

# Kurzreview: Identifikation offener Forschungsfragen zur Integration ethischer Anforderungen in Softwareentwicklungsprozesse

Hochschule Harz  
Stabsstelle Forschung  
Christian Reinboth, M.Sc.  
Version 1.3 vom 26.01.2022

## Inhalt

1 Einführung und Zielstellung.....	2
2 Methodisches Vorgehen .....	4
3 Narrative Synthese .....	6
3.1 Problemwahrnehmung.....	6
3.2 Künstliche Intelligenz.....	7
3.3 Identifizierte Lösungsansätze .....	8
3.3.1 Bildung und Weiterbildung .....	8
3.3.2 Selbstverpflichtungen.....	9
3.3.3 Berufsethische Standards.....	10
3.3.4 (Selbst-)Evaluation.....	10
3.3.5 Transparenz .....	11
3.3.6 Ethical User Stories.....	11
3.3.7 Requirements Engineering .....	12
3.3.8 Ethics Penetration Testing.....	12
3.3.9 Agile Softwareentwicklung.....	12
4 Themenrelevante Forschungsvorhaben.....	13
5 Mögliche Forschungsfragen .....	14
5.1 Inhalte und Wirksamkeit von Selbstverpflichtungen in KMU .....	14
5.2 Umgang mit interkulturellen Differenzen .....	15
5.3 Problem der „fehlenden Werte“ .....	15
5.4 Identifikation von Entscheidungsebenen.....	16
6 Abschließende Empfehlungen.....	16
7 Verwendete Quellen .....	17

## 1 Einführung und Zielstellung

Die gesamtgesellschaftliche Bedeutung von Software nimmt seit etwa zwei Jahrzehnten beständig zu. Auch wenn Anwenderinnen und Anwender dies nicht immer reflektieren, wirkt Software sich an vielen Stellen längst nicht mehr nur auf unternehmerische, sondern auch auf individuelle und teils höchstpersönliche Entscheidungs- und Meinungsbildungsprozesse aus, wobei es neben von den Softwareherstellern beabsichtigten auch zu zahlreichen unbeabsichtigten Effekten kommt [Hussain et al. 2018, Zuber et al. 2021]. Vor diesem Hintergrund stellt sich die Frage, inwieweit menschliche Ethik – und menschliche Werte – bei der Entwicklung von Algorithmen etwa zur Bewerberauswahl in Unternehmen oder zur Verbreitung politischer Inhalte in Sozialen Netzen Berücksichtigung finden [Whittle 2019]. Da selbst hochkomplexe Software, die dem Bereich der Künstlichen Intelligenz zuzuordnen ist, nicht als eigenständiger moralischer Akteur betrachtet werden kann, sondern stets ausführend bleibt, entscheiden letztendlich immer Menschen darüber, ob sie moralisch akzeptabel agiert [Filipovic et al. 2018, Zuber et al. 2021]. Damit unterliegt Software – wie auch sonst „alles, was die zivilisatorische Lebenswelt ausmacht“ [Krüger 2021, S. 366] – den mit menschlichem Handeln inhärent verbundenen Voreingenommenheiten, Befangenheiten, Neigungen und Fehlvorstellungen. Und da es die Komplexität moderner Software sowie die Vielfalt möglicher Einsatzgebiete und Anwendungsszenarien unmöglich machen, die ethische Angemessenheit von Outputs auf Einzelfallbasis zu untersuchen, liegt der Fokus der Values-in-Design-Forschung auf ihrem Entstehungsprozess und damit auf der Frage, wie (und von wem) überhaupt ethische Anforderungen an Software definiert werden, wie mit diesen Anforderungen während der Entwicklung umgegangen und wie ihre Umsetzung überprüft wird.

Die Relevanz dieser Thematik für eine Vielzahl gesellschaftlicher Bereiche soll nachfolgend schlaglichtartig an einzelnen Problembeispielen aufgezeigt werden, wobei bei Interesse bezüglich näherer Ausführungen zu den benannten Exempeln auf die jeweiligen Referenzen verwiesen wird.

- Soll ein selbstfahrendes Auto bei einem nicht mehr zu verhinderndem Unfall das Leben des Passagiers gefährden, um eine größere Anzahl anderer Verkehrsteilnehmer zu retten? Falls ja, anhand welcher Kriterien soll entschieden werden, welchen gefährdeten Personen(gruppen) bei Manövern zur Unfallverhinderung Rettungspriorität eingeräumt wird? [Zwingmann und Gärtner 2021, Järvinen 2017]
- Soll eine Social Media-Plattform psychologische Erkenntnisse über Suchtverhaltensmuster nutzen dürfen, um Anwenderinnen und Anwender durch sogenanntes Addictive Design zu exzessivem und langfristig ungesundem Nutzungsverhalten zu „erziehen“ und von plattformeigenen Belohnungsmechanismen (z.B. Dopaminausschüttung bei „Likes“) abhängig zu machen? [Cemiloglu et al. 2020, Hansen 2019, Krüger 2021]
- Kann die Bewertung von Klausuren und Aufsätzen im Rahmen eines Studiums oder einer Ausbildung oder auch die Bewertung von Bewerbungsunterlagen und Kreditanträgen in einem Unternehmen vollständig oder teilweise durch Künstliche Intelligenz übernommen werden, um eine bewusste oder unbewusste Schlechterstellung etwa aufgrund von Vorurteilen gegenüber Personengruppen auszuschließen? Wie lässt sich vermeiden, dass eine derartige Software, die mit historischen Daten trainiert wird, in der Vergangenheit erfolgte Schlechterstellungen (etwa bei Kreditanfragen von Menschen mit Wohnsitz in einem Niedrigeinkommensgebiet oder bei Bewerbungen von Personen ausländischer Herkunft) in die Zukunft fortschreibt? [Zwingmann und Gärtner 2021, Whittle 2019, Beckert 2021]

- Sollte ein Softwareentwickler mögliche ökologische und klimatische Auswirkungen für den Erfolgsfall einer Software im Vorfeld berücksichtigen müssen? (Beispiele hierfür sind der massive Energieverbrauch beim Mining von Crypto-Währungen wie Bitcoin oder die Verdrängung von ÖPNV-Angeboten durch weniger klimafreundliche Services wie Uber.)
- Wie sollte ein System zur Pflegeunterstützung ausgestaltet sein, das die Trinkmengen älterer Menschen überwacht und diese ggf. zur Einnahme von Getränken auffordert? Wie intrusiv sollte ein solches System handeln und zu welchen Mittel sollte es greifen dürfen, um die Compliance zu verbessern? [Mittelhorn 2019]

Eine Vielzahl aktueller Realbeispiele zeigt, dass die oben aufgeführten Fragen alles andere als rein akademischer Natur sind. Insbesondere beim Einsatz von KI-gestützten Systemen kam es in der Praxis bereits vielfach zu – größtenteils nicht intendierten – ethisch unerwünschten Ergebnissen.

- Während des Hurrikans Irma (2017) registrierten die für die Kalkulation der Ticketpreise eingesetzten Systeme verschiedener Airlines die sprunghaft steigende Nachfrage an Flügen aus dem Katastrophengebiet und erhöhten die Preise um bis zu 600%. [Hussain et al. 2018]
- Die wachsende Popularität der Ferienwohnungs-Plattform Airbnb führte nachweislich zu Preissteigerungen und sozialen Verwerfungen in innerstädtischen Wohnungsmärkten, wodurch es insbesondere zur Verdrängung weniger wohlhabender Personen und Familien aus ihren bisherigen Lebensmittelpunkten kam. [Pretzschner et al. 2021, Zuber et al. 2021]
- Der von Microsoft für Twitter entwickelte AI-Chatbot TayTweets, der durch Interaktionen mit menschlichen Nutzerinnen und Nutzern lernen und die eigenen kommunikativen Fähigkeiten beständig verbessern sollte, musste nach seinem Release im Jahr 2016 bereits nach wenigen Stunden abgeschaltet werden, da er sich rassistische, misogynie, ableistische und antisemitische Ausdrucksweisen angeeignet hatte. Die Frage, ob es den Holocaust gegeben habe, verneinte das System auf Basis des in den Netzdiskussionen innerhalb kürzester Zeit „erlernten“ Wissens. [Strasser und Niedermayer 2021]
- Die österreichische Arbeitsagentur gab 2018 eine Software in Auftrag, die die zukünftigen Arbeitsmarktchancen von Arbeitssuchenden beruhend auf Alter, Ausbildung und anderen demografischen Faktoren einschätzen sollte. Basierend auf der Einteilung in drei Kategorien wurde dann über Unterstützungsmaßnahmen entschieden, wobei die meisten Mittel für Personen in der mittleren Kategorie verausgabt werden sollten. Die Einstufung wirkte sich hier also ganz konkret auf die individuelle Chance auf Unterstützung aus. [Strasser und Niedermayer 2021]
- Von Such- und Empfehlungsalgorithmen auf zahlreichen Online-Plattformen und in Sozialen Netzwerken ist bekannt, dass sie zur Ausbildung von Echokammern und Filterblasen beitragen und teils extreme politische oder religiöse Inhalte empfehlen, wenn diese zu einer Ausweitung der Interaktionszeiten beitragen. [Saurwein et al. 2017]

Wie die Breite der aufgeführten Problembeispiele zeigt, geht die Betrachtung weit über die reine Produkthaftung hinaus – vielmehr geht es um die Frage „welche Regeln gelten sollen, um ethische Maschinen entwickeln zu können und wie diese ausgestaltet sein müssen, damit sie eingehalten werden (können)“ [Zuber et al. 2021, S. 3]. Das hier vorgelegte kurзорische Kurzreview soll einen ersten Überblick des Themenfeldes liefern sowie die Identifikation derzeit offener Fragestellungen von zentraler Bedeutung sowie wesentlicher laufender Projekte im betrachteten Forschungsfeld unterstützen.

## 2 Methodisches Vorgehen

Die diesem Review zugrundeliegende Literaturrecherche wurde zwischen dem 03.01.2022 und dem 04.01.2022 in verschiedenen einschlägigen Publikationsdatenbanken (siehe nachfolgende Ausführungen) durchgeführt. Auf Basis einer oberflächlichen Themenfeldsuche in Google Scholar wurden für die vertiefte Suche insgesamt sechs Suchphrasen in deutscher und englischer Sprache festgelegt.

