

4.

IMPULSE

Technologieentwicklung und Gerechtigkeit im Zeitalter der Digitalisierung¹

Die Diversity-Folgenabschätzung als Instrument zur Auflösung des Collingridge-Dilemmas

Emily Breuer und Olivia Hankins

1. Einleitung

Debatten um die Abwägung der Chancen und Risiken freier technologischer Entwicklung, insbesondere in Bezug auf ethische Auswirkungen von KI auf unsere Gesellschaft und unser ökonomisches Handeln, sind hochaktuell. Schon in den vergangenen Jahren wurden in Europa, als auch den USA Richtlinien für den Umgang und die Entwicklung von KI-Systemen diskutiert und letztendlich sowohl von der Europäischen Union (vgl. EU-Verordnungen 2024) als auch unter dem ehemaligen US-Präsidenten Joe Biden ausgearbeitet (vgl. O'Brian/Parvini 2024). Im US-Wahlkampf 2024 betonte vor allem Kamala Harris die Relevanz der Regulierung von technologischer Innovation, indem sie auf schwerwiegende Auswirkungen von verzerrten KI-Systemen oder deren Missbrauch Bezug nahm.

When a woman is threatened by an abusive partner with explicit deepfake photographs, is that not existential for her? (ebd.)

¹ Workshop auf der #CDRK24 Konferenz unter dem Titel „Diversity-Folgenabschätzung: Ein Instrument, um Diversität, Inklusion und Chancengleichheit zu fördern“ – geleitet von Thomas Bieler (ING Bank), Maria Hinz (Barmer) und Christian Thorun (CDR-Initiative), betreut von Emily Breuer (Universität Bayreuth) und Olivia Hankins (Universität Bayreuth). Der nachfolgende Beitrag ist im Rahmen eines Blockseminars „Ethik der Digitalisierung“ an der Universität Bayreuth entstanden und wurde durch den Workshop auf der #CDRK24 Konferenz inspiriert.

Harris macht deutlich, dass KI-Systeme in ihrer Anwendung einen bedeutenden Einfluss auf die Gesellschaft haben und Fehler in ihren Algorithmen zu schwerwiegenden Folgen führen können, welche unsere demokratische Gesellschaft in Bezug auf ihre Grundpfeiler von Gleichheit und Gerechtigkeit bedrohen können. Das bedeutet, dass in frühen Entwicklungsphasen die Folgen eines solchen Algorithmus nicht adäquat erfasst worden sind. Demgegenüber stehen Haltungen, welche in ethischen Bedenken und frühzeitigen Regulierungen von KI eine Beschränkung des technologischen Fortschritts sehen, da wir die genauen Auswirkungen heute noch nicht exakt beziffern können. So hob Donald Trump nach seiner Wiederwahl die von Joe Biden beschlossenen Richtlinien zum Umgang mit KI, wie ethische Überlegungen zur Kennzeichnung von KI-Inhalten oder Sicherheitsabschätzungen der Systeme für die Gesellschaft, auf, da er diese als Hindernis für technologische Innovation betrachtete (vgl. Spiegel 2025). Dies kündigte er bereits im Wahlkampf mit den Worten: “We will repeal Joe Biden’s dangerous Executive Order that hinders AI innovation, and imposes Radical Leftwing ideas on the development of this technology” (O’Brian/Parvini 2024) an. Hinsichtlich dieser Entwicklung stellt sich auch für Unternehmen die Frage, inwiefern sie an ethischen Überlegungen und frühzeitigen Regulierungen in der Entwicklung von Technologie festhalten sollten, wenn äußere Regularien wegfallen.

In diesem Artikel beabsichtigen wir die Relevanz einer frühzeitigen Berücksichtigung ethischer Prinzipien und Folgenabschätzungen in der Technologieentwicklung zu verdeutlichen und der Frage entgegenzutreten, ob wir tatsächlich Chancen verbauen, wenn wir basierend auf noch nicht vollständig erkennbaren Auswirkungen von Technologien ethische Richtlinien für die Entwicklung setzen. Stattdessen könnten die Regulierungen sogar als Voraussetzung dafür gesehen werden, Chancen überhaupt nutzen zu können – indem sie ethische Risiken und Fehlentwicklungen minimieren. Diese Fragestellung weist auf ein Dilemma hin, welches in der Literatur als Collingridge-Dilemma bekannt ist (vgl. Collingridge 1980). Es besagt, dass die potenziellen Konsequenzen einer Implementierung der Technologie in einem frühen Entwicklungsstadium in ihrem vollen Umfang noch unbekannt sind und sich erst nach der Implementierung zeigen und daher schwer zu berücksichtigen sind. Zeigen sich die Folgen erst nach der Implementierung, so ist ein Gegensteuern aufgrund der bereits vielseitigen Verwurzelung der Technologie schwierig und kostspielig, woraus sich ein Dilemma ergibt. Obwohl die Unsicherheit über die Zukunft das Dilemma schwer aufzulösen scheint, hat ein Versuch höchste Relevanz, womit sich die Frage nach einer Möglichkeit der

Auflösung des Dilemmas ergibt, d.h. wie man ethische Regulationen schon im Entwicklungsprozess einbindet, um schwerwiegende Folgen zu adressieren, die man noch nicht genau kennt.

In diesem Artikel möchten wir das Instrument der Diversity-Folgenabschätzung, welches von der Corporate Digital Responsibility Initiative für Unternehmen entwickelt wurde, als Versuch betrachten, das Collingridge-Dilemma aufzulösen und es in Bezug auf sein philosophisches Fundament analysieren. Die Corporate Digital Responsibility (CDR) Initiative unterstützt Unternehmen, Verantwortung für eine ethisch fundierte Umsetzung der Digitalisierung zu übernehmen, wobei insbesondere der Mensch in den Vordergrund gestellt wird (vgl. CDR-Initiative 2024). Die Diversity-Folgenabschätzung, welche von ihr als Instrument für unternehmensinterne Digitalisierungsprojekte entworfen wurde, soll ethische Überlegungen zu Inklusion und Diversität schon in den Entwicklungsprozess implementieren, um negative Auswirkungen wie einer Diskriminierung von Mitarbeitenden über inadäquate Angepasstheit an Fähigkeiten im Umgang mit digitalen Systemen, vorzubeugen. Im Folgenden werden wir die Diversity-Folgenabschätzung bezüglich ihres Potenzials das Collingridge Dilemma zu adressieren betrachtet. Dabei wird sich trotz einiger Unzulänglichkeiten nicht nur zeigen, dass sie das Dilemma in relevanter Weise adressieren kann, sondern auch dass sie Unternehmen eine gute Grundlage bietet ihrer Verantwortung und eigenem Interesse in der Auflösung des Dilemmas nachzukommen. Ihre gesellschaftliche und ethische Relevanz zeigt sich dabei vor allem auch in ihrer Anwendbarkeit auf Szenarien, wo keine politischen Regularien für Unternehmen seitens der Politik getroffen werden.

2. Die Relevanz der Berücksichtigung ethischer Prinzipien in der Entwicklung von Technologie

2.1 Die Perspektive eines Laissez-faire-Ansatzes und die Regulierung der Entwicklung und des Einsatzes von Künstlicher Intelligenz

Im Zeichen der Debatte um die (ethische) Notwendigkeit von Regularien in der KI-Entwicklung bestehen unterschiedliche Positionen. Dabei werden Regularien einerseits als ein Hemmnis für die Entwicklung fortschrittlicher Technologie gesehen und ethischen Bedenken in der Entwicklung wenig Raum geboten, woraus eine Laissez-Faire-Haltung erwächst, bei welcher technologischer Fortschritt weitgehend ohne regulatorische Eingriffe gefördert wird und damit

keine ethischen Richtlinien für die Unternehmen vorgegeben werden, sodass potenzielle aber noch nicht genau absehbare gesellschaftliche Konsequenzen nicht zwangsläufig Berücksichtigung finden, bzw. erst nach der Implementierung untersucht werden.

Auf der anderen Seite stehen Positionen, welche eine hohe Relevanz in der Vorgabe einheitlicher ethischer Maßstäbe, wie Transparenzanforderungen oder Tests vor der Implementierung, sehen, um ethischen Konsequenzen für die Gesellschaft aus der Entwicklung von KI-Systemen vorzubeugen und Unternehmen Anreize zu setzen ethische Betrachtungen in ihre Entwicklungen einzubeziehen. Als Beispiel lässt sich hier der „AI-Act“ der Europäischen Union anführen, welcher Regulierungen für künstliche Intelligenz umfasst und (ethische) Risikoabwägungen trifft (vgl. Europäisches Parlament 2023). Man versucht Richtlinien für Folgen zu implementieren, deren Ausmaß noch nicht feststeht, um Schaden von der Gesellschaft abzuwenden.

Aus der Unternehmensperspektive erscheint ein Laissez-Faire-Ansatz zunächst attraktiver, da er zulässt, dass sich Technologie frei entwickeln kann, alle Möglichkeiten offenstehen und geringere Kosten in der Entwicklung anfallen, wenn nicht zusätzlich ethische Überlegungen und Richtlinien zu berücksichtigen sind und Ressourcen nicht durch etwaige zukünftige Folgen, welche heute noch nicht im Detail bestimmbar sind, gebunden sind. Es erscheint, als könnte er dazu beitragen, dass bahnbrechende Technologien schneller zur Marktreife gelangen und Wettbewerb sowie wirtschaftliches Wachstum stimuliert werden. Dies bedeutet aber auch bedeutende Risiken in unvorhergesehenen negativen Auswirkungen, da negative Auswirkungen für die Gesellschaft von einem Unternehmen unter dem Laissez-Faire-Ansatz nicht zwangsläufig internalisiert werden, da sie nicht mit übergeordneten Richtlinien an die ethischen Konsequenzen ihrer Entwicklung gebunden sind. Mit der Abschaffung einheitlicher gesetzlicher Richtlinien zur Entwicklung und Einsatz von KI durch Donald Trump stehen Unternehmen in den USA einer Laissez-Faire-Situation gegenüber (vgl. Spiegel 2025), woraus die Frage erwächst, ob dennoch eine Betrachtung ethischer Maßstäbe in der Entwicklung ihrer Technologie eingebunden bleiben sollten.

