, indem eine Risikoanalyse durchgeführt wird, welche fragt, was bei dem Einsatz der Technologie misslingen könnte.¹³³⁶ Raji et al. wenden jedoch ein: „Risk-based assessments can be limited in their ability to capture social and ethical stakes, and they should be complemented by anticipatory

1334 Raji et al. 2020:33f.; Raji et al. reflektieren, dass interne Auditor*innen zwangsläufig ein Interesse haben an einem guten Verlauf der Prüfung ihrer Organisation. Sie empfehlen, dass sich die Auditor*innen ihrer eigenen Vorurteile und Standpunkte sowie derjenigen der Organisation bewusst werden, um zu vermeiden, dass die Prüfungen lediglich dem Reputationsmanagement dienen. Außerdem sollen deshalb laut Raji et al. interne Prüfungen nur ein Aspekt eines umfassenderen Systems von Qualitätskontrollen sein (Raji et al. 2020:42).

1335 Raji et al. 2020:34f.

1336 Ebd.:35, 37

questions such as, ‘what if...?’¹³³⁷ Deshalb ergänzen die Autor:innen die formale Risikobewertung um die Kernanforderung der Antizipation aus der Responsible Innovation Methode. Das Ziel ist dabei, die ethische Vorausschau zu verbessern, indem systematisch nachgedacht wird über das sozio-technische System, in dem ein Produkt eingesetzt wird.¹³³⁸

Der Rahmen für ein internes Audit, den Raji et al. entwerfen, besteht aus fünf Phasen: „(...) Scoping, Mapping, Artifact Collection, Testing and Reflection (SMACTR) (...)“¹³³⁹ In der Scoping-Phase werden Anwendungsfälle definiert sowie die Grundsätze und Werte bestätigt, die die Produktentwicklung leiten sollen. Zudem wird zum einen eine ethische Überprüfung der Anwendungsfälle durchgeführt, um die Frage zu beantworten, wer wahrscheinlich von der auditierten Technologie betroffen sein wird und auf welche Weise. Zum anderen werden die sozialen Auswirkungen bewertet (social impact assessment), indem die Schwere der Risiken eingeschätzt wird und die relevanten sozialen, wirtschaftlichen und kulturellen Auswirkungen und Schäden identifiziert werden.¹³⁴⁰ In der Mapping-Phase wird das Systemumfeld analysiert, werden Beteiligte dokumentiert und mithilfe der Fehlermöglichkeits- und Einflussanalyse (FMEA) potenzielle Fehlerquellen im Design identifiziert.¹³⁴¹ Um Transparenz im Entwicklungsprozess zu schaffen und eine fundierte Prüfung zu ermöglichen wird in der Artifact Collection Phase dokumentiert, wie das zu auditierende Modell erstellt wurde, welche Annahmen während der Entwicklung getroffen wurden und welche Arten von Modellverhalten bei verschiedenen kulturellen, demografischen oder phänotypischen Gruppen auftreten könnten, wie die Daten erhoben wurden und ob sie sich auf Personen beziehen.¹³⁴² Die Testing-Phase dient dazu, mit Hilfe von Simulationen und absichtlicher Fehlerauslösung zu prüfen, ob das System mit ethischen Werten vereinbar ist.¹³⁴³ In der Reflection-Phase wird abschließend eine Risikoanalyse in Bezug auf die algorithmische Nutzung (Algo-

1337 Raji et al. 2020:38

1338 Ebd.:38

1339 Ebd.:37

1340 Ebd.:39

1341 Ebd.:36

1342 Ebd.:40f.

1343 Ebd.:41

rhythmic Use-related Risk Analysis) durchgeführt, welche sich auf die soziale Folgenabschätzung (social impact assessment) aus der Scoping Phase stützen sollte. Außerdem werden Maßnahmen zur Risikominderung entwickelt und alle Nachweise für die Entwicklung des Produktes in Übereinstimmung mit den ethischen Werten der Organisation für den Auditbericht dokumentiert.¹³⁴⁴