Schlagwort-Kombinationen in deutscher Sprache:

- „Ethik“ + „Softwareentwicklung“
- „Ethik“ + „Requirements Engineering“

Schlagwort-Kombinationen in englischer Sprache:

- „Ethics“ + „Software Development“
- „Ethics“ + „Software Engineering“
- „Ethics“ + „Software Design“
- „Ethics“ + „Requirements Engineering“

Auf die Nutzung der naheliegenden Schlagwort-Kombinationen „Ethik“ + „Software“ sowie „Ethics“ + „Software“ wurde verzichtet, da beide Anfragen zu unspezifisch sind und daher zu einer nicht zu bewältigenden Anzahl an Suchergebnissen führten, die nach kurзорischer Durchsicht der jeweils ersten Trefferseiten in vielen Fällen nicht dem inhaltlichen Spektrum dieser Recherche zuzuordnen waren.

Nach den sechs aufgeführten Suchphrasen wurde in den akademischen Suchmaschinen

- Google Scholar (<https://scholar.google.com>),
- ScienceDirect (<https://www.sciencedirect.com>),
- Springer Link (<https://link.springer.com>) und
- BASE (<https://www.base-search.net>)

sowie in den akademischen sozialen Netzwerken

- ResearchGate (<https://www.researchgate.net>) und
- Academia (<https://www.academia.edu>)

und in den akademischen Repositorien

- CORE (<https://hcommons.org/core/>) und
- SSOAR (<https://www.gesis.org/ssoar/home/>)

gesucht. Auf eine Beschränkung auf Datenquellen mit informationswissenschaftlichem Schwerpunkt wurde für diese Übersicht bewusst verzichtet, da der Eingang ethischer Anforderungen in Prozesse der Softwareentwicklung vielfältige geistes-, wirtschafts- und organisationswissenschaftliche Aspekte berührt, die gerade für die Identifikation offener – interdisziplinärer – Fragen von hoher Relevanz sind.

Um in die Grundmenge der nachfolgend gesichteten Quellen aufgenommen zu werden, mussten die über die Suchen aufgefundenen Veröffentlichungen neben einer im Titel oder Abstract bereits klar erkennbaren thematischen Relevanz die nachfolgend aufgeführten Kriterien erfüllen:

- Publikation in deutscher oder englischer Sprache
- Veröffentlichung im Fünf-Jahres-Zeitraum zwischen 2017 und 2021
- Frei (Open Access) oder über das Campusnetz der Hochschule Harz verfügbar

Die wie dargelegt durchgeführte Suche erbrachte – nach Bereinigung um Dopplungen – insgesamt 61 Publikationen mit hoher vermuteter Relevanz, die mittels der Literaturdatenbank Citavi erfasst, kategorisiert und systematisch ausgewertet wurden. Neben klassischer wissenschaftlicher Literatur wie Buchkapiteln, peer-reviewten Beiträgen in Fachzeitschriften sowie Beiträgen zu Tagungsbänden, wurden auch akademische Abschlussarbeiten sowie Dokumente aus dem Feld der Grauen Literatur (wie etwa White Paper und Behörden- oder Projektberichte) berücksichtigt. Nicht aufgenommen wurden dagegen reine Preprints ohne bereits erfolgte Publikation an anderer Stelle.

Die 61 erfassten Werke verteilen sich wie folgt auf diese Kategorien:

- 30 Beiträge in Sammel- oder Tagungsbänden
- 27 peer-reviewte Beiträge in Fachzeitschriften
- Drei Dokumente aus dem Bereich der Grauen Literatur
- Eine Abschlussarbeit (eine Masterarbeit an der Universität von Utrecht)

14 der aufgenommenen Publikationen sind in deutscher, 47 in englischer Sprache erschienen. Der Blick auf die Erscheinungsdaten der Beiträge zeigt eine leichte, wenn auch nicht signifikante Zunahme der als themenrelevant identifizierten Beiträge<sup>1</sup>.

2017	2018	2019	2020	2021
7	14	10	12	18

Der Fokus der nachfolgenden Auswertungen liegt auf Prozessen der Entwicklung legal nutzbarer Software für den nichtmilitärischen Einsatz in demokratisch verfassten Staatsformen. Zu Prozessen der Entwicklung illegaler Software (z.B. im Rahmen des medial stark rezipierten „VW-Skandals“) sei an dieser Stelle auf [Järvinen 2017] und [Rogerson 2017] verwiesen, während [Krüger 2021] einen kurzen Einblick in ethische Fragestellungen im Kontext militärischer Anwendungen und [Filipović et al. 2018] in das dystopische Social Credit-System der chinesischen Zentralregierung bieten. Auch die Umsetzung von Vorgaben der Allgemeinen Datenschutzgrundverordnung (DSGVO, siehe hierzu etwa [Josephina und Andreas 2019] oder [Weyns 2020]) ist kein Betrachtungsgegenstand, auch wenn der Datenschutz natürlich vielfältige ethische Implikationen hat (Autonomie über eigene Daten, Freiheit von verdeckter Meinungsmanipulation etc. pp.). Die Umsetzung von Anforderungen des Datenschutzes an Software ist jedoch gerade in Deutschland ein so weites und kompliziertes Feld, dass sie als eigenständiger Untersuchungsgegenstand zu betrachten ist.

---

<sup>1</sup> Eine feingliedrigere Betrachtung etwa nach Themenschwerpunkten, nach Nationalität oder Affiliation (Hochschulen oder Industrie) von Autorinnen und Autoren sowie nach Zitierungen kann grundsätzlich noch ergänzt werden, empfiehlt sich aber erst im Kontext einer fachlich konkreter spezifizierten Recherche.

### 3 Narrative Synthese

#### 3.1 Problemwahrnehmung

Verglichen mit der Vielzahl an Veröffentlichungen im Themenfeld des Software Engineerings, führte die für dieses Review durchgeführte Suche über einen Zeitraum von fünf Jahren (siehe Abschnitt 2) zu einer überschaubaren Anzahl von Publikationen. Wie die weit umfassendere Recherche von [Perera et al. 2020] zeigt, entspricht dieses Ergebnis dem derzeitigen Status Quo: Nur 16% von 1.350 der in den Jahren 2015-2018 in den Top-Journals zum Software Engineering publizierten peer-reviewed Paper gehen überhaupt auf menschliche Wertvorstellungen und ethische Fragen ein, wobei vor allem Sicherheit und Privatsphäre (kaum aber Werte wie Fairness, Offenheit oder Freiheit) angesprochen wurden.

Während zu Begriffen wie Ethical Software Development oder Value Sensitive Design zwar zunehmend mehr publiziert wird, das Themenfeld der Integration ethischer Zielvorgaben in Softwareentwicklungsprozesse aber im Gesamtvolumen aller Veröffentlichungen weiterhin von eher geringer Bedeutung ist, sind in der unternehmerischen Praxis erhebliche Anstrengungen erkennbar, um das Risiko potentieller imageschädigender Skandale durch die Aufstellung ethischer Leitlinien zu reduzieren. Bereits 2015 hatten 86 der 100 im FTSE 100-Index (Financial Times Stock Exchange Index, der wichtigste britische Aktienindex) gelisteten Unternehmen veröffentlichte „Value Statements“, „Codes of Ethics“ oder „Codes of Conduct“ [Whittle 2019]. Zwar wird – siehe Abschnitt 3.3.2 – die Aussagekraft und Wirksamkeit dieser Selbstverpflichtungen von der überwiegenden Anzahl der gesichteten Publikationen deutlich bestritten, gleichwohl ist ihre Existenz dennoch als Indikator dafür zu betrachten, dass die Problematik in der Wirtschaft erkannt wurde und an Lösungen – ob real wirksam oder primär dem Schutz des eigenen Images und der Verantwortungsdiffusion dienend – gearbeitet wird.

Die Dichotomie zwischen der geringen Erwähnungsdichte des Themenfeldes in einschlägigen akademischen Publikationen auf der einen und den erkennbaren Bemühungen insbesondere großer Unternehmen zur Absicherung des öffentlichen Vertrauens auf der anderen Seite, kann als Hinweis darauf gewertet werden, dass die Problematik zwar in der Praxis erkannt wurde, in der akademischen Ausbildung des IT-Nachwuchses von morgen aber noch nicht in gleichem Ausmaß präsent ist. Auf die sich hieraus ergebenden Lösungsansätze wird in 3.3.1 eingegangen.

Zusammenfassend lässt sich feststellen, dass die – schon vor Jahrzehnten diskutierte – Frage, wie und durch wen ethische Erwägungen Eingang in die Softwareentwicklung finden und wie sich daraus ergebende ethische Anforderungen an das fertige Produkt implementiert, überprüft und im Rahmen der weiteren Nutzung nachgehalten werden, im Zusammenhang mit der wachsenden Zahl von KI-Anwendungen beständig an gesellschaftlicher Relevanz zunimmt. In diesem Kontext ist die Suche nach neuen Ansätzen des ethisch vertretbaren Umgangs mit Chancen und Risiken von KI als „emerging topic“ oder „emerging field“ des Software Engineerings zu betrachten.