Im Folgenden werden wir vor diesem Hintergrund erläutern, weshalb insbesondere für die Unternehmen auch unabhängig von gesetzlichen Vorgaben eine Relevanz in der frühzeitigen Berücksichtigung ethischer Überlegungen und der gesellschaftlichen Folgen von Technologieentwicklungen besteht. Es wird sich zeigen, dass ethische Überlegungen in ihrem eigenen wirtschaftlichen Interesse liegen und eine Laissez-Faire-Situation sie nicht von ihrer gesellschaftlichen Verantwortung in der Technologieentwicklung entbindet.

2.2 Künstliche Intelligenz im Kontext gesellschaftlicher und ethischer Prinzipien

Unsere Gesellschaft basiert in einem hohen Maße auf ethischen Prinzipien, welche unsere politische Organisation, unser Wirtschaften und das Zusammenleben bestimmen. Es bestehen verschiedene Vorstellungen von einer gerechten und moralischen Gesellschaft, wobei wir in diesem Zusammenhang vor allem auf die Vertragstheorie John Rawls' verweisen wollen. Rawls beschreibt die Gesellschaft als kooperatives System, dessen Gerechtigkeitsprinzipien auf der Grundlage eines fairen Entscheidungsprozesses basieren (vgl. Rawls 1999). Er stellt eine hypothetische Ausgangssituation dar, in der alle Mitglieder der Gesellschaft über die grundlegenden Prinzipien sozialer Gerechtigkeit entscheiden, dabei jedoch nicht wissen, welche soziale Position sie später einnehmen werden. Da niemand weiß, ob er oder sie reich oder arm, privilegiert oder benachteiligt sein wird, würden rationale Individuen Prinzipien wählen, die für alle fair sind (vgl. ebd.: 15–19). Unterschiedliche körperliche und soziale Merkmale sind dabei moralisch willkürlich, weshalb kein darauf basierender Anspruch auf mehr Rechte besteht (vgl. Wenar 2021). Dies impliziert ein Prinzip der Gleichheit und Nichtdiskriminierung. Ethische Prinzipien wie Gleichheit und Gerechtigkeit sind somit nicht nur abstrakte Werte, sondern in der gesellschaftlichen Übereinkunft unseres gemeinsamen Wirtschaftens und Zusammenlebens verankert. Auch wenn diese Überlegungen in einem vorgesellschaftlichen Zustand hypothetisch sind und keine tatsächliche vertragliche Vereinbarung besteht, zeigen sie auf, dass Gerechtigkeitsprinzipien rational fundiert sind und aus ihr für die Mitglieder eine Verantwortung ihrer Bewahrung und Verwirklichung erwächst. Dies ist keine rein politische und institutionelle Verantwortung, welche allein die Wahrung von Rechten und einer staatlichen Struktur betrifft, sondern betrifft jede unserer Handlungen, welche eine Wirkung auf gesellschaftliche Zusammenhänge hat, unter anderem auch die Entwicklung von Technologie.

Es besteht ein Bewusstsein darüber, dass neue technologische Entwicklungen unsere Gesellschaft prägen und verändern, woraus auch ein verantwortungsvoller Umgang hinsichtlich ihrer Berührung ethischer Prinzipien erwächst. Dies kann derzeit vor allem vor dem Hintergrund der Digitalisierung erkannt werden. So findet beispielsweise ein großer Teil unserer Kommunikation mittlerweile in digitalen Räumen statt, die häufig als frei zugängliche und gleichberechtigte Plattformen des Meinungsaustauschs illustriert werden. Dennoch sind Plattformen, wie Facebook oder X (ehemals Twitter) ebenso über technische Strukturen wie Algorithmen geprägt, wodurch Aufmerksamkeit und Reichweite gelenkt werden und nicht allein in der Hand des Nutzers liegen.

Die neue Technologie formt damit die Struktur unserer gesellschaftlichen Kommunikation, woraus aber auch eine Verantwortung in der Entwicklung erwächst, ethische Prinzipien, auf welchen unsere Gesellschaft baut, auch im digitalen Raum zu erhalten und diese in die Technologie zu implementieren, damit beispielsweise Diskriminierungen im digitalen Raum ebenso sozial sanktioniert werden und sich nicht strukturell verfestigen (vgl. Seeliger/Sevignani 2022: 3–16).

Ein weiteres Beispiel besteht in KI-gestützten Bewerbermanagementsystemen, welche Bewerbungsverfahren für eine offene Stelle optimieren sollen. Auch hier gilt es in ihrer Entwicklung ethische Prinzipien, wie der Inklusion und Nicht-Diskriminierung, hinsichtlich des Geschlechts, der Herkunft oder einer Behinderung zu berücksichtigen, da sie auf unsere gesellschaftlichen Strukturen, wie die Zusammensetzung des Personals und Bewerbungschancen wirken. KI-Systeme basieren dabei auf menschlichen Daten, wodurch die Gefahr besteht, dass bestehende strukturelle soziale oder politische Voreingenommenheit, die in der Gesellschaft bestehen, in maschinell lernende Algorithmen als Muster wahrgenommenen und verfestigt werden, sodass sich Entscheidungsprozesse darauf berufen (vgl. Ruttkamp-Bloem 2023: 20–23). Wenn der Fall besteht, dass diese Systeme mit voreingenommenen Daten trainiert wurden, sodass z. B. in der Vergangenheit überwiegend männliche Bewerber bevorzugt wurden, woraufhin die entwickelte KI Frauen oder andere Gruppen systematisch benachteiligt, so betrifft dies die Strukturen unserer Gesellschaft und in einem größeren Kontext die strukturelle Verankerung der ethischen Prinzipien. Das heißt, dass die Technologie nicht nur Entscheidungskriterien für die Vergabe von Jobs berührt, sondern über mögliche Diskriminierungen auch fundamentale gesellschaftliche Prinzipien der Gleichberechtigung oder Inklusion verletzen kann. Dies wird bei einer langfristigen Implementierung nicht nur einzelne Jobvergaben beeinflussen, sondern auch die Umsetzung ethischer Prinzipien prägen und damit die strukturelle Art und Weise, wie Entscheidungen in unserer Gesellschaft getroffen werden. Die Relevanz der Implementierung ethischer Überlegungen und Folgenabschätzungen bei der Entwicklung von Technologie besteht somit darin, dass sie bedeutende Einflüsse auf unsere gesellschaftlichen Strukturen und die Verwirklichung unserer ethischen Prinzipien haben, welche auf eine gerechte Gesellschaft abzielen.

Die Verantwortung, die mit der Wirkung von Technologieentwicklungen auf die gesellschaftliche Struktur einhergeht, liegt dabei jedoch nicht allein in der Vorgabe von Regularien seitens politischer Institutionen, sondern auch bei jedem Unternehmen (vgl. Strasser/Niedermayer 2021),

da Unternehmen in einem hohen Maße mit der Gesellschaft verflochten sind und in einem Wechselspiel stehen. Ihre Produkte und Dienstleistungen, d.h. auch ihre technologischen Innovationen stellen Lösungen und Angebote für die Bewältigung gesellschaftlicher Prozesse und Bedürfnissen dar wie z. B. die Informationsbeschaffung mit Hilfe intelligenter Suchmaschinen oder KI-basierter Auswahlsysteme für Jobs oder Studienplätze. Diese betreffen mitunter bedeutende und lebensverändernde Angelegenheiten, da Entscheidungsprozesse der Systeme uns direkt betreffen oder wir aufbauend auf unserem Informationsstand politische, ökonomische, aber auch individuelle Entscheidungen treffen, wodurch die Ausgestaltung der Technologie durch die Entwickler relevanten Einfluss auf das Leben ihrer Nutzer und darüber hinaus langfristig auf die gesellschaftliche Struktur hat.

2.3 Unternehmensethik im Kontext von KI

Eine Implementierung ethischer Prinzipien in der Entwicklung von KI ist jedoch nicht nur aufgrund der gesellschaftlichen Verantwortungsbeziehung eines Unternehmens relevant, sondern ist auch in ihrem eigenen Interesse begründet. Auch wenn eine Laissez-Faire-Haltung hinsichtlich Innovationen vielversprechender erscheinen mag, wird ein Unternehmen langfristig von einer frühzeitigen Implementierung ethischer Überlegungen profitieren. Die Einbindung ethischer Überlegungen in die Entwicklung von Technologie ist auch eine Strategie der Risikominimierung von gesellschaftlichen Folgen und damit auch für die Nutzer. Dazu gehören nicht nur rechtliche Aspekte, wie etwa die Haftung für diskriminierende oder fehlerhafte KI-Entscheidungen, sondern auch wirtschaftliche Risiken. Investitionen in Technologien, die sich später als ethisch problematisch und gesellschaftlich inakzeptabel erweisen, können zu finanziellen Verlusten führen und das Vertrauen der Öffentlichkeit sowie die Reputation eines Unternehmens erheblich schädigen. Auch in Bezug auf die Marktakzeptanz spielt die frühzeitige Einbindung ethischer Überlegungen eine entscheidende Rolle, da Verbraucher und Unternehmen zunehmend Wert auf verantwortungsbewusste Technologien legen, die Transparenz, Fairness und gesellschaftliche Werte berücksichtigen, woraus sich eine sinkende Wettbewerbsfähigkeit ergeben kann. So zeigen Anissa Dakhli's Untersuchungen zur Beziehung zwischen der sozialen Verantwortung und der finanziellen Leistungsfähigkeit von Unternehmen, dass ein positiver Einfluss auf die finanzielle Leistung eines Unternehmens besteht, wenn ethisches Verhalten und Transparenz in seinen Aktivitäten berücksichtigt werden (vgl. Dakhli 2021). Die Berücksichtigung ethischer Prinzipien ist

dabei jedoch nicht nur eine Reaktion auf gesellschaftliche Entwicklungen, sondern auch in der Anreizstruktur eines Unternehmens verankert. Eine diverse und inklusive Unternehmenskultur trägt dazu bei, vielfältige Perspektiven und Innovationsideen zu fördern, was die Wettbewerbsfähigkeit langfristig stärken kann. Auch wenn Gerechtigkeitsprinzipien in erster Hinsicht auch als Einschnitt in die Wirtschaftlichkeit eines Unternehmens begriffen werden können, z. B. über Quoten zur Einstellung bestimmter Personengruppen. So führt, wie Scott E. Page untersucht hat, ein diverses Umfeld langfristig zu einer besseren Anpassungsfähigkeit, kreativeren Lösungsansätzen und einer erhöhten Innovationskraft (vgl. Page 2007).