Dadurch, dass die Risikoanalyse aus einer sozialen Folgenabschätzung (social impact assessment) besteht, wird deutlich, dass der Einfluss der Technologie auf alle Betroffenen berücksichtigt wird und somit der Vorteil für alle von der Technologie betroffenen Personen das Ziel sein soll (ii). Die Verantwortung für die Durchführung des internen Audits liegt eindeutig bei dem Audit-Team (iii). Doch die Verantwortung für das Vorlegen einer Erklärung oder Bestätigung der ethischen Ziele, Standards und KI-Prinzipien liegt beim Produkt- und Entwicklungsteam. Es wird aber nicht spezifiziert, wie diese Teams die ethischen Ziele, Standards und KI-Prinzipien identifizieren sollen. Ohne eine Anleitung lastet diese schwierige Aufgabe auf den Schultern von Mitarbeitenden, die auch noch andere Aufgaben haben und vermutlich keine Erfahrung haben im Identifizieren von ethischen Zielen. In ihrem Beispiel haben sich Raji et al. an den fünf am häufigsten genannten KI-Prinzipien aus dem Literatur Review von Jobin et al. orientiert.¹³⁴⁵ Meine Kritik an diesem Vorgehen ist der Einwand, dass die Häufigkeit der Nutzung bestimmter Prinzipien nicht allein ihre Begründung darstellen sollte.

Hier könnte man meinen, dass die ethischen Überlegungen dieser Arbeit wieder die Lösung sein könnten, wie bei den Ethischen Leitlinien und beim Value-Sensitive Design. Aber beim internen Audit geht es nicht darum konzeptionell alle für die Technologie und ihren Kontext wichtigen ethischen Werte zu identifizieren, sondern es geht darum die ethischen Werte der Organisation festzustellen und die Technologie daran auszurichten. Das heißt es bedarf eines Reflektionsprozesses innerhalb der betreffenden Organisation über die eigenen ethischen Ziele, Standards und KI-Prinzipien. Das interne Audit von Raji et al. bietet durch die klar beschriebenen fünf Phasen (SMACTR) konkrete praktische Anleitung (i), aber es fehlen

1344 Raji et al. 2020:41f.

1345 Jobin et al. 2019; Raji et al. 2020:38

Instruktionen für das Produkt- und Entwicklungsteam, wie es zu einer Erklärung oder Bestätigung der ethischen Ziele, Standards und KI-Prinzipien der betreffenden Organisation kommen kann.¹³⁴⁶

Wenn man die Methode des internen Audits uminterpretiert und dafür nutzt zu prüfen, ob die Technologie mit allgemeingültigen ethischen Werten übereinstimmt, anstatt mit den ethischen Werten der Organisation, könnten die Werte mit Hilfe der in dieser Arbeit entwickelten deontologischen Risikoethik begründet werden und es wären damit alle allgemeinen Kriterien erfüllt. Trotzdem würde die Methode des internen Audits an mindestens einem der spezifischen Kriterien für die ethische Gestaltung eines Datenökosystems scheitern: es wird nicht ermöglicht, die Auferlegung von Risiken zu regulieren (I). Das interne Audit betrachtet in der Scoping-Phase ausführlich wer wahrscheinlich von der auditierten Technologie betroffen sein wird und auf welche Weise, wie schwer die Risiken sind und welche sozialen, wirtschaftlichen und kulturellen Auswirkungen sie haben können und führt in der Reflection-Phase auf dieser Grundlage eine Risikoanalyse durch – dadurch wird die Risikoauferlegung problematisiert. Aber dabei wird nicht klar, unter welchen Umständen das Risiko zulässig ist¹³⁴⁷, sodass die Auferlegung von Risiken nicht reguliert werden kann. Da somit Kriterium I) nicht vollständig erfüllt werden kann, ist auch das interne Audit nicht ausreichend zur Umsetzung von ethischen Überlegungen in die Praxis.

9.3.3 Ethische Risikoanalyse

Hansson hat, angelehnt an das Drei-Parteien-Modell aus Kapitel 3.2.1, das er mit Hermansson entwickelt hat, eine ethische Risikoanalyse (eRA) entwickelt. Er kritisiert, dass der Fokus bei der modernen Risikoanalyse fast ausschließlich auf den Wahrscheinlichkeiten und der Schwere der unerwünschten Folgen liegt. Seine

¹³⁴⁶ Raji et al. 2020:39

¹³⁴⁷ Raji et al. schlagen vor eine Risikoschwelle zu definieren, unterhalb derer das Risiko als tolerierbar angesehen wird (Raji et al. 2020:42), aber es bleibt unklar, wie diese Schwelle definiert werden könnte. Außerdem wäre es, wie unter 4.3 erläutert, laut Hansson sehr schwierig, diese Grenze zu ziehen, weil eine pauschale Risikogrenze nicht in der Lage wäre, die Vorteile der jeweiligen risikobehafteten Handlung einzubeziehen (Hansson 2003:299).