### 3.2 Künstliche Intelligenz

Bevor nachfolgend auf gängige Vorgehensweisen zur Lösung ethischer Dilemmata eingegangen wird, soll kurz dargelegt werden, warum Anwendungen mit Künstlicher Intelligenz<sup>2</sup>, wie sie im Weiteren immer wieder referenziert werden, besondere Herausforderungen für die ethisch verantwortliche Entwicklung von Software präsentieren. Hierfür lassen sich primär fünf Ursachen ausmachen:

- Da es sich bei KI-Anwendungen um „lernende Software“ handelt, die sich permanent an immer wieder neue Inputs anpasst, kann sie im Gegensatz zu klassischer Software mit dem Release nicht als fertiggestellt gelten. Vielmehr ist es erforderlich, ihre Entwicklung im Einsatz auch aus ethischer Perspektive weiterhin zu beobachten und zu evaluieren. Grundsätzlich kann nicht ausgeschlossen werden, dass ethisch zunächst korrekt handelnde KI erst nach Jahren des Einsatzes problematische Entscheidungsmuster entwickelt [Beckert 2021]. Bei Anwendungen, die Big Data und KI kombinieren, steht zudem zu befürchten, dass die Wirkmächtigkeit eines Systems überhaupt erst im praktischen Einsatz erkennbar wird: „The actual impacts of the use of technology tend to reveal themselves only in a stage of entrenchment in society where it has become very hard to adapt the design.“ [Dechesne 2020, S. 2]
- Sogenannte schwache KI, die anhand von historischen oder fiktiven Trainingsdaten lernt, schreibt einen in diesen Daten vorhandenen Bias – etwa eine frühere Diskriminierung von Menschen nach Herkunft, Aussehen oder Wohnsitz – vermutlich schlicht fort. [Krüger 2021, Zwingmann und Gärtner 2021]. Bei der Entwicklung bzw. beim Training solcher Systeme spielen daher Auswahl und Qualität von Trainingsdaten eine entscheidende Rolle und sind somit bei ethischen Erwägungen entsprechend mitzuberücksichtigen.
- Wirkt sich ein solcher Bias auf die Entscheidungen einer KI aus, lässt sich selbst eine als solche erkannte Verzerrung nicht immer unmittelbar korrigieren, da Systeme „in vielen Fällen [...] in kritischen Prozessen und mit enormen Datenmengen arbeiten, die nicht ohne Weiteres und auch nicht von den Personen, die die Abweichung bemerken, auf die Schnelle angepasst werden können.“ [Krüger 2021, S. 374] Hinzu kommt, dass „bei Deep Learning-Modellen, die mit vielen Schichten arbeiten, nicht mehr nachvollzogen werden [kann], an welcher der unzähligen Abzweigungen im neuronalen Netz welcher Entscheidungspfad beschritten wurde.“ [ebenda, S. 377] Die Entscheidungen der „Black Box“ KI können somit unter Umständen selbst durch die Entwickler des Systems nicht aufgeklärt werden. Die Forderung nach Entscheidungstransparenz und -nachvollziehbarkeit gehört daher zu den zentralen ethischen Anforderungen an Künstliche Intelligenz (siehe 3.3.4).

---

<sup>2</sup> Der KI-Begriff fasst in diesem Kontext nach [Radomski 2020] „verschiedene Methoden und Technologien zusammen, die kognitive Prozesse mit Computersystemen nachbilden. Dazu zählen unter anderem Fuzzy Logic, Künstliche Neuronale Netze, Expertensysteme, Maschinelles Lernen, Deep Learning, heuristische Suchen, Clustering und Klassifizierung, Mustererkennung sowie Ontologien und logisches Schließen.“ (S. 151) Dies umfasst hypothetische zukünftige Anwendungen sogenannter starker KI, die Lösungen für zuvor unbekannte Probleme finden kann, wie auch die schon weit verbreiteten Anwendungen schwacher KI, die mit Daten auf die Lösung spezifischer Aufgaben trainiert wird [Zwingmann und Gärtner 2021]. Die wichtigsten aktuellen Einsatzgebiete schwacher KI sind digitale Sprachassistenzen auf Smartphones sowie Sprachübersetzungssoftware.

- Probleme, die bei KI-Entscheidungen auftreten, betreffen oftmals nicht nur die direkten Anwenderinnen und Anwender der Systeme, sondern einen weit darüberhinausgehenden Personenkreis, der dem Einsatz der KI gar nicht ausdrücklich zugestimmt hat oder ggf. nicht einmal davon weiß. [Guizzardi et al. 2020] verweisen in diesem Zusammenhang beispielhaft auf andere Verkehrsteilnehmerinnen und -teilnehmer, die durch die Entscheidungen eines autonom navigierenden Fahrzeugs unmittelbar (und ggf. gravierend) mitbetroffen sein können.
- Schlussendlich ist festzustellen, dass sich die Leistungsfähigkeit von KI etwa alle zwei Jahre verdoppelt (Huang's Law, vgl. [Strasser und Niedermayer 2021]), weshalb sowohl die Verfügbarkeit als auch die Breite möglicher Einsatzszenarien in den kommenden Jahren rasch zunehmen werden. Die Branche ist vor diesem Hintergrund gefordert, besonders zeitnah Lösungen für die Vielzahl sich abzeichnender ethischer Fragestellungen zu entwickeln.

Aufgrund dieser Besonderheiten herrscht weitgehender Konsens darüber, dass KI-basierende Systeme „gesellschaftliche Implikation haben, die über herkömmliche Software-basierte Systeme hinausgehen“ [Beckert 2021, S. 17], weshalb die Entwicklung von KI-Anwendungen neue ethische Prinzipien erforderlich macht [Weyns 2020]. Da KI kognitiv nicht eigenständig ethisch handeln kann [Zuber et al. 2021, Zwingmann und Gärtner 2021], die Entscheidungen aber dennoch nicht mehr vom Nutzer und auch nicht mehr durch die Entwickler getroffen werden, verlagert sich die Ebene der ethischen Betrachtung in die Konzeptionsphase der Systeme [Järvinen 2017]. Eine Minderheitenmeinung vertreten [Vakkuri et al. 2020b] und [Pretschner et al. 2021], die so große Schnittmengen zwischen klassischer Softwareentwicklung und KI-Entwicklung sehen, dass sie separate ethische Leitlinien für verzichtbar halten.

Für Diskussionen über den ethisch vertretbaren Einsatz von KI auf nationaler Ebene von herausragender Bedeutung sind die 2019 durch die unabhängige High Level Expert Group on AI der Europäischen Kommission verabschiedeten Leitlinien, in denen sieben zentrale Anforderungen skizziert werden: (1) Vorrang menschlichen Handelns und menschlicher Kontrolle, (2) technische Stabilität und Sicherheit, (3) Achtung von Privatsphäre und Datenschutz, (4) (Entscheidungs-)Transparenz und Erklärbarkeit, (5) Diskriminierungsfreiheit und Fairness, (6) soziale und ökologische Verträglichkeit sowie (7) Rechenschaftspflicht [Dechesne 2020, Beckert 2021, Guizzardi et al. 2020].

### 3.3 Identifizierte Lösungsansätze

Die im Rahmen des Literaturreviews identifizierten Lösungsansätze werden nachfolgend ohne Gewichtung oder sonstige Bewertung in Reihenfolge ihrer Position entlang der Entwicklungskette aufgeführt – angefangen bei der Integration ethischer Inhalte in die Ausbildung von IT-Nachwuchskräften und Selbstverpflichtungen und Selbstevaluationen von Softwareherstellern und Entwicklerteams über an verschiedenen Stellen im Entwicklungsprozess zu nutzende Tools und Frameworks bis hin zur Vollintegration ethikbezogener Prozessschritte in agile Entwicklungsverfahren.

#### 3.3.1 Bildung und Weiterbildung

In einer überraschend großen Zahl der gesichteten Veröffentlichungen wird vorgeschlagen, die Verankerung ethischer Prinzipien in der Softwareentwicklung nicht erst über die Entwicklungsverfahren selbst, sondern bereits über die Integration ethischer Inhalte in die Aus- und Weiterbildung von IT-Fachkräften und damit über die Förderung berufsethischer Prinzipien bzw. einer Professionsethik zu erreichen. Wie [Gracia-Holgado et al. 2021], [Hans et al. 2021] und [Marebane et al. 2021] am Beispiel der IT-Ausbildung in spanischen und südafrikanischen Universitäten zeigen, kommen berufsethische Inhalte in der akademischen Lehre derzeit vielfach noch zu kurz. Gleichzeitig stießen Versuche mit der Integration solcher Inhalte durch [Gracia-Holgado et al. 2021] an der Universität Salamanca sowie von

[Pretschner et al. 2021] an der Universität München auf ein großes fachliches Interesse und eine positive Resonanz seitens der Studierendenschaft.

Die frühzeitige Vermittlung solcher Inhalte insbesondere an praktischen Beispielen und im Rahmen von fachpraktischen Gruppenarbeiten und Rollenspielen wird dann auch von einer Vielzahl der gesichteten Autorinnen und Autoren empfohlen [Filipović et al. 2018, Järvinen 2017, Larson und Miller 2017, Hansen 2019, Kamthan und Shahmir 2021]. Konkrete Vorschläge zur didaktischen Umsetzung, zur zeitlichen Verankerung im Curriculum sowie zur formativen Evaluation entsprechender Lehrinhalte finden sich in [Miñano et al. 2017], [Norberto und Iordanis 2017], [Järvinen 2017] sowie [Larson und Miller 2017]. Erklärtes Ziel ist die Ausbildung ethisch sensibilisierter und im beruflichen Umfeld resilient agierender Softwareexpertinnen und -experten, die eine Umsetzung fragwürdiger Software im Zweifelsfalle aktiv verweigern und durch ihre Marktmacht und Expertenstellung entsprechenden Druck auf Arbeitgeberinnen und Arbeitgeber aufbauen. [Hansen 2019] weist darauf hin, dass ethische Fragen zur Entwicklung und zum Einsatz von Software im Sinne eines besseren öffentlichen Verständnisses ethischer Herausforderungen auch Gegenstand von Lehrveranstaltungen in Studiengängen ohne Fokus auf Software Engineering (z.B. Betriebswirtschaftslehre oder technischer Journalismus) sein sollten.

### 3.3.2 Selbstverpflichtungen

Ein Großteil aller gesichteten Veröffentlichungen greift das Thema unternehmerischer Selbstverpflichtungen zur Berücksichtigung ethischer Fragestellungen in der Softwareentwicklung auf. Wie bereits in 3.1 erwähnt wurde, hat ein erheblicher Anteil insbesondere der großen Softwarehersteller entsprechende Codes of Ethics (CoE), Codes of Conduct (CoC) oder Ethical Guidelines veröffentlicht [Whittle 2019, Pretschner et al. 2021, Hagendorff 2020, Vakkuri et al. 2021]. Dabei sind derartige Selbstverpflichtungen kein Kind der KI-Ära, sondern bereits seit den 1990ern bekannt – so wurde etwa 1997 die erste Fassung des noch bis heute relevanten CoE der US-amerikanischen Berufsfeldgesellschaften ACM (Association for Computing Machinery) und IEEE (Institute of Electrical and Electronic Engineers Computer Society) veröffentlicht [Aberkane 2018, Ozkaya 2019].