Durch die ethische Fundierung unserer Gesellschaft wird die Entwicklung neuer Technologie, welche im Kontext der Digitalisierung gesellschaftliche Strukturen neugestaltet, an die Berücksichtigung ethischer Folgen gebunden sein und die Entwicklung von z. B. ethischen KI-Systemen einen Anspruch darstellen. Damit wird auch der wirtschaftliche Erfolg der Unternehmen und ihre Wirtschaftsweise im Kontext der Digitalisierung von ihrer Implementierung ethischer Folgenabschätzungen abhängen (vgl. Lauer 2021: 21–23). Dies stellt jedoch eine bedeutende Herausforderung dar, da die genauen Auswirkungen einer Technologie sich erst in ihrer zukünftigen Implementierung zeigen und im Entwicklungsprozess, zwar teilweise kalkuliert, aber schwer reguliert werden können. Dies ist vor allem im Fall der KI-Systeme ersichtlich, da ihre Entscheidungsprozesse zwar auf ausgewählten Trainingsdaten basieren, sie jedoch kontinuierlich über die Interaktion mit weiteren Daten hinzulernen, was die Abschätzung ihrer Entwicklung und Ausgestaltung und damit auch ihrer gesellschaftlichen Auswirkungen erschwert. Wir kennen die genauen Folgen damit erst, wenn sie entstehen, was es schwer macht, Regulierungen präzise genug zu gestalten, sodass sie die Folgen abfangen, bzw. ethische Bedenken adäquat auf die Zukunft auszurichten. Damit stellt sich die Frage, wie wir heute alle ethischen Folgen der Technologie abschätzen können, die erst morgen sichtbar sein werden. Kann dieses Dilemma über die Implementierung ethischer Überlegungen in der Technologie gelöst werden?

Die Erläuterungen dieses Kapitels haben gezeigt, dass auch wenn seitens der Politik eine Laissez-Faire-Haltung bezüglich der ethischen Regularien besteht, für Unternehmen eine Relevanz in der frühzeitigen Berücksichtigung ethischer Auswirkungen ihrer Technologieentwicklung besteht, welche nicht nur in ihrer gesellschaftlichen Verantwortung, sondern auch in einem ökonomischen Eigeninteresse begründet liegt. Die unternehmerische Praxis eines regulatorischen Ansatzes über die Implementierung von ethischen Prinzipien in die Entwicklung von Technologie

beinhaltet dennoch Unsicherheiten hinsichtlich ihrer Umsetzung, da es in einem frühen Stadium der Entwicklung schwer ist die Folgen abzuschätzen, während sie im Nachhinein eine zu starke Verankerung und auch schon Auswirkungen aufweisen. Dieses so genannte Collingridge-Dilemma wird im Folgenden erläutert.“

3. Das Collingridge-Dilemma

Das Collingridge-Dilemma, welches auf David Collingridge zurück geht, ist vor allem im Bereich der Folgen und Risikoabschätzung neuer Entwicklungen bedeutsam, insbesondere in Bezug auf deren Wirkung hinsichtlich der Ausgestaltung und Beförderung einer gerechten und inklusiven Gesellschaft (vgl. Collingridge 1980). Dieses Dilemma besteht darin, dass früh im Entwicklungsprozess der Technologie die potenziellen Konsequenzen dieser Technologie noch nicht bekannt sind, sondern die Effekte erst zu einem späteren Zeitpunkt nach ihrer Implementierung sichtbar werden. Zu diesem Zeitpunkt ist es allerdings schwierig, die Technologie noch zu regulieren, da Prozesse schon fest verankert sind. Kurz gesagt ist dieses Dilemma ein Problem der Zeit und Technologieentwicklung (vgl. Demos Helsinki 2022). Im Collingridge-Dilemma gibt es zwei Bedingungen, welche bestehen müssen, um die unerwünschten Konsequenzen einer Technologie vermeiden zu können. Zum einen muss Wissen darüber bestehen, dass eine Technologie schwerwiegende Konsequenzen hat oder in Zukunft haben wird. Zweitens muss es möglich sein, die Technologie im späteren zu verändern, um diese Konsequenzen zu beheben. In der Praxis bleiben diese Bedingungen jedoch häufig unerfüllt, weshalb wir einem Dilemma gegenüberstehen.

Ein konkretes Beispiel, das dieses Dilemma eindeutig illustriert, besteht in der Entwicklung des Autos. Im Jahr 1908 waren Veränderungen im Design und der Benutzung von Autos einfach umzusetzen, wobei Prognosen zu späteren sozialen Konsequenzen noch schwer möglich waren. Die heutzutage auftretenden Folgen wie eine erhebliche Luftverschmutzung sowie soziale Veränderungen wie die Prägung unseres Mobilitätsverhalten, der Stadtarchitektur und einer Beförderung sozialer Ungleichheiten waren im Detail nicht vorherzusagen. Heute sind diese Folgen fast jeden Tag sichtbar und effektive Änderungen schwer durchzuführen, da sie sich fest in unserem gesellschaftlichen Zusammenleben verankert haben. Dieses Dilemma besteht auch bezüglich ei-

ner Folgenabschätzung zu ethischen Konsequenzen der Digitalisierung. Wir kennen beispielsweise heute die Auswirkungen einer zunehmenden Etablierung von KI-Assistenten auf unsere Gesellschaft nicht im Detail, da diese erst sichtbar werden, nachdem die Technologie schon in der Gesellschaft und Wirtschaft verankert ist. Trotz frühen Bemühungen zur Regulierung von KI-Systemen bleibt eine effektive Regulierung kompliziert, weil man in Anbetracht der Unsicherheit der zukünftigen gesellschaftlichen Struktur und der Etablierung der Systeme nicht weiß, wo der Fokus der Regulierung liegen sollte. Dieser kann lediglich über heutige Normen und Wertvorstellungen abgeschätzt werden. Das Dilemma von Collingridge kann damit auch folgendermaßen formuliert werden:

Either one attempts to control technology by early prediction of consequences, or one attempts to control technology by reacting to consequences as they unfold. If the former, then the attempt fails because of predictive unreliability. If the latter, then the attempt fails because of developed rigidity. Therefore, either the attempt to control technology fails because of predictive unreliability, or the attempt to control technology fails because of developed rigidity (Croy 1996: 2).

Das Collingridge-Dilemma betrifft damit den Zeitverlauf und den Entwicklungsprozess. In der Literatur bestehen unterschiedliche Lösungsansätze des Dilemmas. Unter anderen von Daniel Tigard, welcher diskutiert, inwiefern eine Brücke zwischen den beiden Zeitpunkten in der Technologieentwicklung und ihren Wissensständen geschlagen werden kann, sodass Folgen in frühen Stadien regulierbar sind und das Dilemma gelöst werden kann (vgl. Tigard 2022: 12–19). Er verweist unter der Bezeichnung “embedded ethics” auf flexible Ansätze, welche eine Vielzahl an Perspektiven in den Entwicklungsprozess einbeziehen, um mögliche Konsequenzen über die darin enthaltenen unterschiedlichen Erfahrungen und Bedürfnisse zu erkennen (vgl. ebd.: 7–20). Dabei soll insbesondere ein Bewusstsein für ethische und soziale Wirkungen von Technologie eingebettet werden und Sensibilität gegenüber Fehlern und Verzerrungen in Daten und Technologie auf Grundlage eines Abgleichs mit ethischen Prinzipien entstehen (vgl. ebd.: 17–27). Mit der Berücksichtigung ethischer Überlegungen und ihrer Einbettung in den frühen Entwicklungsprozess tut sich dabei jedoch eine weitere Dimension des Dilemmas auf und zeigt, dass diese Brücke zwischen beiden Zeitpunkten des Dilemmas auch mit einer solchen Einbettung schwer überwindbar scheint. Dies ist der Fall, weil technologische Entwicklungen

gesellschaftliche Umstände beeinflussen und dabei auch Normen und Werte formen. Wenn wir jedoch unsere heutigen Wertmaßstäbe anlegen, berücksichtigen wir nicht jene, die in der Zukunft gelten werden (vgl. Kudina/Verbeek 2019).