eRA hingegen soll stattdessen die Fragen des Handelns und der zwischenmenschlichen Beziehungen betrachten, die den Kern von ethischen Diskussionen über Risiken bilden. Dabei soll die eRA von unparteiischen Akteuren durchgeführt werden, wie zum Beispiel akademischen, gemeinnützigen und in einigen Ländern staatlichen Organisationen. Um Risiken auf ethische Art und Weise zu analysieren, werden drei Schritte durchlaufen: Zuerst werden die Personen identifiziert, die eine primäre ethisch relevante Rolle in Bezug auf das Risiko haben. Dann werden ihre Rollen, Rollenkombinationen, Wechselbeziehungen und potenziellen Konflikte beschrieben. Abschließend werden diese Konflikte mit einem Schwerpunkt auf Fragen nach Gerechtigkeit, Rechten und sozialer Macht analysiert.¹³⁴⁸

Hansson kritisiert die Ansätze und Methoden, welche das Einbeziehen von Stakeholdern empfehlen (z.B. Human-centered Design, Responsible Innovation, Value Sensitive Design), da seiner Meinung nach Stakeholder aus instrumentellen oder strategischen Gründen in die Analyse einbezogen werden und nicht aus ethischen Gründen. Zudem unterscheidet die Stakeholder-Analyse seiner Ansicht nach nicht zwischen verschiedenen Gruppen von Stakeholdern: „From an ethical point of view, the large differences among stakeholders in terms of risk exposure, power, and resources have to be highlighted.“¹³⁴⁹ Außerdem nimmt die Stakeholder-Analyse typischerweise eine Insider-Perspektive ein, während eine ethische Risikoanalyse laut Hansson eine unparteiische Outsider-Perspektive einnehmen sollte. Und die Stakeholder-Analyse versucht einen Konsens zwischen den Beteiligten zu erzielen, während eine ethische Analyse auf die unterschiedliche Verhandlungsmacht der verschiedenen Stakeholder aufmerksam macht und nicht unbedingt einen Konsens anstrebt.¹³⁵⁰

In dem ersten Schritt einer ethischen Risikoanalyse müssen die verschiedenen Stakeholdergruppen identifiziert werden. Dabei gibt es laut Hansson drei unterschiedliche Formen der unmittelbaren Beteiligung von Menschen an einer Risikosituation, die schon aus dem Drei-Parteien-Modell aus Kapitel 3 bekannt sind: die Risikoexponierten, die Entscheidungstragenden und die durch das Risiko Be-

1348 Hansson 2018:1820f.

1349 Ebd.:1821

1350 Ebd.:1821f.

günstigten. Risikoexponierte können nicht immer individuell identifiziert werden, aber Gruppen, die von dem Risiko betroffen sind, können in die eRA miteinbezogen werden. Dabei sind die Gruppen in unterschiedlicher Weise und Höhe dem Risiko ausgesetzt: „They may also differ widely in their awareness of the risk, their capacity to make themselves heard, and the resources they can muster for furthering their interests.“¹³⁵¹ Dadurch, dass die Gruppen der Risikoexponierten identifiziert werden, kann sichergestellt werden, dass ihre Rechte und Interessen beachtet werden.

Eine Situation, die für die einen ein Risiko birgt, kann für andere von Vorteil sein. Es ist wichtig, diese Gruppe von Begünstigten zu identifizieren, um festzustellen, warum das Risiko in Kauf genommen wird. Zusätzlich werden im ersten Schritt der eRA die Entscheidungstragenden identifiziert: „(...) we are looking for those persons who (knowingly or due to ignorance for which they can be held responsible) make decisions that contribute causally to a risk exposure.“¹³⁵² Dies kann schwierig sein, da auch Hintergrundentscheidungen, zum Beispiel durch externe Hersteller und Regulierungsbehörden, die Bedingungen schaffen, welchen die Risikoexponierten ausgesetzt sind. Durch solche Hintergrundentscheidungen haben Risikoexponierte dann eventuell nur begrenzte Möglichkeiten, sich gegen das Risiko zu entscheiden.