Ein wirkungsvoller CoC sollte idealerweise nicht versuchen, die an ihn gebundenen Akteure unter Androhung beruflicher Konsequenzen zu disziplinieren, sondern das Individuum durch Bereitstellung eines ethischen Rahmenwerks in die Lage versetzen, selbstverantwortlich im Sinne des CoC zu handeln. Darüber hinaus sollte er anderen Stakeholdern Leitlinien zur Beurteilung von Entscheidungen liefern und zum Image des jeweiligen Unternehmens bzw. der jeweiligen Institution sowie zur Herausbildung berufsethischer Standards bzw. einer Professionsethik beitragen (siehe 3.3.3) [Hagendorff 2020, Gogoll et al. 2021]. Bis auf [Karim et al. 2017] („Software engineering code of ethics is acknowledged as an important area of practice that has a huge impact on how software affects our daily life“, S. 290) schätzen alle 19 gesichteten Beiträge, die CoCs erwähnen, diese als weitgehend wirkungslos ein.

Kritisiert wird dabei vor allem die schlechte Umsetzbarkeit eher vage formulierter ethischer Grundprinzipien in konkrete Projektentscheidungen und damit deren mangelnde Praxistauglichkeit – gerade dort, wo zwischen verschiedenen Werten abgewogen werden muss oder unmittelbar wirtschaftliche Interessen berührt sind [Dechesne 2020, Gogoll et al. 2021, Maas 2021, Dobbe et al. 2021, Pretschner et al. 2021, Kamthan und Shahmir 2021, Weber 2019]. „Wie es typisch für Debatten auf Basis ethischer Prinzipien ist, einigt man sich sehr schnell auf allgemeine abstrakte Aussagen. Sobald allerdings die Sachverhalte im Detail beleuchtet werden und verschiedene Sichtweisen auf die Themen eingenommen werden, zeigt sich der Diskussions- und Klärungsbedarf.“ [Hansen 2019, S. 18]

Hinzu kommen die mangelnde Verbindlichkeit vieler CoC, in denen meist keine Aussagen zu möglichen Konsequenzen von Fehlhandlungen getroffen werden [Vakkuri et al. 2021], die vielfach fehlende Priorisierung von als gleichrangig behandelten, teils aber im Konflikt zueinanderstehenden Zielvorgaben [Gogoll et al. 2021, Pretschner et al. 2021] sowie die Schwierigkeit, dass – etwa im Kontext der Entwicklung von KI-Anwendungen – vollkommen neu auftretende und zuvor nicht bedachte Fragestellungen von existierenden CoC aufgrund der oft rasanten Entwicklungen im IT-Bereich nicht immer inhaltlich sinnvoll abgedeckt werden [Gogoll et al. 2021, Beckert 2021].

Im Zuge einer umfangreichen empirischen Verhaltensstudie mit 63 Softwareentwicklerinnen und -entwicklern durch [McNamara et al. 2018] zeigte sich dann auch, dass die ausdrückliche Aufforderung, ein allgemein formuliertes CoC zu berücksichtigen, im Vergleich mit einer ohne CoC ausgestatteten Kontrollgruppe nicht zu anderen Entscheidungen führte: “Explicitly instructing participants to consider [a CoC] in their decision making had no observed effect when compared with a control group.” (S. 729) Auch [Vakkuri et al. 2021] gelangten bei einem Survey aktueller KI-Projekte zu dem Schluss, dass die in CoC sowie in der aktuellen wissenschaftlichen Literatur zur Softwareethik diskutierten Prinzipien in der Praxis keine Rolle zu spielen scheinen. Autoren wie [Gogoll et al. 2021] empfehlen vor diesem Hintergrund die Abkehr von allgemeinen CoC zugunsten der direkten Integration ethikzentrierter Diskussions- und Entscheidungsfindungsprozesse in – insbesondere agile – Entwicklungsframeworks.

### 3.3.3 Berufsethische Standards

Eine über die Formulierung von Selbstverpflichtungen für einzelne Unternehmen oder Organisationen deutlich hinausgehende Frage ist die der Notwendigkeit von berufsethischen Standards für die Softwareentwicklung analog etwa zum Hippokratischen Eid in der Medizin oder dem Pressekodex im Journalismus. Erste Ansätze hierfür stammen aus der Science-Fiction-Literatur der Mitte des 20. Jahrhunderts (z.B. die drei Gesetze der Robotik von Isaac Asimov) und ziehen sich bis zu aktuellen Versuchen, insbesondere allgemeine Prinzipien etwa zum Umgang mit KI zu formulieren (z.B. Digital Manifest von 2015, Asilomar AI-Prinzipien von 2017 oder Hippokratischer Eid für AI-Praktiker von 2018) [Filipović et al. 2018]. All diesen Anstrengungen ist jedoch gemein, dass sie aufgrund fehlender Verbindlichkeit bzw. nichtexistenter Beschwerde- und Sanktionsmöglichkeiten sowie der Heterogenität der Zielgruppe (im Gegensatz zu Medizinerinnen und Medizinern durchlaufen Personen, die Handy-Apps entwickeln, vollkommen unterschiedliche Entwicklungs- und (Selbst-)Qualifikationspfade) bislang kaum eine erkennbare Regelungswirkung entfalten. Gerade in hochdynamischen und Start-Up-geprägten Branchenfeldern der IT ist das Bild vom autodidaktischen Ein-Mann-Garagenunternehmen hochpräsent und daher zu erwarten, dass Berufsethiken, die während eines IT-Studiums vermittelt oder in Berufsverbänden oder Großkonzernen entwickelt und gelebt werden, deutlich weniger Strahlkraft entwickeln.

### 3.3.4 (Selbst-)Evaluation

Ein weiterer nennenswerter Ansatz liegt in der Selbstermächtigung von IT-Professionals, an sie herangetragene Wünsche und Anforderungen eigenständig auf ethische Problemstellungen zu evaluieren, Änderungen auszuhandeln oder sich im Zweifelsfall konkreter Aufgaben zu verweigern. Um diesen Zustand erreichen zu können, benötigen einzelne Akteure sowie auch Teams das notwendige Rüstzeug, um eigene ethische Zielvorstellungen durchdringen, strukturieren, formulieren und untereinander abstimmen zu können. Mit MEESTAR (Modell zur ethischen Evaluierung sozio-technischer Arrangements, [Weber 2019]), Values Q-Sort [Winter et al. 2021], Epos (Erfassung ethischer Positionen von Forschenden, Entwickelnden und Anwendenden, [Brandenburg et al. 2021]), der Ethical Matrix oder dem frei nutzbaren Ethics Canvas<sup>3</sup> [Weber 2019] werden in den gesichteten Veröffentlichungen gleich mehrere Instrumente zur Selbstevaluation ethischer Positionen und Zielstellungen durch Einzelpersonen oder Teams vorgestellt und bewertet. Diskutiert wird auch die Möglichkeit, entsprechende Evaluationen

---

<sup>3</sup> <http://www.ethicscanvas.org>

durch externe Expertinnen und Experten vornehmen und Entwicklerteams ggf. sogar ethisch zertifizieren zu lassen [Brandenburg et al. 2018, Beckert 2021] – ein Ansatz, der durch [Vakkuri et al. 2020b] explizit abgelehnt wird: „Quality in software development cannot be outsourced and neither can ethics.“ (S. 56) [Zuber et al. 2021] plädieren bezüglich interner wie externer Evaluationen für eine angepasste Erwartungshaltung: Analog zu CoC sei davon auszugehen, dass das dabei entstehende, eher abstrakte (Selbst-)Bild „weder bei einer Werte-Priorisierung oder einer Konfliktbewältigung noch bei der Formulierung der Verhältnismäßigkeiten oder gar bei der Einbettung ethischer Deliberationszyklen in betriebswirtschaftliche Arbeitsprozesse“ eine über grundsätzliche Denk- und Diskussionsanstoße hinausgehende Hilfe bietet (S. 9).

### 3.3.5 Transparenz

Zu den am häufigsten gestellten Anforderungen an moderne Softwaresysteme gehört die Transparenz (vgl. [Gogoll et al. 2021], die diesen Wert in 73 von 84 analysierten CoCs ausfindig machen konnten). Dahinter verbirgt sich der Wunsch, bei „Anwendungsfällen mit Auswirkungen auf Menschen“ eine „Erklärbarkeit des Modell-Outputs im Einzelfall“ grundsätzlich immer zu ermöglichen [Zwingmann und Gärtner 2021, S. 115]. Gerade mit Blick auf KI-Anwendungen gilt diese Anforderung als essentiell: „The very foundation of AI ethics: If we cannot understand how the systems work, we cannot make them ethical either“ (von Turilli und Floridi in 2009, zitiert nach [Vakkuri et al. 2020]). Besonders in diesem Einsatzbereich ist Transparenz jedoch nicht einfach zu erreichen, da sie einen Grad an Verständnis der Abläufe voraussetzt, der bei selbstlernender KI nicht in jedem Anwendungsfall erreicht werden kann [Hansen 2019, Filipović et al. 2018]. Vor diesem Hintergrund etabliert sich die Konzeption von Verfahren zur Entwicklung von Explainable AI – erklärbarer KI – derzeit als eigenständiges Forschungsfeld. Mit der Rationalisierung, der Counterfactual Analysis und der Layer-wise Relevance Propagation werden in [Krüger 2021] drei solcher Verfahren exemplarisch vorgestellt.

### 3.3.6 Ethical User Stories

Der Umgang mit User Stories oder auch Use Cases ist bereits seit den 1990er Jahren (Extreme Programming) fester Bestandteil so gut wie aller agilen Methoden der Softwareentwicklung (siehe 3.3.9). Im Kern handelt es sich um eine kurze narrative Beschreibung der Interaktion mit einem zu entwickelnden Softwaresystem aus Sicht einer Anwenderin bzw. eines Anwenders. Dabei wird in knappen Worten dargestellt, wer mit dem System etwas tut, was getan wird und welcher Nutzen dadurch entsteht [Kamthan und Shahmir 2021]. Grundsätzlich können User Stories dabei helfen, abstrakte Funktionsentwürfe am Beispiel zu konkretisieren und damit Diskussions- und Klärungsbedarf auch in ethischen Fragen zu identifizieren [Hansen 2019]. Es besteht allerdings das große Risiko, durch die voreingenommene oder ungenügend sensitive Formulierung und Priorisierung von User Stories durch Entwicklerinnen und Entwickler einen Bias in den Entwicklungsprozess zu tragen, der zu ethisch fragwürdigen Entscheidungen führt [Zwingmann und Gärtner 2021, Zuber et al. 2021].