Kudina und Verbeek, adressieren dies in Abgrenzung zur wirtschaftlichen Variante des Dilemmas als ethische Version, welche darauf basiert, dass die Technologie oftmals auch unsere herangezogenen Werte verändert (vgl. ebd.). Dabei besteht das Problem, dass ethische Regularien entweder zu früh, d.h. bevor klar ist, wie die Technologie Bewertungsmaßstäbe beeinflussen wird, eine Bewertung der Technologie vornehmen oder zu spät, wo die ethischen Auswirkungen zwar bekannt sind, die Technologie aber bereits etabliert und kaum noch veränderbar ist (vgl. ebd.: 2–5). Laut den Autoren gibt es zwei Wege dieses Dilemma zu adressieren. Zum einen über einen technno-moralischen Weg, der versucht, die ethischen Aspekte einer Technologie vorherzusagen und zu antizipieren (vgl. ebd.: 292–295). Dabei ist dieser Weg kaum zuverlässig, weil er keine Dynamiken zwischen Technologie und Moralität adressiert, sondern nur ihre Implikationen. Zum anderen besteht der sozio-technische Weg, welcher Voraussagen über die Zukunft als spekulativ zurückweist und stattdessen Technologieentwicklungen als soziale Experimente auffasst, welche ethische Überlegungen in ihrer Durchführung erfordern. Er versucht damit, den Prozess technologischer Entwicklung zu regulieren (vgl. ebd.: 3). Die Autoren selbst plädieren, basierend auf ihrer These, dass sich normative Werte zusammen mit neuer Technologie verändert haben, für ein Vorgehen, das sich auf die Dynamiken der Interaktion zwischen Mensch und Technologie fokussiert und die ethischen Wechselwirkungen analysiert (vgl. ebd.: 1–3).

[W]hen technologies are used, they typically do not play a role as technological “objects” in interaction with human “subjects”; rather, they are “mediators” of the relation between users and their environment (ebd.: 7).

Wenn man die Entwicklung Technologies somit als ein dynamisches System versteht, gibt es mehr Möglichkeiten die Auswirkungen der Technologie auf die Gesellschaft zu verstehen. Nicht nur die Gesellschaft, sondern auch die Technologieentwicklung, sind kein stagnierendes Objekt, sondern verändern sie sich stetig und beeinflussen sich gleichzeitig.

Die Relevanz der Auflösung des Collingridge-Dilemmas ist eng mit der Relevanz des Einbezugs ethischer Überlegungen in die Entwicklung von Technologie verknüpft. Unternehmen haben ein Interesse daran, dass ihre Technologie zuverlässig funktioniert und in Zukunft zu keinen

negativen Auswirkungen führt, welche die Effizienz und Anwendbarkeit der Technologie gefährden, als auch daran, dass sich keine negativen Effekte auf die Gesellschaft ergeben, sodass sie ihrer Verantwortung nachkommen. Das Collingridge-Dilemma fasst dabei die Problematik zusammen, vor welcher Entwickler in der Entwicklung stehen. In einem frühen Stadium, vor der Implementierung sind Folgen noch nicht sichtbar und sind schwer zu regulieren, da sie sich erst zeigen, wenn die Technologie implementiert und eine Korrektur schwierig oder sehr kostspielig wird. Es liegt somit im Interesse eines Unternehmens, dieses Dilemma zu lösen, um möglichst geringe Risiken einzugehen und schwerwiegende Folgen zu vermeiden.

Im Folgenden wollen wir einen Ansatz vorstellen, welcher genau diesen Anspruch verfolgt und sich zur Aufgabe macht ethische Überlegungen zu Folgen technologischer Entwicklungen auf die Diversität in den Entwicklungsprozess zu implementieren, um diese zu regulieren und im späteren nicht auf schwerwiegende Konsequenzen für die Gesellschaft als auch für das Unternehmen zu stoßen. Wir werden diesen, als einen Versuch der Lösung des Collingridge-Dilemmas diskutieren und aufzeigen, inwiefern er das Dilemma auflösen kann und wo Verbesserungspotenzial oder auch Schwierigkeiten bestehen.

4. Die Diversity-Folgenabschätzung

4.1 Die Diversity-Folgenabschätzung der Corporate Digital Responsibility Initiative

Die Corporate Digital Responsibility Initiative ist eine Initiative des Bundesministeriums für Umwelt, Naturschutz, nukleare Sicherheit und Verbraucherschutz und bietet Unterstützung sowie ein Austauschforum für eine verantwortungsvolle Umsetzung der digitalen Transformation in Unternehmen. Dabei stehen neben der digitalen Veränderung insbesondere die sozialen und ethischen Auswirkungen dieses Wandels im Vordergrund und sollen Berücksichtigung finden. Im Jahr 2024 veröffentlichte die CDR-Initiative ein Instrument zur Diversity-Folgenabschätzung, für die Förderung von Diversität, Inklusion und Chancengleichheit im Prozess der Digitalisierung, welche Unternehmen als Leitfaden dienen soll (vgl. CDR-Initiative 2024). Dabei besteht die Kernidee darin, dass schon im Entwicklungsprozess digitaler Produkte und unternehmensinternen Dienstleistungen eine Vielzahl an Perspektiven und ethischen Überlegungen zur Betroffenheit implementiert werden, um Risiken nach der Implementierung

schon in der Entwicklung zu erkennen und besser abschätzen zu können (vgl. ebd.: 8–11). Darin lässt sich ein Kalkül des Versuchs einer Auflösung des Collingridge-Dilemmas, erkennen, da versucht wird Folgen, die eigentlich erst nach der Implementierung sichtbar werden und dann schwer zu verbessern sind über eine frühzeitige Berücksichtigung ethischer Standards als auch den Einbezug vieler Perspektiven, Betroffenheiten und Erfahrungen abzuschätzen. Unsere These lautet somit, dass die Diversity-Folgenabschätzung versucht, das Collingridge-Dilemma aufzulösen. Anlässlich einer genaueren Betrachtung der Diversity-Folgenabschätzung soll in diesem Teil zunächst das von der CDR-Initiative herausgegebene Papier zur Diversity-Folgenabschätzung dargestellt werden.

Laut der CDR-Initiative besteht eine gesetzliche, gesellschaftliche und wirtschaftliche Verantwortung, sich mit ethischen Konsequenzen in der Technologieentwicklung zu beschäftigen (vgl. ebd.: 9). Die Diversity-Folgenabschätzung ist dabei als Instrument geeignet, da sie einen einfachen Prozess zur Sensibilisierung für die Thematik der Inklusion, eine Identifikation von Verbesserungspotenzialen sowie einen bedeutsamen Beitrag in der Einbindung ethischer Überlegungen zum wirtschaftlichen Erfolg eines Unternehmens bietet (vgl. ebd.). Adressiert an potentielle unternehmerische Nutzer formulieren dies folgendermaßen:

Die Diversity-Folgenabschätzung ermöglicht es Ihnen, im Rahmen einer Vorabkontrolle zu überprüfen, welche Auswirkungen die Einführung von neuen digitalen Lösungen auf Ihre Mitarbeitenden, Kund*innen und andere Stakeholder hat (CDR-Initiative 2024: 10).

Die Diversity-Folgenabschätzung wird in drei Phasen gegliedert. Die erste Phase, die Vorbereitungsphase, ist dabei in vier Schritte unterteilt, wobei insbesondere die Festlegung des Anwendungsfalls und der zu untersuchenden Diversitäts-Dimensionen sowie die Identifikation der relevanten Stakeholder zu nennen sind. Um einen spezifischen Anwendungsfall wie z. B. die Entwicklung einer Mitarbeiterplattform, und die damit verbundenen Auswirkungen in Bezug auf Diversität und Inklusion klar abzugrenzen, gibt die Initiative den Unternehmen einen Fragekatalog an die Hand, welcher spezifische Fragen nach der Zielsetzung oder erhofften Verbesserungen adressiert (vgl. ebd.: 15–21). In einem weiteren Schritt werden dann die Diversitätsdimensionen, welche im konkreten Fall untersucht werden sollen, festgelegt, d.h. es werden Dimensionen, wie Alter, Geschlecht oder die Vereinbarkeit von Beruf und Familie

festgelegt, bezüglich welcher man eine Folgenabschätzung durchführen möchte, definiert. Wesentlich für die erfolgreiche Durchführung der Folgenabschätzung ist dabei, einen konkreten Anwendungsfall zu behandeln und zunächst nur einige und nicht zu hohe Ziele zu stecken, um diese effektiv zu bearbeiten, auch wenn im Prinzip noch viel mehr Dimensionen des Falls betrachtet werden könnten (vgl. ebd.: 15). Im Anschluss daran gilt es, die relevanten Stakeholder in Bezug auf ihre Betroffenheit im konkreten Anwendungsfall zu identifizieren. Mit der Bezeichnung der relevanten Stakeholder ist daran zu erinnern, dass die Diversity-Folgenabschätzung nur in einem interdisziplinären Team erfolgreich ist, wobei relevante Stakeholder zum Beispiel Mitarbeiter des Bereichs Human Ressource (HR), Diversitätsbeauftragte, Menschen aus den Dimensionen-Zielgruppen oder der Betriebsrat sein können. Damit alle Stakeholder an einem Workshop beteiligt werden können, gilt es seitens des Unternehmens eine geeignete Agenda und ein Format festzulegen.

Die zweite Phase beinhaltet zwei Schritte, die sich mit dem eigentlichen Workshop zur Folgenabschätzung bei der Entwicklung von digitalen Tools beschäftigen (vgl. ebd.: 22–24). Laut dem Papier sollte der Workshop die Relevanz der Thematik praxisnah abbilden und den konkreten Anwendungsfall in der Diskussion mit allen Stakeholdern hinsichtlich der Folgen für die Diversität analysieren (vgl. ebd.). Weiterhin sollten die zentralen Ergebnisse nach dem Workshop gesammelt und zusammengefasst werden, wobei nicht nur die (kritischen) zentralen Ergebnisse, sondern auch die Zielsetzung des Workshops und priorisierte Verbesserungsmaßnahmen berücksichtigt werden sollten.