Im zweiten Schritt der eRA wird betrachtet, ob manche Akteure mehrere dieser Rollen einnehmen, ob sich verschiedene Akteure gegenseitig beschränken und ob sie voneinander abhängig sind. Dabei stellt sich heraus, dass bestimmte Rollenkombinationen problematischer sind als andere. So ist es zum Beispiel ethisch hochproblematisch, wenn ein Akteur ausschließlich die Rolle der Risikoexponierten erfüllt, während andere Akteure über das Risiko entscheiden und davon profitieren. Hansson begründet das so: „(...) a disadvantage to one person is not necessarily outweighed by a larger advantage to someone else.“¹³⁵³ Selbst wenn bestimmte Akteure einen großen Vorteil von dem Risiko haben, ist es nicht unbedingt ethisch gerechtfertigt, dieses Risiko anderen Akteuren aufzuerlegen.

Weniger problematisch wird es, wenn die Risikoexponierten gleichzeitig auch die durch das Risiko Begünstigten sind. Hier sind

1351 Hansson 2018:1822

1352 Ebd.:1823

1353 Ebd.:1826

die Ergebnisse vorteilhaft, solange der Nutzen einen möglichen Schaden überwiegt. Ein Nachteil ist jedoch, dass die Entscheidung, welche diese Risikoexponierten betrifft, immer noch von anderen getroffen wird. Ein Beispiel dafür wäre eine paternalistische Entscheidung im medizinischen Kontext: Bewusstlose Patient:innen profitieren von einer medizinischen Maßnahme und tragen das Risiko, dass die erwünschte Wirkung ausbleibt, aber die Entscheidung treffen Ärzt:innen.

Eine andere Möglichkeit wäre es, dass die Risikoexponierten auch gleichzeitig die Entscheidungstragenden sind und zum Beispiel in einen Fluss springen, um eine andere Person vor dem Ertrinken zu retten. So lange sich der Akteur frei entscheiden kann, das Risiko einzugehen, ist diese Rollenkombination ethisch nicht problematisch – auch wenn der Akteur nicht selbst von dem Risiko profitiert.

Es ist sogar möglich, alle drei Rollen zu verkörpern. Zum Beispiel, wenn ein Akteur entscheidet, abseits der Pisten Ski zu fahren, nimmt er das erhöhte Risiko eines Sturzes aktiv auf sich, ist aber auch derjenige, der davon profitiert. Die Kombination aller drei Rollen klingt vorerst ideal, aber es können trotzdem noch andere Akteure in negativer Weise betroffen sein. Beim Skifahren abseits der Piste kann zum Beispiel eine Lawine ausgelöst werden, die andere Menschen einem Risiko aussetzt, das sie nicht gewählt haben und von dem sie nicht profitieren.

Außerdem können andere Entscheidungstragende die Möglichkeiten des Akteurs beschränken. So stellt sich zum Beispiel die Frage, welche Verantwortung eine Person trägt, die nicht alle Pisten so abgesperrt hat, dass niemand abseits der Piste Skifahren kann. Diese Person wäre in dem Fall ausschließlich in der Rolle eines Entscheidungstragenden. Ihre Aufgabe wäre es, unparteiisch zu sein und die Interessen der Risikoexponierten und Begünstigten auszubalancieren. Dies ist eine verantwortungsvolle Aufgabe, da die von der Entscheidung betroffenen Personen die Entscheidung nicht beeinflussen können.

Auch wenn Akteure nur die Rolle von Begünstigten verkörpern, muss über die Verteilung der Vor- und Nachteile nachgedacht werden. Für den begünstigten Akteur ist dies eine vorteilhafte Position, aus gesellschaftlicher Sicht erscheint das jedoch unfair für diejenigen, die das Risiko tragen. Wenn aber ein Akteur sowohl zu den Entscheidungstragenden als auch zu den Begünstigten gehört, ist das

die problematischste Rollenkombination aus ethischer Sicht. Dieser Akteur entscheidet über das Risiko und profitiert davon, dass es eingegangen wird, aber geht das Risiko nicht selbst ein. Dadurch entsteht der Anreiz, anderen mehr Risiken aufzuerlegen, um selber davon zu profitieren, ohne Nachteile dadurch zu erleiden.¹³⁵⁴