Vor diesem Hintergrund finden sich in der betrachteten Literatur mehrere Vorschläge zur ethischen Absicherung von User Stories. Diese kann beispielsweise dadurch erreicht werden, dass die jede User Story obligatorisch begleitenden Akzeptanzkriterien um eine mandatorische ethische Dimension erweitert oder sogar spezielle „Ethicality Stories“ verfasst werden [Kamthan und Shahmir 2021]. Mit ECCOLA existiert ein in der Praxis bereits erprobtes Instrument zur Entwicklung von stärker wertezentrierten und an menschlichen Wünschen und Bedürfnissen jenseits der technischen Erwartungen an ein Produkt ausgerichteten User Stories, das in [Vakkuri et al. 2021] und insbesondere in [Halme et al. 2021] detailliert beschrieben wird. Dabei handelt es sich um ein mit dem Scrum Poker vergleichbares Kartenspiel mit 21 Karten, die mögliche Zielvorgaben wie menschliche Selbstbestimmtheit, Transparenz, Erklärbarkeit, Datenschutz oder ökologische Nachhaltigkeit repräsentieren und die flexibel auf

Sprints und Reviews verteilt werden können. Das Verfahren eignet sich insbesondere zur Dokumentation von Entscheidungen mit ethischen Dimensionen im Sinne einer späteren Nachvollziehbarkeit. [Halme et al. 2021] beschreiben das Spiel zusammenfassend als “in Agile methodology, a continuum for ethical building blocks in form of an analysis tool” (S. 40).

### 3.3.7 Requirements Engineering

Als Requirements Engineering wird – kurz zusammengefasst, die in der Praxis verfügbaren Definitionen weichen in Details voneinander ab – das systematische Vorgehen bei der Erarbeitung von Spezifikationen und Anforderungen an ein Softwareprodukt unter Einbeziehungen aller betroffenen Stakeholder bezeichnet, mithin also die Anforderungsdefinition, die Anforderungsanalyse und das Anforderungsmanagement [Guizzardi et al. 2020]. Als eine der frühesten Phasen eines Softwareprojekts ist das Requirements Engineering von entscheidender Bedeutung für die gesamte Projektausrichtung – auch mit Blick auf mögliche ethische Implikationen der zu entwickelnden Software [Dechesne 2020, Kamthan und Shahmir 2021]. Autoren wie [Jarvinen 2017] und [Guizzardi et al. 2020] plädieren deshalb für eine Integration ethikorientierter Deliberationsprozesse in diese Projektphase – eine Integration, die durch überschaubare Adaptionen existierender Instrumente des Requirements Engineering abgebildet werden kann [Whittle 2019]. Aufgrund der in 3.2 kurz angerissenen Spezifika und besonderen Potentiale von KI-Projekten mahnen jedoch [Ahmad et al. 2021] diesbezügliche Limitierungen des klassischen Requirements Engineering an. Auch [Aberkane 2018] hält den gegenwärtigen Erkenntnisstand zur Integration ethischer Fragestellungen in Methoden des Requirements Engineering grundsätzlich noch für nicht ausreichend („[W]ithin the domain of requirements engineering, there is a lack of scientific literature that addresses ethical awareness or that identifies ethical issues.“, S. 10).

### 3.3.8 Ethics Penetration Testing

Die Grundidee der Ethics Penetration Tests leitet sich aus den lang etablierten Penetrationstests im Bereich der IT-Sicherheit ab, in deren Rahmen Angriffe von außen auf ein System durch entsprechend kundige Expertinnen und Experten simuliert werden. In Umkehr der bisher angerissenen Verfahren sind hier also nicht die Entwicklerinnen und Entwickler aufgefordert, Systeme unter ethischen Gesichtspunkten zu designen, vielmehr werden Externe hinzugezogen, um Ideen zum gezielten Missbrauch von Systemen zu entwickeln und auf diesem Wege Schwachstellen aufzudecken und iterativ zu schließen. [Behrendt 2020], der dieses Verfahren vorschlägt, schreibt dazu: “The point is not to pass all ethics pen tests (this is impossible), the point is to get better through trying” (S. 4-5).

### 3.3.9 Agile Softwareentwicklung

Unter den Autorinnen und Autoren der ausgewerteten Publikationen besteht weitgehende Einigkeit darüber, dass agile Verfahren der Softwareentwicklung wie etwa Scrum in besonderer Weise dazu geeignet sind, die Berücksichtigung ethischer Anforderungen während des Entwicklungsprozesses zu fördern. Ursächlich hierfür ist vor allem, dass agile Verfahren den ständigen und offenen Austausch zwischen den Teammitgliedern – auch über normative Fragen – unterstützen, die flexible Umsetzung von Änderungen und Herangehensweisen von Sprint zu Sprint gestatten und grundsätzlich zur fachlichen Ermächtigung der Entwicklerinnen und Entwickler gegenüber übergeordneten Instanzen (im Vergleich mit hierarchischen Modellen) beitragen [Gogoll et al. 2021, Pretschner et al. 2021, Rogerson 2017]. Durch die entsprechende Anpassung agiler Verfahren wird es somit möglich, die „ethische Deliberation [zum] integralen Bestandteil der Produktentwicklung“ zu machen [Zuber et al. 2021, S. 11].

Zur Frage, in welcher Form dies idealerweise geschehen sollte, finden sich in den gesichteten Veröffentlichungen unterschiedliche Ansätze. So schlagen etwa [Whittle 2019] und [Pretschner et al. 2021] die Einführung neuer Scrum-Rollen („Values Guardian“, „Embedded Ethicist“) vor, die wahlweise durch

externe Expertinnen und Experten oder durch entsprechend (weiter-)qualifizierte Mitglieder des Kern-Scrumteams ausgefüllt werden könnten. Andere Ansätze empfehlen die Ergänzung der gängigen User Stories durch Value Stories (siehe 3.3.5) u.a. durch Nutzung von Kartenspielen wie ECCOLA oder den Einsatz von Frameworks wie EDAP (Ethical Deliberation in Agile Software Processes), die u.a. die Agile-konforme Konkretisierung normativer Ziele aus CoC in praktische Anforderungen sowie die Selbstreflexion des Entwicklerteams über ethische Entscheidungen unterstützen [Gogoll et al. 2021]. Auch andere Ethik-zentrierte SE-Frameworks wie HCAI (Hard Choices in Artificial Intelligence, siehe [Dobbe et al. 2021]) oder VDH (Value Design Hub, siehe [Hussain et al. 2018]) sind explizit Agil-geeignet.

Bezeichnenderweise kommt eine empirische Untersuchung der Integration ethischer Komponenten in agile Methoden von [Abdulhalim et al. 2018] zu dem Ergebnis, dass sich ein statistisch signifikanter Zusammenhang zwischen der Inklusion ethischer Tools in agilen Methoden und den drei Qualitätsparametern Zeitplaneinhaltung, Produktfunktionalität und Kosten nachweisen lässt. Dies verleiht entsprechenden Tools aus der Managementperspektive nicht nur mit Blick auf die Vermeidung von Ethik-Skandalen und die Sicherung der Mitarbeiterzufriedenheit, sondern auch aus rein wirtschaftlicher Sicht eine erhebliche Relevanz.

## 4 Themenrelevante Forschungsvorhaben

Folgende laufende Forschungsvorhaben mit besonderer Relevanz (ausschließlich im deutschsprachigen Raum) konnten im Rahmen der durchgeführten Literaturrecherche identifiziert werden. Für eine umfassendere Auflistung von Projekten zur KI-Ethik sei auf [Beckert 2021] verwiesen.

- Das Projekt „Ethik und Softwareentwicklung“<sup>4</sup>, ein Forschungsvorhaben des bidt (Bayerisches Forschungsinstitut für Digitale Transformation) der Ludwig-Maximilians-Universität München und der Technischen Universität München. Projektziel ist die Entwicklung eines möglichst universell einsetzbaren Schemas, das eine normativ wünschenswerte Ausgestaltung von Softwaresystemen ermöglichen soll. Das Projekt entwickelte u.a. bereits das EDAP-Schema zur Integration ethischer Entscheidungsfindungskomponenten in Scrum.
- Das Projekt „Ethik der Algorithmen“<sup>5</sup> der Bertelsmann-Stiftung. Projektziel ist hier die Sensibilisierung sowohl der Gesellschaft als auch direkter Stakeholder für Fragen der ethisch wünschenswerten Ausgestaltung von Softwaresystemen durch Veranstaltungen, die Förderung von Kleinvorhaben und Diskursen sowie die Veröffentlichung einer Open Access-Schriftenreihe (seit 2017) mit themenrelevanten Beiträgen.
- Das Projekt „ExamAI“<sup>6</sup> der Gesellschaft für Informatik e.V. Im Rahmen dieses Projekts forscht ein interdisziplinäres Team an mehreren Hochschulen zu Kontroll- und Testverfahren für KI anhand der beiden konkreten Anwendungsfälle eines Systems für das Personalmanagement sowie eines Systems zur Mensch-Maschine-Kooperation in der industriellen Fertigung.

---

<sup>4</sup> <https://www.bidt.digital/ethikundsoftwareentwicklung/>

<sup>5</sup> <https://www.bertelsmann-stiftung.de/de/unsere-projekte/ethik-der-algorithmen>

<sup>6</sup> <https://testing-ai.gi.de/>

- Das Projekt „GOAL“<sup>7</sup> (Governance von und durch Algorithmen) der Westfälischen Wilhelms-Universität Münster und des Karlsruher Instituts für Technikfolgenabschätzung und Systemanalyse (ITAS). Dieses Vorhaben beschäftigt sich mit der Frage, wie die Governance von risikobehafteten Algorithmen gestaltet und wie Algorithmen selbst in Governancefunktionen eingebunden werden können. Im Detail greift dieses Vorhaben viele auch in diesem Kurzreview angerissene Fragestellungen wie etwa das „Right to Explantion“, den Umgang mit Wertpluralismus und Wertekonflikten sowie die Identifikation von Biases in KI-Entscheidungsprozessen auf.