In der dritten Phase, der Umsetzungs- und Reflexionsphase, sollen die Erfahrungen des Workshops reflektiert und die entwickelten Verbesserungsmaßnahmen implementiert werden (vgl. ebd.: 25–26). Weil jedes Unternehmen und ihr jeweiliger Anwendungsfall und die Verbesserungsmaßnahmen unterschiedlich sind, gibt es keine konkreten Empfehlungen seitens der Initiative, wie die Verbesserungsmaßnahmen implementiert werden sollten. Nach einiger Zeit sollten die Unternehmen jedoch erneut nachfragen, inwieweit die Verbesserungsmaßnahmen geholfen haben und die Ergebnisse und Erfahrungen mit der Diversity-Folgenabschätzung kritisch reflektieren, wobei die Stakeholder auch in Bezug auf ein Feedback eingebunden werden sollten (vgl. ebd.: 25 f.). Die Methode einer Diversity-Folgenabschätzung kann sich so immer weiterentwickeln, wozu zudem eine kurze Checkliste nach allen drei Phasen besteht. Die CDR-Initiative beschreibt auch zwei Beispiele Anwendungsfälle von der

ING und der Barmer. Um besser zu verstehen, wie diese Diversity-Folgenabschätzung aufgebaut ist, betrachten wir das Fallbeispiel der ING (vgl. ebd.: 18). Das Unternehmen nutzt ein internes digitales Tool für die Abrufung eines Gesundheitsbudgets für Mitarbeiter, welches Informationen und Möglichkeiten zu angebotenen Gesundheitsmaßnahmen bereitstellt. Im ersten Teil der Diversity-Folgenabschätzung konnten die relevanten Stakeholder identifiziert werden, die in der Nutzung des Tools wegen ihres Alters, Geschlechts oder anderen Diversitätsdimensionen benachteiligt werden könnten. Zum Beispiel waren einige digitale Icons in ihrer Bedeutung nicht für jeden selbstverständlich oder einige Schriftarten für einige schwer zu lesen (vgl. ebd.). Damit wurden Probleme identifiziert, die ohne einen konkreten Einbezug vielfältiger Perspektiven und Bedürfnisse unerkannt geblieben wären, da sie für Unbetroffene nicht ersichtlich waren. Nach dem Workshop standen damit konkrete Verbesserungen des digitalen Tools, hinsichtlich seiner gleichberechtigten Nutzbarkeit bereit. Die ING berichtet diesbezüglich:

Die Diversity-Folgenabschätzung hat sich dabei als Methode bewährt und uns – entgegen unserer ursprünglichen Einschätzung – Verbesserungsbedarf bei der Gestaltung des Tools und der enthaltenen Texte aufgezeigt (ebd.).

Dieses Beispiel zeigt, wie wichtig Diversity-Folgenabschätzung in der Technologieentwicklung ist, da ohne eine spezifische Betrachtung von Perspektiven und Betroffenheiten in Bezug auf die Diversität und Inklusion mögliche Folgen und Verbesserungspotenziale unentdeckt bleiben. Wenn die Technologie sich ohne Diversitätsmaßnahmen entwickelt, ist es für einige Menschen in der Gesellschaft nicht nutzbar und sie werden exkludiert, was in gravierenden Folgen für unsere Gesellschaft und die Unternehmenskultur münden kann. Vor dem Hintergrund aktueller Technologieneuerungen ist beispielsweise die große Frage, ob KI sich hinsichtlich Diversität und Inklusion adäquat entwickelt und wenn nicht, wie man diese adressieren könnte. Die Diversity-Folgenabschätzung bietet hierbei einen bedeutsamen Ansatz, indem schon im Entwicklungsprozess über eine Implementierung vielseitiger Perspektiven und der Berücksichtigung von Diversitätsdimensionen die Folgen der Entwicklung berücksichtigt werden. Die Diversity-Folgenabschätzung gliedert sich in drei Phasen, welche in der folgenden Abbildung dargestellt sind (vgl. Abbildung 1).



ABBILDUNG 1: PHASEN DER DIVERSITY-FOLGENABSCHÄTZUNG
 (QUELLE: CDR-INITIATIVE 2024: 12)

4.2 Die Diversity-Folgenabschätzung als Instrument zur Auflösung des Collingridge-Dilemmas

Wie schon erwähnt, ist das Collingridge-Dilemma ein Problem der Zeit und Technologieentwicklung, wobei es unterschiedliche Ansätze gibt, das Dilemma zu adressieren und aufzulösen. Wir möchten in diesem Kapitel zeigen, inwiefern der Diversity-Folgenabschätzung dies mit Bezug auf die Ansätze von Tigard sowie von Kudina und Verbeek gelingt.

Die Diversity-Folgenabschätzung versucht über den frühzeitigen Einbezug relevanter Stakeholder vielfältige Perspektiven in die Entwicklung einzubeziehen und damit ethische Konsequenzen früh abschätzbar zu machen. Sie werden bereits im Entwicklungsprozess über die verschiedenen Blickwinkel und ethischen Ansprüche implementiert, sodass die Technologie über den konkreten Einbezug von Diversitätsdimensionen zukünftige Folgen in der Entwicklung abzuschätzen versucht. Damit folgt die Diversity-Folgenabschätzung dem Ansatz Tigards, indem versucht wird, eine Brücke zwischen den beiden Zeitpunkten des Dilemmas zu schlagen und ethische Überlegungen in die Entwicklung einzubetten (vgl. Tigard 2022). Mit diesem Ansatz stellt sich allerdings erneut die Frage nach der Auflösung der ethischen Version des Dilemmas, d.h. jenes Dilemma, welches berücksichtigt, dass auch ethische Normen, welche implementiert werden, nicht jene sein werden, welche im späteren unsere Wertvorstellungen abbilden (vgl. Kudina/Verbeek 2019). Die Diversity-Folgenabschätzung scheint dies schwerlich vollständig auflösen zu können. Mit Blick auf den interaktiven Ansatz von Kudina und Verbeek, welcher die Wechselwirkungen von Mensch und Technologie in den Blick nimmt, lässt sich ein genauerer Blick wagen (vgl. ebd.).

Die Diversity-Folgenabschätzung der CDR-Initiative ist flexibel aufgebaut, was sich anhand vieler Feedbackprozesse und individuellen Ansätzen für das jeweilige Unternehmen und Produkt zeigt und womit es keinen einzigartigen Weg für die Förderung von Diversität gibt. Über die stetigen Reflexionsphasen und erneute Test-Phasen und Rücksprachen mit Stakeholdern kann die Diversity-Folgenabschätzung eine gewisse Dynamik adressieren und Veränderungen in Wertvorstellungen oder neue Diversitätsdimensionen einbeziehen. So können potenzielle Risiken identifiziert und notwendige Anpassungen vorgenommen werden, bevor die Technologie umfassend in die Unternehmensstrukturen integriert ist. Dennoch wird die Diversity-Folgenabschätzung in diesem Aspekt keine vollständige Auflösung des ethischen Dilemmas erreichen, da sie die Dynamik zwischen Technologie und Mensch und die daraus folgenden Veränderungen in Wert- und Inklusionsvorstellungen noch zu wenig antizipiert. Doch auch in Bezug auf die ursprüngliche Version des Collingridge-Dilemmas, d.h. dass die ethischen Konsequenzen der Technologie heute noch nicht regulierbar sind, da sie sich erst morgen zeigen, dann jedoch schwer reguliert werden können, bestehen Unzulänglichkeiten bezüglich dessen, ob die Diversity-Folgenabschätzung das Dilemma auflösen kann. Wir wollen diese im Folgenden darstellen.

5. Herausforderungen bei der Auflösung des Collingridge-Dilemmas durch die Diversity-Folgenabschätzung

5.1 Machtstrukturen und Hierarchien

Wir möchten in diesem Kapitel die Diversity-Folgenabschätzung, nicht ausschließlich als unternehmensinternes Vorhaben verstehen, sondern auch als ein Instrument, welches im Allgemeinen auf Technologieentwicklungen anwendbar ist. Diese umfassen damit auch auf jene, welche gesellschafts- und konsumentenorientiert sind, wie etwa die marktorientierte Entwicklung von KI-Systemen. Wir halten dies für relevant, da die Diversity-Folgenabschätzung eine gute Grundlage bietet, um ethische Überlegungen und Folgenabschätzungen schon in den Entwicklungsprozess zu integrieren und damit als Lösungsansatz für das auftretende Collingridge-Dilemma betrachtet werden kann. Die Diversity-Folgenabschätzung versucht über einen frühen Einbezug von diversen Perspektiven in den Entwicklungsprozess ethische Folgen zu erkennen und zu berücksichtigen, sodass zukünftige Folgen für die Gesellschaft und das Unternehmen abschätzbar werden und reguliert werden können. Die erfolgreiche Auflösung des Collingridge-Dilemmas durch diesen Ansatz ist jedoch an weitere Faktoren geknüpft, welchen die Diversity-Folgenabschätzung noch nicht in vollem Umfang begegnen kann und damit weiterhin Risiken bestehen Folgen in einem frühen Stadium nicht regulieren zu können, da sie in ihrer Tragweite nicht erkannt wurden.