Im dritten Schritt der eRA können verschiedene Instrumente zur ethischen Analyse der jeweiligen Rollenkombinationen eingesetzt werden. Hansson schlägt vor, dass für Akteure, die sowohl dem Risiko ausgesetzt sind, als auch von ihm profitieren, eine individuelle Abwägung zwischen Risiko und Nutzen durchgeführt wird. Eine solche Abwägung kennt man zum Beispiel aus dem medizinischen Kontext, in welchem im Idealfall die Patientin selbst die Risiken und den Nutzen einer bestimmten Behandlung abwägt, um möglichst ein Gleichgewicht zu erreichen, das auf den eigenen Werten und Prioritäten basiert.¹³⁵⁵

Für den Fall, dass diese individuelle Analyse eine ungleiche Verteilung zwischen Risiko und Nutzen ergibt, schlägt Hansson die Durchführung einer Verteilungsanalyse vor. Für eine Verteilungsanalyse müssen quantitative Schätzungen oder Indikatoren sowohl für die Risiken als auch für den Nutzen erstellt werden. Mit diesen soll festgestellt werden, wie die Verteilung der neuen Risiken und Nutzen mit bereits bestehenden Ungleichheiten zusammenhängt. So zeigt sich, ob die Rollenverteilung fair ist. Denn es zeigt sich, ob die Risiken vorrangig von sowieso schon benachteiligten Gesellschaftsgruppen getragen werden und ob sowieso schon privilegierte Gesellschaftsgruppen davon profitieren. Laut Hansson ist die Durchführung dieser Verteilungsanalyse auch bei allen anderen Rollenkonstellationen sinnvoll, außer wenn ein Akteur ausschließlich die Rolle der Entscheidungstragenden einnimmt.¹³⁵⁶

Wenn nun das Risiko nicht durch den Nutzen aufgewogen wird oder wenn eine Gruppe von Risikoexponierten niemals von dem Risiko profitiert, sollte laut Hansson analysiert werden, ob ihre moralischen Rechte verletzt wurden. Denn, wie in Kapitel 5 ausführlich besprochen, ist es laut Hansson nur dann zulässig, ein Risiko von außen aufzuerlegen, wenn dies Teil eines fairen und für beide

1354 Hansson 2018:1824ff.

1355 Ebd.:1826

1356 Ebd.:1827

Seiten vorteilhaften Austauschs von Risikoexpositionen ist.¹³⁵⁷ Insbesondere, wenn es machtlose, risikoexponierte Personen gibt oder Entscheidungstragende, die davon profitieren, dass andere einem Risiko ausgesetzt sind, dann ist laut Hansson zusätzlich eine Analyse der zugrunde liegenden Machtstruktur erforderlich. Er schlägt vor, diese Frage zu stellen: Was für ein System der Entscheidungsfindung für diese Art von Risiko würden wir aufbauen, wenn wir von vorne anfangen würden? Auf dieser Grundlage können Vor- und Nachteile des aktuellen Systems diskutiert und Alternativen vorgeschlagen werden.¹³⁵⁸

Schritt drei der eRA kann außerdem die Identifikation von weiteren für die Risikoanalyse relevanten Akteuren beinhalten. So können zum Beispiel wissenschaftliche Expert:innen oder Journalist:innen als unabhängige Informationslieferant:innen fungieren, und Personen wie Rechtsanwält:innen und Lobbyist:innen können als Vertreter:innen von Risikoexponierten, Entscheidungstragenden oder Begünstigten auftreten. Wissenschaft spielt eine Rolle, wenn es darum geht, Risiken zu erkennen, zu bestätigen und ihr Ausmaß zu bestimmen. Journalist:innen spielen eine Rolle, da die Öffentlichkeit häufig vor allem über die Medien über potenzielle Risiken informiert wird. Problematisch wird es, wenn Wissenschaftler:innen oder Journalist:innen Verbindungen zu mächtigen, durch das Risiko begünstigten Akteuren haben. Deshalb bewertet die eRA, ob jegliche Informationslieferant:innen den von ihnen erwarteten ethischen Grundsätzen gerecht werden. Bei dem Einsatz von Vertreter:innen ist es die Aufgabe der eRA zu überprüfen, ob auch hier moralisch zulässig gehandelt wird.