## 5 Mögliche Forschungsfragen

### 5.1 Inhalte und Wirksamkeit von Selbstverpflichtungen in KMU

Mit Blick auf den oben dargelegten internationalen Stand der Forschung sowie auf die Ausrichtung aktueller themenrelevanter Forschungsvorhaben auf nationaler Ebene, konnten insbesondere die in diesem sowie in den drei nachfolgenden Textabschnitten kurz vorgestellten vier Fragekomplexe als gegenwärtig noch bestehende Erkenntnislücken identifiziert werden.

Wie in 3.3.2 dargelegt, existiert in der Praxis bereits eine Vielzahl von Selbstverpflichtungen einzelner Unternehmen sowie berufsethischer Leitlinien zur Berücksichtigung ethischer Fragestellungen in der Softwareentwicklung. Wie [Zuber et al. 2021] feststellen, fehlt eine systematische Auswertung dieser permanent wachsenden Zahl von Leitlinien bislang allerdings vollkommen. Nach [Beckert 2021] oder [Vakkuri, Kemell et al. 2020] sind Leitlinien zudem fast ausschließlich für große und international tätige Unternehmen untersucht (siehe z.B. die Evaluation von 22 CoC durch [Hagendorff 2020], die diese aus Transparenzgründen publizieren bzw. sich an entsprechenden Fachforen beteiligen. Weit weniger klar ist das Bild bei der Vielzahl von KMU und insbesondere kurzlebigen Start-Ups, die in Marktbereichen wie Mobile Apps oder Social Media-Anwendungen ja aber durchaus erhebliche Innovationsschübe auslösen und aufgrund der Skalierungseffekte gerade bei KI-Anwendungen schnell rasant wachsen oder marktdisruptiv wirken können [Zwingmann und Gärtner 2021]. Gerade für Start-Ups ist – etwa durch [Vakkuri et al. 2020a] – belegt, dass ethische Erwägungen bei der Konzeption neuer KI-Anwendungen häufig deutlich hinter marktwirtschaftlichen Überlegungen zurücktreten. Tatsächlich unterstreicht das Beispiel der diskriminierenden Personalauswahl-KI des bayerischen Start-Ups Retorio, dass das gerade für junge Unternehmen attraktive Ziel, mit einer neuen Anwendung als erstes am Markt zu sein („First Mover“), einer gründlichen Prüfung ethischer Implikationen erheblich im Weg stehen kann – dies für den Vertrieb des Produkts aber auch keinesfalls nachteilig sein muss [Beckert 2021].

Konkret untersucht werden könnten in diesem Zusammenhang insbesondere drei Aspekte: Viele Autorinnen und Autoren (z.B. [Hansen 2019], [Dechesne 2020] und [McNamara et. al. 2018]) mutmaßen oder belegen am Beispiel einzelner Selbstverpflichtungen, dass diese oft in abstrakten Überlegungen verhaftet bleiben und keine Möglichkeiten zur Sanktionierung bei Verstößen beinhalten, weshalb sie auf die konkrete Umsetzung von Softwareprojekten kaum Wirkung entfalten. Darüber hinaus merkt z.B. [Beckert 2021] an, dass viele Selbstverpflichtungen noch nicht mit Blick auf die sich neu öffnende Problemdimension autonom entscheidender KI entstanden sind und in diesem Anwendungsbereich daher als überaltet zu gelten haben. Alle drei Aspekte (Praxisnähe, Sanktionsmechanismen und KI-Bezug) ließen sich mit einer allgemeinen deskriptiven Auswertung zu Selbstverpflichtungen und internen Richtlinien verbinden (Umfang, Geltungsbereich, Erarbeitung im Haus oder mit einem externen Partner etc. pp.). Eine derartige Übersicht für deutsche KMU (quantitative Erfassung wie auch qualitative Auswertung) würde eine klar identifizier- und definierbare Erkenntnislücke schließen.

---

<sup>7</sup> <https://goal-projekt.de/>

## 5.2 Umgang mit interkulturellen Differenzen

Ein bislang noch weitgehend wenig untersuchtes Phänomen ist der Einfluss divergierender ethischer Vorstellungen sowohl auf Softwareprojekte mit internationaler Ausrichtung als auch auf international arbeitende Entwicklerteams. Da für Konzepte wie Gerechtigkeit, Gleichheit oder Fairness global gesehen (sowie ja auch schon mit Blick auf nur eine einzelne Gesellschaft) keine „universalen, klar umrissenen Grenzen“ [Krüger 2021, S. 379] existieren sowie auch deren relative Bedeutung schwankt, können sich in kulturdiversen Kontexten Missverständnisse und Fehlinterpretationen ergeben, die für die Stakeholder nicht unmittelbar ersichtlich sind [Winkler 2018, Alsanoosy et al. 2018]. So weist etwa [Berendt 2020] darauf hin, dass viele Selbstverpflichtungen zum Umgang mit KI darauf drängen, dass diese dem „Allgemeinwohl“ dienen soll – gerade dieser Begriff aber auch kulturell bedingt ganz unterschiedlich definiert wird. Konkrete Beispiele für solche Konflikte finden sich bei [Massago et al. 2018], mit dem erheblichen Einfluss kultureller Unterschiede auf das Requirements Engineering setzen sich [Alsanoosey et al. 2018] auseinander und belegen diesen unter anderem anhand qualitativer Interviews mit Softwareentwicklerinnen und -entwicklern in Australien und Saudi-Arabien.

In der gesichteten Literatur werden im Kontext dieser Problematik verschiedene relevante Forschungsfragen angesprochen: Wie können agile Methoden adaptiert oder ergänzt werden, um kulturell diverse ethische Anforderungen seitens der späteren Nutzerinnen und Nutzer oder seitens der Mitglieder eines Entwicklerteams abzubilden? Sollte es Kundinnen und Kunden aus unterschiedlichen Kulturreihen möglich sein, bei der Beauftragung von Software einen von mehreren ethischen Bezugsrahmen für die Umsetzung zu wählen [Järvinen 2017]? Sollten Entwicklungsteams nicht nur disziplinär [Beckert 2021], sondern auch kulturell diversifizierter sein, um möglicherweise extern zu erwartende Konflikte frühzeitig internalisieren und lösen zu können? Auch dieses Themenfeld ist reich an potentiell interessanten Fragestellungen und eignet sich von den hier aufgeführten vermutlich am besten für ein Projekt mit internationaler Ausrichtung und jenseits der Zielgruppe deutscher KMU und Start-Ups.

## 5.3 Problem der „fehlenden Werte“

Ein definitorisches Problem, das im Rahmen der betrachteten Literatur häufig angesprochen wird, liegt in der mangelnden Grenzziehung zwischen den Begriffen „Ethik“ und „Werte“. Zwar hängen beide Begriffe sehr eng zusammen, sind jedoch nicht als synonym zu betrachten. Bei menschlichen Werten handelt es sich – in unzureichender definitorischer Tiefe – um oft tief verwurzelte Überzeugungen oder Ideale, die der Unterscheidung zwischen als gut und schlecht empfundenen Prozessergebnissen dienen und die damit Implikationen haben können, die im allgemeinen Sprachgebrauch als ethisch relevant betrachtet werden würden, die derartige Implikationen aber auch nicht haben müssen [Winter et al. 2019]. Beispielhaft sei hier auf die Werte „Fairness“ und „Gerechtigkeit“ sowie „Harmonie“ und „Treue“ verwiesen. Während die ersten beiden Werte mit Blick etwa auf die Arbeitsweise eines KI-gestützten Personalauswahlsystems unmittelbare Relevanz entfalten, ist dies bei den zwei nachfolgend genannten Werten zumindest oberflächlich betrachtet nicht der Fall.

Diese zunächst irrelevant erscheinende Unterscheidung weist auf eine nicht unerhebliche Erkenntnislücke hin, die insbesondere von [Winter et al. 2018 und 2019] angesprochen wird: Die menschliche Existenz wird von einer Vielzahl von Werten (je nach kulturellem Kontext einigen hundert) beeinflusst, von denen scheinbar nur einige (wie eben die angesprochene Fairness) Eingang in Überlegungen zur Entwicklung von Software finden. In welchem Umfang und auf welchem Wege aber finden andere menschliche Werte Eingang in solche Prozesse? „While the influential field of Value Sensitive Design, for example, focuses predominantly on values with ‚ethical impact‘ [...], to not consider ‚nonethical‘ values seriously fails to account for a significant part of the picture of [Software Engineering], and the complex decision-making processes and values trade-offs that may result.“ [Winter et al. 2019, S. 20]

Insbesondere [Winter et al. 2018] sehen hier ein wesentliches Wissensdefizit – und auch ein emergentes Forschungsfeld: „It seems that little research on the measurement of human values in Software Engineering exists. [...] This paper makes the case for the study of human values in Software Engineering as a highly important emerging area of research with significant societal implications.“ (S. 2/19)

#### 5.4 Identifikation von Entscheidungsebenen

Auch wenn einige der unterschiedlichen Ebenen, auf denen im Rahmen eines Softwareprojekts ethische Entscheidungen getroffen werden können, im Rahmen der gesichteten Literatur und damit auch im Rahmen dieses Kurzreviews bereits Erwähnung fanden (z.B. Management, Auftraggeber, spätere Nutzer), lag doch der Fokus aller bisherigen Ausführungen auf der ethikbezogenen Ausbildung, Befähigung und Ermächtigung der Entwicklerinnen und Entwickler von Software. Einige der betrachteten Publikationen werfen dennoch die Frage auf, ob die primäre Verantwortungszuschreibung an die Ebene der technischen Expertinnen und Experten zu bejahen ist. [Richter 2020] bezeichnet dies als das „Problem der vielen Hände“ – eine Phrase, mit der ausgedrückt werden soll, dass „die Verantwortungszuschreibung in der Softwareentwicklung durch den Umstand erschwert wird, dass nicht bestimmte, sondern oft eine Vielzahl von Akteuren am System als Endprodukt beteiligt waren.“ (S. 342)

Im Falle von KI-Anwendungen könnten Entwickler vielfach nicht einmal vorhersagen, welche Ergebnisse zu erwarten sind, bevor der praktische Einsatz nicht bereits angelaufen ist. Während [Vakkuri et al. 2019] für eine stärkere Inverantwortungnahme der Managementebene plädieren, argumentieren [Gogoll et al. 2021], diese Ebene habe „little control and influence in the development of a software product“ (S. 1087). [Järvinen 2017] weist darauf hin, dass viel neue Software auf existierenden Programmen und frei oder kommerziell verfügbaren Frameworks aufbaut und stellt die Frage, inwiefern sich (auch moralische) Verantwortung über Entwicklergenerationen hin verteilen kann – und von [Krüger 2021] kommt der Hinweis, dass bei mit historischen oder fiktiven Daten trainierten KI-Systemen denjenigen, die diese Daten auswählen, zusammenstellen, sichten und filtern, am Ende vermutlich sehr viel mehr Einfluss auf ein Entwicklungsprojekt ausüben, als die eigentlichen Entwickler.