Dies ist insbesondere in Bezug auf die Berücksichtigung unternehmensinterner Machtstrukturen der Fall. Macht und Hierarchien im Unternehmen können ein Problem für die effiziente Durchführung der Diversity-Folgenabschätzung darstellen, da in hierarchisch organisierten Unternehmen Entscheidungsträger oft großen Einfluss darauf haben, welche Risiken und ethischen Folgen überhaupt thematisiert werden. Dies kann zur Folge haben, dass bestimmte Folgen oder Betroffenheiten nicht erkannt und nicht reguliert werden, wodurch sich kostspielige und gravierende ethische Folgen bei der Implementierung der Entwicklungen ergeben können. Das Collingridge-Dilemma wird dementsprechend nicht gelöst.

In genauerer Betrachtung der von der CDR-Initiative vorgestellten Diversity-Folgenabschätzung zeigt sich dies insbesondere in der ersten Phase der Identifikation der Stakeholder und der zweiten Phase der Durchführung des Workshops, wo Machtgefälle und Hierarchien im Unternehmen ein Hemmnis der erfolgreichen Umsetzung darstellen können. In der Diversity-

Folgenabschätzung, wie sie die Initiative vorschlägt wird insbesondere auf eine verantwortungsvolle Durchführung der Folgenabschätzung basierend auf einem Top-down-Ansatz, d.h. dass die Bewertung und Umsetzung von Maßnahmen hauptsächlich von führenden Akteuren und zentralen Entscheidungsträgern gesteuert werden, gesetzt, was durchaus ihre Berechtigung hat, da Führungskräfte über eine gute Unternehmenskenntnis und Erfahrung in der Risikobewertung verfügen (vgl. CDR-Initiative 2024: 15–17). Damit besteht jedoch die Gefahr, dass Stakeholder nicht einbezogen werden, weil die Hierarchien den Blickwinkel beschränken können, so dass Stakeholder nur auf bestimmten Hierarchieebenen verortet werden. Dabei haben die Umsetzenden zwar mit den vorgeschlagenen Diversitäts-Dimensionen eine Richtlinie und Orientierung, wobei gerade diese ebenfalls eine Beschränkung darstellen kann, wenn zu wenige Dimensionen seitens der Leitenden identifiziert und für relevant befunden wurden. In Anbetracht einer erfolgreichen Abschätzung aller relevanter Folgen sollte eine Einschätzung in der Identifikation der Stakeholder sich nicht allein an der Relevanzeinschätzung der mit der Diversity-Folgenabschätzung Betrauten oder entlang der schon identifizierten Dimensionen orientieren, sondern sich hierarchiefreier gestalten. Um dies zu erreichen, sollten in der Ermittlung der relevanten Stakeholder partizipative Formate, wie breite Umfragen oder offene Gesprächsformate eingesetzt werden, sodass zunächst jedem eine Partizipationschance eingeräumt wird.

Uns ist bewusst, dass auch dies zu Verzerrungen führen kann, wenn sich bestimmte Stakeholder überproportional beteiligen und andere nicht, weshalb wir eine dynamische und zweigleisige Strategie, über eine Relevanzeinschätzung auf Grundlage der Diversitätsdimensionen seitens der mit der Folgenabschätzung betrauten Personen und zudem über partizipative Prozesse unter allen Mitarbeitern vorschlagen. Dynamisch meint hier, dass beide Prozesse miteinander im Austausch und in gegenseitiger Korrektur stehen.

Die Grundidee eines gleichberechtigten Einbeugs aller Mitarbeiter ist auch in der Workshophase zu beachten. Hier ist vor allem bedeutsam, dass es sich um ein Kommunikationsformat argumentativen Charakters unter Beteiligung unterschiedlicher Perspektiven und Schwerpunktsetzungen, wie etwa stärker mitarbeiterorientierten Forderungen oder finanziellen Betrachtungen handelt. Um dies in einen ethischen Kontext zu setzen, spricht dies die Diskursethik von Jürgen Habermas an, die eine Orientierung an einer idealen Sprechsituation vorschlägt (vgl. Habermas 2022). Es geht darum, dass ein legitimer Konsens auf rationaler Diskussion und dem bestmöglichen Ausgleich der Interessen aller Beteiligten basiert, anstatt durch einseitige Machtinteressen

oder emotionale Überzeugungsstrategien geformt zu werden (vgl. Seiler 2014: 36). Eine ideale Sprechsituation beinhaltet demnach unter anderem eine gleichberechtigte und hierarchiefreie Kommunikationsstruktur, in welcher unabhängig von der Person die Güte des Arguments von Bedeutung ist (vgl. ebd.: 32–36). Dies ist bedeutsam, damit sich das Ergebnis nicht allein aus Partikularinteressen zusammensetzt, sondern in einem Konsens begründet liegt. In Bezug auf die Workshophphase bedeutet dies, dass eine Relevanz darin besteht, diese hierarchiefrei zu organisieren, sodass es nicht zur Unterdrückung einzelner Perspektiven und keiner Berücksichtigung kommt. Dies kann erreicht werden, indem nicht nur alle Teilnehmenden mit gegenseitigem Respekt und einer Gesprächsbereitschaft in den Dialog treten, sondern auch über eine neutrale Leitung des Diskurses durch eine Person, welche keine Anliegen in der Debatte vertritt, sondern sich der Wahrung einer gleichberechtigten Diskurssituation verschreibt.

5.2 Zukünftige Stakeholder

Während sich die CDR-Initiative in Bezug auf Hemmnisse für eine Diversity-Folgenabschätzung schon in Teilen mit Macht und Hierarchieaspekten beschäftigt, möchten wir auf einen weiteren sehr relevanten Aspekt aufmerksam machen, welcher bislang keine Berücksichtigung findet und die erfolgreiche Auflösung des Collingridge-Dilemmas insbesondere bei der Übertragung des Ansatzes der Diversity-Folgenabschätzung auf konsumentenorientierte Technologieentwicklung gefährdet. Dabei handelt es sich um die Berücksichtigung von Folgen auf zukünftige Generationen oder im unternehmensinternen Kontext auch auf zukünftige Mitarbeiter.

Unsere Beziehung zu zukünftigen Generationen ist dabei jedoch von Asymmetrien geprägt. Neben der Tatsache, dass keine gegenseitige Interaktion stattfinden kann, besteht zudem ein asymmetrisches Machtverhältnis, indem wir mit unseren Entscheidungen maßgeblichen Einfluss auf die Lebensbedingungen, als auch die Anzahl an zukünftigen Menschen haben (vgl. Meyer 2021). Seitens zukünftiger Generationen gibt es keine Handlungsmöglichkeit ihre Anliegen zu vertreten, da sie noch nicht existieren, aber auch, weil ihre Existenz in einem hohen Maße von uns abhängt. Dies bedeutet, dass wir in unseren Entscheidungen und Entwicklungen nicht nur einen Einfluss auf ihre Existenz oder ihre Lebensbedingungen haben, sondern auch darauf, was diese Menschen in Bezug auf ihre Identität ausmacht und in welchen gesellschaftlichen Strukturen sie leben (vgl. ebd.). Auch wenn ihre bloße Existenz in der Zukunft sie nicht in ein Verantwortungsverhältnis zu uns setzt, so besteht dennoch eine Verantwortung basierend auf der Annahme,

dass zukünftige Generationen basierend auf ihren zukünftigen Interessen in der Zukunft Rechte haben werden und wir diese Rechte über die Beeinflussung ihrer Interessen mit unseren Entscheidungen berühren (vgl. Hoerster 1991: 98–102).

Auch unsere Entscheidungen bezüglich der Entwicklung neuer Technologie berühren die Lebenswirklichkeit zukünftiger Generationen. Nicht nur über die pure Existenz der neuen Technologie, sondern, wie wir zu Beginn des Artikels argumentieren, auch darüber, dass die Entwicklung neuer Technologie die Strukturen und Prinzipien der Gesellschaft beeinflussen kann. Entwickeln wir beispielsweise KI-Systeme, welche auf vorurteilsbehafteten Daten basieren und diskriminierende Strukturen in der Gesellschaft über ihren Einsatz in Entscheidungsprozessen, wie der Jobvergabe dahingehend prägen, dass Chancengleichheit beispielsweise in Abhängigkeit vom Geschlecht definiert wird, so wird sich dies in einer heute noch nicht vollständig bestimmbaren Weise auf die zukünftige Gesellschaft auswirken und ihre Zuschreibung von Rechten bestimmen.

Wir tragen demnach in der Entwicklung eine Verantwortung für die Auswirkungen der Technologie auf zukünftige Generationen, wodurch, wie Emma Ruttkamp-Bloem anführt, zukünftige Generationen in der Technologieentwicklung als Stakeholder berücksichtigt werden sollten (vgl. Ruttkamp-Bloem 2023: 20–24). Im Kontext des Collingridge-Dilemmas ist dieser Aspekt höchst relevant, da die Auswirkungen unserer heutigen Entscheidungen über technologische Entwicklungen noch nicht genau bestimbar sind. Dennoch besteht eine Verantwortung, diese Auswirkungen zu berücksichtigen, da ihre Auswirkungen oft schwer umkehrbar sind, sobald sich die Technologie etabliert hat und auch weil wir zu einem gewissen Grad beeinflussen, welche Form der gesellschaftlichen Prinzipien fortgeführt wird. Eine Auflösung des Dilemmas muss diese Auswirkungen deshalb über die Aufnahme zukünftiger Generationen als Stakeholder berücksichtigen. Dabei wird in einer der Einbeziehung von zukünftigen Stakeholdern auch eine verstärkt zukunftsorientierte Sichtweise implementiert, die Risiken nicht nur in Bezug auf uns einbezieht, sondern auch auf zukünftige Generationen.