Die ethische Risikoanalyse bietet dank ihrer drei Schritte eine konkrete praktischen Anleitung für die Ausrichtung technischer Entwicklungen an ethischen Maßstäben (i). Durch den Fokus auf die Fragen, wer durch das Risiko begünstigt wird, wird deutlich, dass der Einfluss der Technologie auf alle Betroffenen berücksichtigt wird und der Vorteil für alle von der Technologie betroffenen Personen das Ziel ist (ii). Die Verantwortung für die ethische Gestaltung wird unparteiischen Akteuren zugeschrieben, damit eine Outsid-

1357 Hansson 2018:1827

1358 Ebd.

er-Perspektive eingenommen werden und auf die unterschiedliche Verhandlungsmacht der verschiedenen Stakeholder aufmerksam gemacht werden kann (iii). Bei den unparteiischen Akteuren kann es sich zum Beispiel um akademische, gemeinnützige oder staatliche Organisationen handeln. Verantwortlich sollen also menschliche Akteure sein, aber es bleibt vage, welche bestimmten menschlichen Akteure die Verantwortung übernehmen sollen, aber es werden zumindest mehrere Vorschläge gemacht, die sich noch ausgestalten ließen. Somit erfüllt die ethische Risikoanalyse alle allgemeinen Kriterien für die ethische Gestaltung einer Technologie.

Konkret in Bezug auf die ethische Gestaltung eines Datenökosystems wäre es mit Hilfe der eRA zumindest möglich, Verletzung von Privatsphäre und Abwesenheit von Transparenz als Risiken zu identifizieren. Denn dank der eRA könnte man erkennen, dass Datenquellen das Risiko der Verletzung ihrer Privatsphäre und Datennutzenden das Risiko fehlender Transparenz auferlegt wird. Da Datenquellen und Datennutzende nicht über das ihnen auferlegte Risiko entscheiden können, handelt es sich bei diesem Akteur um machtlose, risikoexponierte Personen. Deshalb wäre hier laut Schritt drei der eRA eine Analyse der zugrunde liegenden Machtstruktur erforderlich.¹³⁵⁹ Sie schlägt aber keine konkreten Schritte vor, welche die Ableitung von administrativen Anforderungen zur Umsetzung von Privatsphäre für die Datenquellen (II) und zur Umsetzung von Transparenz für die Datennutzenden (III) ermöglichen würde, so dass auch ihre Erfüllung schwierig wäre.

Ich stimme Hansson zu, dass beim Vorkommen von „powerless risk-exposed persons“¹³⁶⁰ die Machtverhältnisse, bzw. wie ich sagen würde die Möglichkeiten zur Kontrolle, analysiert werden sollten. Ich denke aber, dass es nicht reicht, sich das ideale System auszumalen und Alternativen zum aktuellen System zu diskutieren. Das hieße im Kontext von Verwaltungsleistungen, dass man das gesamte deutsche Verwaltungssystem hinterfragen müsste – das würde nur unser System destabilisieren und nicht das Problem der fehlenden Risiko-Kontrolle lösen. Ich argumentiere, dass wir stattdessen Bedingungen definieren müssen, unter welchen es zulässig ist, anderen ein Risiko aufzuerlegen.

1359 Hansson 2018:1827

1360 Ebd.

Meiner Meinung nach müsste also in Schritt drei der eRA analysiert werden, unter welchen Umständen das Risiko zulässig ist. Denn nur so ist es möglich, die Auferlegung von Risiken zu regulieren (I). Laut Hansson ist es nur dann zulässig, ein Risiko von außen aufzuerlegen, wenn dies Teil eines fairen und für beide Seiten vorteilhaften Austauschs von Risikoexpositionen ist.¹³⁶¹ Ich habe jedoch in Kapitel 5 argumentiert, dass es für eine zulässige Risikoauferlegung nicht reicht, wenn ein Risiko durch einen eigenen Vorteil kompensiert wird. Meine Ableitung einer Regel für fairen und für beide Seiten vorteilhaften Austausch von Risikoexpositionen aus Kants dritter Formel des Kategorischen Imperativs kann an dieser Stelle klarere Antworten geben und die Regulierung von Risikoauferlegung ermöglichen. Deshalb entwickle ich im Folgenden auf Basis der eRA sowie der ethischen Überlegungen in dieser Arbeit eine eigene Umsetzungsmethode zur ethischen Gestaltung eines Datenökosystems.

1361 Hansson 2018:1827.