Die oft mangelnde Trennung zwischen der Ethik von Business-Entscheidungen eines Softwarekonzerns und der Ethik von Entscheidungen im Rahmen der Softwareentwicklung [Gogoll et al. 2021] sowie die bereits in 5.3 angesprochene Vermischung von Ethik und den diese Ethik determinierenden Werten führt dazu, dass dem „Problem der vielen Hände“ eine gewisse definitorische Unschärfe anheim wohnt, die schwierig in konkrete Forschungsfragen aufzulösen ist. Davon unbenommen kommt in Teilen der gesichteten Literatur deutlich zum Ausdruck, dass die sich aus fast allen der in Abschnitt 3 angesprochenen Lösungsansätze manifestierende Zuschreibung der primären Verantwortung für die Lösung ethischer Fragestellungen an die Ebene der Entwickler durchaus kritisch zu hinterfragen ist. Auch die hieraus sprechende Erkenntnislücke bietet Raum für Projektansätze, deren weitere Konkretisierung jedoch eine vertiefende Betrachtung der Literatur zwingend voraussetzt.

### 6 Abschließende Empfehlungen

Ausgehend von den Ergebnissen des in diesem Bericht kurz zusammengefassten Literaturreviews sowie weiteren kurSORischen Recherchen zu Zielen und Forschungsfragen von gegenwärtig im gleichen Themenfeld laufenden Projekten, wird hinsichtlich der begründeten Darstellung von Erkenntnislücken empfohlen, eine Antragstellung zu einem der folgenden Themenkomplexe zu erwägen:

1. Struktureller Umgang mit ethischen Fragestellungen bei der Softwareentwicklung in deutschen KMU. Welcher Anteil an KMU / Start-Ups der deutschen Softwarelandschaft verfügt über eine veröffentlichte oder interne Selbstverpflichtung? Lässt sich eine kritische Menge

an Selbstverpflichtungen akquirieren, die z.B. auf verbindende Inhalte, Wording, Aussagen zu Sanktionen bei Verstößen, allgemeine Zielstellung etc. hin verglichen werden können? Wie stellt sich der Umgang mit ethischen Fragestellungen bei IT-Start-Ups dar, die über keine Ressourcen für die professionelle Leitbildentwicklung verfügen? Lassen sich hierzu ggf. Interviews mit Gründerinnen und Gründern aus dem Kreise der Alumni durchführen?

2. Möglichkeiten der Adaption agiler Methoden der Softwareentwicklung zur Integration kulturell bedingter Wertdifferenzen in Entscheidungsfindungsprozesse. Wie lassen sich kulturell bedingte Wertdifferenzen – sowohl mit Blick auf die Entwicklung von Software für diversifizierte internationale Zielmärkte als auch mit Blick auf die internationale Zusammensetzung von Entwicklerteams – im Rahmen agiler Frameworks abbilden? Können bestehende Verfahren um neue Rollen oder Prozessschritte ergänzt werden, um Wertunterschieden Rechnung zu tragen? Lassen sich ggf. deutsche KMU ermitteln, die eine solche Differenzierung in abgeschlossenen oder laufenden Projekten umsetzen mussten oder müssen – und falls ja, wären diese Unternehmen bzw. die jeweiligen Entwicklerteams bereit, an einer qualitativen Studie (qualitative Befragung, moderierte Gruppendiskussion, Delphi-Verfahren) teilzunehmen?

Mit Blick auf die Möglichkeiten der Beschaffung von Rohdaten innerhalb der Hochschule sollte darüber hinaus ggf. auch das folgende Thema in Betracht gezogen werden, auch wenn ein entsprechender Forschungsbedarf hier ausschließlich von [Nittala et al. 2018] attestiert wird.

3. Grundlegende Einstellung zum Umgang mit ethischen Fragestellungen unter Studierenden verschiedener Informatik-Fachrichtungen sowie unterschiedlicher Studienjahrgänge an der Hochschule Harz. Mit welchen Einstellungen und Überzeugungen zu diesem Themenkomplex starten Studierende entsprechender Studiengänge (z.B. Medieninformatik, Wirtschaftsinformatik, Verwaltungsinformatik) im ersten Semester in ihr Studium? Lässt sich ermitteln, welche Faktoren diese Einstellungen und Überzeugungen wesentlich beeinflusst haben? Wie verändern sie sich im Verlauf des Studiums – etwa vor und nach einem Praxissemester oder vor und nach einer Vorlesung mit dezidiert ethischen Inhalten? Lassen sich signifikante Unterschiede zwischen Studierenden unterschiedlicher Fachrichtungen, Semester oder zwischen Bachelor- und Masterstudiengängen feststellen? Worauf sind diese zurückführbar? Haben die Studierenden den Eindruck, im Rahmen ihres Studiums ausreichend über den Themenkomplex der Implementierung ethischer Softwareanforderungen informiert zu werden? Welche Lehr- und Lernformate würden sie sich hierzu ggf. noch wünschen?

## 7 Verwendete Quellen

Abdulhalim, Hisham; Lurie, Yotam; Mark, Shlomo (2018): Ethics as a Quality Driver in Agile Software Projects. In: *JSSM* 11 (01), S. 13–25. DOI: 10.4236/jssm.2018.111002.

Aberkane, Abdel-Jaouad (2018): Exploring Ethics in Requirements Engineering. Masterarbeit. Utrecht University, Utrecht.

Ahmad, Khlood; Bano, Muneera; Abdelrazek, Mohamed; Arora, Chetan; Grundy, John (2021): What's up with Requirements Engineering for Artificial Intelligence Systems? In: 2021 IEEE 29th International Requirements Engineering Conference (RE). 2021 IEEE 29th International Requirements Engineering Conference (RE). Notre Dame, IN, USA, 20.09.2021 - 24.09.2021: IEEE, S. 1–12.

Alsanoosy, Tawfeeq; Spichkova, Maria; Harland, James (2018): Cultural Influences on the Requirements Engineering Process: Lessons Learned from Practice. In: 2018 23rd International Conference on Engineering of Complex Computer Systems (ICECCS). 2018 23rd International Conference on Engineering of Complex Computer Systems (ICECCS). Melbourne, VIC, 12.12.2018 - 14.12.2018: IEEE, S. 61–70.

Beckert, Bernd (2021): Vertrauenswürdige künstliche Intelligenz. In: *TATuP* 30 (3), S. 17–22. DOI: 10.14512/tatup.30.3.17.

Berendt, Bettina (2020): (De)constructing ethics for autonomous cars: A case study of Ethics Pen-Testing towards “AI for the Common Good”. In: *irie* 28. DOI: 10.29173/irie381.

Brandenburg, Stefan; Schott, Ronja; Minge, Michael (2018): Zur Erfassung der ethischen Position in der Softwareentwicklung.

Cemiloglu, Deniz; Arden-Close, Emily; Hodge, Sarah; Kostoulas, Theodoros; Ali, Raian; Catania, Maris (2020): Towards Ethical Requirements for Addictive Technology: The Case of Online Gambling. In: 2020 1st Workshop on Ethics in Requirements Engineering Research and Practice (REthics). 2020 1st Workshop on Ethics in Requirements Engineering Research and Practice (REthics). Zurich, Switzerland, 31.08.2020 - 31.08.2020: IEEE, S. 1–10.

Dechesne, Francien (2020): Requirements engineering for moral considerations in algorithmic systems: RE’20 Conference Keynote. In: 2020 IEEE 28th International Requirements Engineering Conference (RE). 2020 IEEE 28th International Requirements Engineering Conference (RE). Zurich, Switzerland, 31.08.2020 - 04.09.2020: IEEE, S. 1–2.

Dobbe, Roel; Krendl Gilbert, Thomas; Mintz, Yonatan (2021): Hard choices in artificial intelligence. In: *Artificial Intelligence* 300, S. 103555. DOI: 10.1016/j.artint.2021.103555.

Filipović, Alexander; Koska, Christopher; Paganini, Claudia; Bertelsmann Stiftung (2018): Ethik für Algorithmiker. Arbeitspapier. Bertelsmann Stiftung. Gütersloh.

Gogoll, Jan; Zuber, Niina; Kacianka, Severin; Greger, Timo; Pretschner, Alexander; Nida-Rümelin, Julian (2021): Ethics in the Software Development Process: from Codes of Conduct to Ethical Deliberation. In: *Philos. Technol.* 34 (4), S. 1085–1108. DOI: 10.1007/s13347-021-00451-w.

Guizzardi, Renata; Amaral, Glenda; Guizzardi, Giancarlo; Mylopoulos, John (2020): Ethical Requirements for AI Systems. In: Cyril Goutte und Xiaodan Zhu (Hg.): *Advances in Artificial Intelligence*. Cham: Springer International Publishing (12109), S. 251–256.

Hagendorff, Thilo (2020): The Ethics of AI Ethics: An Evaluation of Guidelines. In: *Minds & Machines* 30 (1), S. 99–120. DOI: 10.1007/s11023-020-09517-8.

Halme, Erika; Vakkuri, Ville; Kultanen, Joni; Jantunen, Marianna; Kemell, Kai-Kristian; Rousi, Rebekah; Abrahamsson, Pekka (2021): How to Write Ethical User Stories? Impacts of the ECCOLA Method. In: Peggy Gregory, Casper Lassenius, Xiaofeng Wang und Philippe Kruchten (Hg.): *Agile Processes in Software Engineering and Extreme Programming*, Bd. 419. Cham: Springer International Publishing (Lecture Notes in Business Information Processing), S. 36–52.

Hans, Robert T.; Marebane, Senyeki M.; Coosner, Jacqui (2021): Computing Academics’ Perceived Level of Awareness and Exposure to Software Engineering Code of Ethics: A Case Study of a South African University of Technology. In: *IJACSA* 12 (5). DOI: 10.14569/IJACSA.2021.0120570.