Die Diversity-Folgenabschätzung bezieht zukünftige Generationen, oder in ihrem unternehmensinternen Kontext zukünftige Mitarbeiter*innen, nicht als Stakeholder mit ein, was dazu führt, dass die Folgenabschätzung bezüglich einer möglichen Auflösung des Collingridge-Dilemmas Schwächen aufweist. Wir geben zu, dass es eine Herausforderung darstellt Interessen und Folgen für zukünftige Generationen abzuschätzen, dennoch sehen wir eine Relevanz in diesem

Ansatz, da das Risiko verringert werden kann, dass sich problematische Technologien unumkehrbar etablieren. In einem praktischen Sinne ist die Berücksichtigung zukünftiger Stakeholder herausfordernd, da die Perspektiven zukünftiger Stakeholder, welche in der Diversity-Folgenabschätzung frühzeitig implementiert werden sollen, damit sich ethische Auswirkungen einer Technologie abschätzen lassen, noch nicht direkt bestimbar sind. Dennoch kann wahrscheinlich allein das Bewusstsein über zukünftige Stakeholder schon eine bedeutende Rolle in der Verbesserung der Folgenabschätzung spielen, da zukünftige Risiken nicht mehr allein in Bezug auf den Nutzen für heutige Stakeholder diskontiert werden, sondern auch auf Zukünftige, was den zukünftigen Gefahren mehr Gewicht im Entscheidungsprozess verleihen würde. Im unternehmensinternen Kontext würde dies bedeuten, dass man z. B. vor dem Hintergrund des demographischen Wandels potenzielle Perspektiven zukünftiger jungen Generationen in einer alternden Gesellschaft betrachtet und in Bezug auf Diskriminierungsrisiken für Personalstrategien evaluiert. Die Diversity-Folgenabschätzung sollte demnach an die Verantwortung gegenüber zukünftigen Generationen appellieren und zukünftige Generationen als festen Stakeholder ausweisen, sodass diese Perspektive Berücksichtigung findet.

Im Kontext eines Unternehmens können im Einbezug von Interessen zukünftiger Generationen jedoch auch Anreizprobleme entstehen, da die Kosten für zukünftige Generationen nicht mehr die eigenen sind. Die Diversity-Folgenabschätzung wird diese Anreizprobleme nicht vollständig lösen können, da sie auf der Grundlage operiert die Relevanz in der Berücksichtigung ethischer Überlegungen in der Folgenabschätzung, neben dem Verweis auf die unternehmerische Verantwortung, an unternehmenseigene Anreize, wie ihr Profitieren von einer diversen Unternehmenskultur, zu knüpfen. Während die anderen Punkte, welche für die Relevanz einer Berücksichtigung ethischer Überlegungen sprechen, sich neben der unternehmerischen Verantwortung in der Gesellschaft auch an ein Eigeninteresse knüpfen lassen, so ist letzteres in Bezug auf zukünftige Generationen schwerer zu argumentieren, was die Auflösung des Collingridge-Dilemmas gefährdet. In diesem Punkt ist es bezüglich der Auflösung des Collingridge-Dilemmas schlussendlich wahrscheinlich effektiver übergeordnete und institutionelle Regulierungen zu implementieren, um die Berücksichtigung zukünftiger Stakeholder sicherzustellen, was nicht bedeutet, dass ein Nicht-Bestehen solcher Regulierungen Unternehmen von ihrer Verantwortung entbindet. Es bedeutet lediglich, dass die Auflösung des Collingridge-Dilemmas nicht vollständig in heutigen Interessen begründet liegt und damit aus wirtschaftlichen Anreizen nur so weit bestreitbar ist, als

dass Unternehmen von einer stärker zukunftsorientierten Schätzung und Strategie profitieren werden. Dies bedeutet, dass eine Auflösung des Collingridge-Dilemmas auch dadurch geprägt ist, dass es unternehmerische als auch politische Beiträge gibt, was die Relevanz von Regularien in der Technologieentwicklung statt eines Laissez-Faire-Ansatzes verstärkt.

Die Diversity-Folgenabschätzung kann im Allgemeinen als gute Grundlage zum Versuch einer Auflösung des Collingridge-Dilemmas verstanden werden, wobei sie Verbesserungspotenzial hinsichtlich ihrer Sensitivität gegenüber Macht und Hierarchiestrukturen in einem Unternehmen als auch ihrer Zukunftsorientiertheit aufweist. Insbesondere der letzte Aspekt hat jedoch auch gezeigt, dass nicht nur die Diversity-Folgenabschätzung in der Berücksichtigung der Verantwortung zukünftiger Stakeholder vor einer Herausforderung steht, sondern ebenfalls die Möglichkeit der Auflösung des Collingridge-Dilemmas betroffen ist, da Perspektiven zukünftiger Stakeholder noch nicht direkt identifizierbar sind und damit erst in der Zukunft sichtbar sein werden.

6. Ethische Verortung

Unsere Überlegungen zu Gerechtigkeit und Diversität sowie die Prinzipien der Diversity-Folgenabschätzung lassen sich auf bedeutende ethische Hintergründe zurückführen. Bezüglich des Diskurses zur Gerechtigkeit kommt der Theorie von John Rawls eine besondere Bedeutung zu, wobei ebenso Immanuel Kants Moraltheorie und die Überlegungen Jürgen Habermas' zur demokratischen Gesellschaft Implikationen für die Unternehmensethik und Technologieentwicklung mit sich bringen. Die ethischen Aspekte der Diversity-Folgenabschätzung lassen sich demnach folgendermaßen darstellen.

Schon zu Beginn dieses Artikels haben wir die Frage, warum wir uns mit der Diversity-Folgenabschätzung befassen sollten, adressiert. Die Antwort auf diese Frage haben wir vor allem auf Überlegungen zu einer fairen Gesellschaft und unsere Verantwortung diese zu fördern gestützt, welche sich unter anderem auf die Theorie der Gerechtigkeit des amerikanischen Philosophen John Rawls zurückführen lassen. Dabei wird eine gerechte Gesellschaft mit freien und gleichen Bürgern an Prinzipien der Fairness geknüpft, die die Gleichheit der Grundfreiheiten und eine gerechte Verteilung von Ressourcen und Chancen sicherstellen sollen. Diese Prinzipien stellen ein Fundament für eine gerechte Gesellschaft dar, in der alle Bürger gleiche Chancen erhalten, unabhängig von ihrer Herkunft, sozialen Stellung oder anderen Merkmalen (vgl. Wenar 2021). In

der Umsetzung geht damit eine Verantwortung einher, systemische Barrieren abzubauen, die bestimmte Gruppen benachteiligen. Dies erfordert eine aktive Förderung von Inklusion und Gerechtigkeit, indem Ressourcen und Chancen auf eine Weise verteilt werden, die die Bedürfnisse und Rechte aller berücksichtigt. Wenn wir Rawls' Theorie auf den digitalen Bereich anwenden, dann ist ein gerechtes digitales System dasjenige, welches von allen Menschen frei und gleichberechtigt genutzt werden kann und weder im Zugang noch in der Nutzung Benachteiligungen beinhaltet. Die Fairness im Kontext der Digitalisierung und insbesondere auch in der Entwicklung von KI bedeutet damit, dass alle Menschen die gleiche Möglichkeit haben, sie zu benutzen, ohne dass Vorurteile basierend auf Geschlecht, Herkunft oder digitaler Kompetenz sie exkludieren. Wie wir gesehen haben, versucht die Diversity-Folgenabschätzung genau dies, indem sie in der Technologieentwicklung unterschiedliche Bedürfnisse berücksichtigt, um spätere Benachteiligungen zu verhindern.

Ein zweiter Aspekt unseres Beitrags besteht in der Frage, wie Unternehmen ethische Überlegungen in die Entwicklung ihrer Technologie im Sinne einer Auflösung des Collingridge-Dilemmas einbinden können. Die Diversity-Folgenabschätzung stellt dabei eine bedeutende Grundlage dar. Dabei ist jedoch zu bemerken, dass allein die Umsetzung dieser noch kein Nachkommen der moralischen Verantwortung eines Unternehmens darstellt. Laut Kant wird moralische Qualität nicht durch die Handlungen selbst bestimmt, sondern durch den guten Willen, d.h. die Absicht, die hinter einer Handlung steht. Er argumentiert, dass Handlungen nicht vollkommen gut oder schlecht sein können, sondern stattdessen ein guter oder schlechter Wille besteht (vgl. Kant 1781/1990; Hübner 2014). Anders gesagt, sollte jedes moralische Gesetz eine richtige Basis haben, wobei sein moralischer Wert genauer in der Maxime, die ihm zugrunde und in ihm verwirklicht ist, liegt (vgl. ebd.: 169). Damit liegt die Verantwortung eines Unternehmens auch darin eine ernsthafte moralische Betrachtung vorzunehmen und sich selbst an moralische Prinzipien zu binden, statt lediglich auf externe Ansprüche zu reagieren, was auch von der CDR-Initiative betont wird (vgl. CDR-Initiative 2024: 8–12). Angewandt auf den Kontext der digitalen Fairness und Diversity-Folgenabschätzung bedeutet dies, dass Unternehmen Diversität und Inklusion nicht nur als Ziel anstreben sollten, sondern dies auch aus einem echten ‚guten Willen‘ heraus tun. Dies ist bedeutsam, damit die Diversity-Folgenabschätzung erfolgreich sein kann und gesellschaftliche Veränderungen ernsthaft angegangen werden können.

Ein dritter Aspekt betrifft die Frage, warum wir eigentlich die Stakeholder benötigen. Können wir nicht lediglich eine Checkliste nutzen, um die Probleme zu identifizieren? Die Antwort auf diese Frage liegt in der Bedeutung demokratischer Prozesse und der Bedeutung des Gesprächs für eine effiziente Verbesserungskultur und Förderung von Inklusion und Diversität. Jürgen Habermas argumentiert, dass nur durch den Dialog, eine nötige Kohärenz und Legitimität der Entscheidungen erreicht werden kann, um die Herausforderungen der modernen Welt zu bewältigen und eine gerechtere, integrativere Zukunft zu gestalten (Habermas 2022). Dies bedeutet, dass angesichts der Vielfalt in unserer Gesellschaft ein kontinuierlicher Dialog notwendig ist, sodass sichergestellt ist, dass die unterschiedlichen Perspektiven in Lösungen berücksichtigt werden und keine Exklusion stattfindet. Demzufolge sind die Stakeholder wichtig für die Diversity-Folgenabschätzung, weil es ohne ihre Einbindung an demokratischen Strukturen, wie der des Meinungsaustauschs und des Einbeugs von verschiedenen Perspektiven als Korrektivfunktion fehlt. Damit stützt sich die Diversity-Folgenabschätzung auf bedeutende ethische Prinzipien, was sie zu einem wichtigen Instrument für die Förderung von Inklusion und Gerechtigkeit.

7. Schlussbemerkung

Die Berücksichtigung von Diversität und Gerechtigkeit in der Entwicklung technologischer Neuerungen stellt ein bedeutsames Thema für die Gesellschaft, als auch für Unternehmen dar, weil Technologieentwicklungen einen bedeutsamen Einfluss auf unsere gesellschaftlichen und ökonomischen Strukturen und Wertvorstellungen haben können. Politische Organisationen, wie die Europäische Union reagieren auf diese Zusammenhänge, aufgrund auch negativer und Diskriminierender Wirkungen mit ethischen Regularien zum Einsatz und Entwicklung von Technologie. Doch auch einzelnen Unternehmen kommt unabhängig von politischen Regelungen eine Verantwortung in der Implementierung ethischer Standards in der Entwicklung zur Risikominimierung zu. Ihre Herausforderung besteht dabei in der Auflösung des Collingridge-Dilemmas. Wir haben in diesem Artikel die Diversity-Folgenabschätzung, als einen Lösungsansatz betrachtet, wobei sich ergab, dass sie dieses Dilemma weitestgehend adressieren kann, insbesondere deshalb, weil sie eine Basis für eine unternehmerische Beschäftigung mit den ethischen Folgen in der Technologieentwicklung bietet und damit auch im Falle einer politischen Laissez-Faire-Situation wirksam ist. Dabei besitzt sie das Potenzial auf konsumentenorientierte

Technologieentwicklungen, statt nur unternehmensinterne Anwendungen ausgeweitet zu werden. Dennoch weist sie Schwächen in Bezug auf den Einbezug zukünftiger Stakeholder und ihrer Sensitivität gegenüber Machtstrukturen auf, welche es insbesondere auch vor dem Hintergrund der Lösung der ethischen Variante des Collingridge-Dilemmas über den Einbezug eines Bewusstseins für Werteveränderungen durch Technologie und zukünftige Stakeholder anzugehen gilt. Insgesamt stellt die Diversity-Folgenabschätzung jedoch bereits einen guten Weg in der Entwicklung inklusiver Technologien dar, indem sie versucht beide Pole des Dilemmas zu verbinden, wodurch sie einen wichtigen Schritt darstellt auch in der Zukunft die ethischen Konsequenzen der Digitalisierung zu adressieren und Unternehmen ein Instrument zur Mitarbeit an einer gerechten Gesellschaft zu bieten.

Literaturverzeichnis

- O'Brian, M. / Parvini, S. (2024): How Harris and Trump Differ on Artificial Intelligence Policy, APNews, URL: <https://apnews.com/article/ai-artificial-intelligence-kamala-harris-trump-vance-1c94ba2ee6a1410e61d980faa51b19e4> (aufgerufen am: 06/02/2025).
- Collingridge, D. (1980): The Social Control of Technology, New York: St. Martin's Press.
- CDR-Initiative (Hrsg.) (2024): Diversity-Folgenabschätzung. Praxisleitfaden zur Förderung von Diversität, Inklusion und Chancengleichheit in der Digitalisierung, Berlin, URL: <https://cdr-initiative.de/news/diversity-folgenabschaetzung> (aufgerufen am: 17/02/2025).
- Croy, M. J. (1996): Collingridge and the Control of Educational Computer Technology, in: Techné: Research in Philosophy and Technology, Jg. 1 / Nr. 3–4, 107–115.
- Dakhli, A. (2021): The Impact of Corporate Social Responsibility on Firm Financial Performance: Does Audit Quality Matter?, in: Journal of Applied Accounting Research, Jg. 23 / Nr. 5, 950–976.
- Demos Helsinki (2022): What Is the Collingridge Dilemma and Why Is It Important for Tech Policy?, URL: <https://demoshelsinki.fi/what-is-the-collingridge-dilemma-tech-policy/> (aufgerufen am: 05/02/2025).
- Europäisches Parlament (2023): KI-Gesetz: erste Regulierung der künstlichen Intelligenz, 08.06.2023, URL: <https://www.europarl.europa.eu/topics/en/article/20230601STO93804/eu-ai-act-first-regulation-on-artificial-intelligence> (aufgerufen am: 23/03/2025).
- Europäische Union (2024): Verordnung (EU) 2024/1689 des Europäischen Parlaments und des Rates vom 13. Juni 2024 zur Festlegung harmonisierter Vorschriften für künstliche

- Intelligenz, in: Amtsblatt der Europäischen Union, URL: <https://eur-lex.europa.eu/eli/reg/2024/1689/oj?locale=de> (aufgerufen am: 23/03/2025).
- Friedrich-Ebert-Stiftung (2022): Jürgen Habermas: Ein neuer Strukturwandel der Öffentlichkeit und die deliberative Politik, URL: <https://www.fes.de/akademie-fuer-soziale-demokratie/buch-essenz/juergen-habermas-2022-ein-neuer-strukturwandel-der-oeffentlichkeit-und-die-deliberative-politik> (aufgerufen am: 04/02/2025).
- Habermas, J. (2022): Reflections and Hypotheses on a Further Structural Transformation of the Political Public Sphere, in: Theory, Culture & Society, Jg. 39 / Nr. 4, 145–171.
- Hoerster, N. (1991): Abtreibung im säkularen Staat. Argumente gegen den § 218, Frankfurt: Suhrkamp.
- Hübner, D. (2014): Deontologie: Das richtige Handeln, in: Hübner, D. (Hrsg.): Einführung in die philosophische Ethik, Göttingen: Vandenhoeck & Ruprecht, 151–205.
- Kant, I. (1781/1990): Kritik der reinen Vernunft [KrV], hg. von Raymund Schmidt, Hamburg: Felix Meiner 1990.
- Kudina, O. / Verbeek, P. (2019): Ethics from Within: Google Glass, the Collingridge Dilemma, and the Mediated Value of Privacy, in: Science, Technology, and Human Values, Jg. 44 / Nr. 2, 291–314.
- Lauer, D. (2021): You Cannot Have AI Ethics Without Ethics, in: AI Ethics, Jg. 1 / Nr. 9, 21–25.
- Meyer, L. (2021): Intergenerational Justice, in: Zalta, E. N. (Hrsg.): The Stanford Encyclopedia of Philosophy, URL: <https://plato.stanford.edu/archives/sum2021/entries/justice-intergenerational> (aufgerufen am: 17/02/2025).
- Page, S. E. (2007): The Difference: How the Power of Diversity Creates Better Groups, Firms, Schools, and Societies, Princeton: Princeton University Press.
- Rawls, J. (1999): A Theory of Justice: Revised Edition, Cambridge: Harvard University Press.
- Ruttkamp-Bloem, E. (2023): Intergenerational Justice as Driver for Responsible AI, Conference Paper, in: Pillay, A. / Jembere, E. / Gerber, A. J. (Hrsg.): Artificial Intelligence Research, SACAIR 2023, Communications in Computer and Information Science, Jg. 1976, Cham: Springer, 12–28.
- Seeliger, M./ Sevignani, S. (2022): A New Structural Transformation of the Public Sphere? An Introduction, in: Theory, Culture & Society, Jg. 39 / Nr. 4, 3–16.
- Seiler, C. (2014): Die Diskursethik (Habermas), in: Die Diskursethik im Spannungsfeld von Systemtheorie und Differenzphilosophie, Wiesbaden: Springer VS, 31–52.
- Spiegel (2025): Trump streicht Vorgaben für Kontrolle von künstlicher Intelligenz. Trotz Warnungen vor immensen Risiken, URL: <https://www.spiegel.de/netzwelt/trump-streicht->

vorgaben-fuer-kontrolle-von-kuenstlicher-intelligenz-a-fbed93e3-fe70-46f3-ab1e-891d7b81714f (aufgerufen am: 08/02/2025).

Strasser, K. / Niedermayer, B. (2021): Unvoreingenommenheit von Künstliche-Intelligenz-Systemen. Die Rolle von Datenqualität und Bias für den verantwortungsvollen Einsatz von künstlicher Intelligenz, in: Altenburger, R. / Schmidpeter, R. (Hrsg.): CSR und Künstliche Intelligenz. Management-Reihe Corporate Social Responsibility, Wiesbaden: Springer Gabler, 121–135.

Tigard, D. (2022): Embedded Ethics as Preparatory Regulation of Technology: a New Solution to the Collingridge Dilemma?, in: Lawrence, D. / Morley, S. (Hrsg.): Novel Being. Regulatory Approaches for a Future of New Intelligent Life. Elgar Law, Technology and Society series, Cheltenham: Edward Elgar Publishing, 12–28.

Wenar, L. (2021): John Rawls, in: Zalta, E. N. (Hrsg.): The Stanford Encyclopedia of Philosophy, URL: <https://plato.stanford.edu/archives/sum2021/entries/rawls/> (aufgerufen am: 17/02/2025).