Hansen, Marit (2019): Die Empfehlungen der Datenethikkommission: Bedeutung für die Informatik. In: *FlfF-Kommunikation* (04/2019), S. 17–21.

Hussain, Waqar; Mougouei, Davoud; Whittle, Jon (2018): Integrating social values into software design patterns. In: Yuriy Brun, Brittany Johnson und Alexandra Meliou (Hg.): *Proceedings of the International Workshop on Software Fairness. ICSE '18: 40th International Conference on Software Engineering*. Gothenburg Sweden, 29 05 2018 29 05 2018. New York, NY, USA: ACM, S. 8–14.

Järvinen, Hannu-Matti (2017): Ethics as a skill of software engineer? 45th SEFI Conference. Azores, 2017.

Josephina, Antoniou; Andreas, Andreou (2019): Case Study – The Internet of Things and Ethics. In: *The ORBIT Journal* 2 (2), S. 1–29. DOI: 10.29297/orbit.v2i2.111.

Kamthan, Pankaj; Shahmir, Nazlie (2021): On Ethically-Sensitive User Story Engineering. In: 2021 4th International Conference on Computer Science and Software Engineering (CSSE 2021). CSSE 2021: 2021 4th International Conference on Computer Science and Software Engineering. Singapore Singapore, 22 10 2021 24 10 2021. New York, NY, USA: ACM, S. 71–79.

Karim, Nor Shahriza Abdul; Ammar, Fahda Al; Aziz, Romana (2017): Ethical Software: Integrating Code of Ethics into Software Development Life Cycle. In: 2017 International Conference on Computer and Applications (ICCA). 2017 International Conference on Computer and Applications (ICCA). Doha, United Arab Emirates, 06.09.2017 - 07.09.2017: IEEE, S. 290–298.

Krüger, Sven (2021): KI und Verantwortung. In: Sven Krüger (Hg.): *Die KI-Entscheidung*. Wiesbaden: Springer Fachmedien Wiesbaden, S. 355–571.

Larson, David K.; Miller, Keith W. (2017): Action ethics for a software development class. In: *ACM Inroads* 8 (1), S. 38–42. DOI: 10.1145/3043949.

Maas, Jonne (2021): A Neo-republican Critique of AI Ethics. In: *Journal of Responsible Technology*, S. 100022. DOI: 10.1016/j.jrt.2021.100022.

Marebane, Senyeki M.; T., Robert (2021): Software Engineering Ethics Competency Gap in Undergraduate Computing Qualifications within South African Universities of Technology. In: *IJACSA* 12 (4). DOI: 10.14569/IJACSA.2021.0120474.

Massago, Yogi; Lapasini Leal, Gislaine Camila; Balancieri, Renato; Cardoza Galdamez, Edwin Vladimir (2018): Towards classifying sociocultural aspects in global software development. In: *JISTEM* 15, S. 1–19. DOI: 10.4301/S1807-1775201815007.

McNamara, Andrew; Smith, Justin; Murphy-Hill, Emerson (2018): Does ACM's code of ethics change ethical decision making in software development? In: Gary T. Leavens, Alessandro Garcia und Corina S. Păsăreanu (Hg.): *Proceedings of the 2018 26th ACM Joint Meeting on European Software Engineering Conference and Symposium on the Foundations of Software Engineering. ESEC/FSE '18: 26th ACM Joint European Software Engineering Conference and Symposium on the Foundations of Software Engineering*. Lake Buena Vista FL USA, 04 11 2018 09 11 2018. New York, NY, USA: ACM, S. 729–733.

Miñano, Rafael; Uruburu, Ángel; Moreno-Romero, Ana; Pérez-López, Diego (2017): Strategies for Teaching Professional Ethics to IT Engineering Degree Students and Evaluating the Result. In: *Science and engineering ethics* 23 (1), S. 263–286. DOI: 10.1007/s11948-015-9746-x.

Misselhorn, Catrin (2019): Moralische Maschinen in der Pflege? Grundlagen und eine Roadmap für ein moralisch lernfähiges Altenpflegesystem. In: Christiane Woopen und Marc Jannes (Hg.): Roboter in der Gesellschaft. Berlin, Heidelberg: Springer Berlin Heidelberg, S. 53–68.

Nittala, Swetha; Zephrin, Tasha; Howland, Shiloh; Kim, Dayoung; Katz, Andrew; Jesiek, Brent (2018): Investigating Influences on First-year Engineering Students' Views of Ethics and Social Responsibility. In: 2018 ASEE Annual Conference & Exposition Proceedings. 2018 ASEE Annual Conference & Exposition. Salt Lake City, Utah: ASEE Conferences.

Norberto, Patrignani; Iordanis, Kavathatzopoulos (2017): On the Difficult Task of Teaching Computer Ethics to Engineers. In: *The ORBIT Journal* 1 (1), S. 1–10. DOI: 10.29297/orbit.v1i1.20.

Ozkaya, Ipek (2019): Ethics Is a Software Design Concern. In: *IEEE Softw.* 36 (3), S. 4–8. DOI: 10.1109/MS.2019.2902592.

Perera, Harsha; Hussain, Waqar; Whittle, Jon; Nurwidayantoro, Arif; Mougouei, Davoud; Shams, Rifat Ara; Oliver, Gillian (2020): A study on the prevalence of human values in software engineering publications, 2015 -- 2018. In: Gregg Rothermel und Doo-Hwan Bae (Hg.): Proceedings of the ACM/IEEE 42nd International Conference on Software Engineering. ICSE '20: 42nd International Conference on Software Engineering. Seoul South Korea, 27 06 2020 19 07 2020. New York, NY, USA: ACM, S. 409–420.

Pretschner, Alexander; Zuber, Niina; Gogoll, Jan; Kacianka, Severin; Nida-Rümelin, Julian (2021): Ethik in der agilen Software-Entwicklung. In: *Informatik Spektrum* 44 (5), S. 348–354. DOI: 10.1007/s00287-021-01390-8.

Radomski, Sabine (2020): Ethikregeln für Künstliche Intelligenz – für die Revision geeignet? In: Aleksandra Sowa (Hg.): IT-Prüfung, Datenschutzaudit und Kennzahlen für die Sicherheit. Wiesbaden: Springer Fachmedien Wiesbaden, S. 151–193.

Richter, Andreas (2020): Transparenz als Mittel der Verantwortlichkeit bei KI-gestützten Systemen. In: Norbert Gronau, Moreen Heine, K. Pousscchi und H. Krasnova (Hg.): WI2020 Community Tracks: GIT Verlag, S. 341–350.

Rogerson, Simon (2017): Is professional practice at risk following the Volkswagen and Tesla motors revelations? In: *The ORBIT Journal* 1 (1), S. 1–15. DOI: 10.29297/orbit.v1i1.34.

Saurwein, Florian; Just, Natascha; Latzer, Michael (2017): Algorithmische Selektion im Internet. In: *kommges* 18 (2). DOI: 10.15460/kommges.2017.18.2.586.

Strasser, Katrin; Niedermayer, Bernhard (2021): Unvoreingenommenheit von Künstliche-Intelligenz-Systemen. Die Rolle von Datenqualität und Bias für den verantwortungsvollen Einsatz von künstlicher Intelligenz. In: Reinhard Altenburger und René Schmidpeter (Hg.): CSR und Künstliche Intelligenz. Berlin, Heidelberg: Springer Berlin Heidelberg (Management-Reihe Corporate Social Responsibility), S. 121–135.

Vakkuri, Ville; Abrahamsson, Pekka (2018): The Key Concepts of Ethics of Artificial Intelligence. In: 2018 IEEE International Conference on Engineering, Technology and Innovation (ICE/ITMC). 2018 IEEE International Conference on Engineering, Technology and Innovation (ICE/ITMC). Stuttgart, 17.06.2018 - 20.06.2018: IEEE, S. 1–6.

Vakkuri, Ville; Kemell, Kai-Kristian; Abrahamsson, Pekka (2019): Implementing Ethics in AI: Initial Results of an Industrial Multiple Case Study. In: Xavier Franch, Tomi Männistö und Silverio Martínez-

Fernández (Hg.): Product-Focused Software Process Improvement, Bd. 11915. Cham: Springer International Publishing (Lecture Notes in Computer Science), S. 331–338.

Vakkuri, Ville; Kemell, Kai-Kristian; Jantunen, Marianna; Abrahamsson, Pekka (2020a): "This is just a prototype" How ethics are ignored in software startup-like environments. In: Viktoria Stray, Rashina Hoda, Maria Paasivaara und Philippe Kruchten (Hg.): Agile Processes in Software Engineering and Extreme Programming. Cham: Springer International Publishing (383), S. 195–210.

Vakkuri, Ville; Kemell, Kai-Kristian; Jantunen, Marianna; Halme, Erika; Abrahamsson, Pekka (2021): ECCOLA — A method for implementing ethically aligned AI systems. In: *Journal of Systems and Software* 182. DOI: 10.1016/j.jss.2021.111067.

Vakkuri, Ville; Kemell, Kai-Kristian; Kultanen, Joni; Abrahamsson, Pekka (2020b): The Current State of Industrial Practice in Artificial Intelligence Ethics. In: *IEEE Softw.* 37 (4), S. 50–57. DOI: 10.1109/MS.2020.2985621.

Weber, Karsten (2019): Methoden der ethischen Evaluation von IT. In: *INFORMATIK 2019: 50 Jahre Gesellschaft für Informatik – Informatik für Gesellschaft*, S. 431–444. DOI: 10.18420/INF2019\_WS47.

Weyns, Danny (2020): Towards a code of ethics for autonomous and self-adaptive systems. In: Shinichi Honiden, Elisabetta Di Nitto und Radu Calinescu (Hg.): Proceedings of the IEEE/ACM 15th International Symposium on Software Engineering for Adaptive and Self-Managing Systems. SEAMS '20: IEEE/ACM 15th International Symposium on Software Engineering for Adaptive and Self-Managing Systems. Seoul Republic of Korea, 29 06 2020 03 07 2020. New York, NY, USA: ACM, S. 163–165.

Whittle, Jon (2019): Is Your Software Valueless? In: *IEEE Softw.* 36 (3), S. 112–115. DOI: 10.1109/MS.2019.2897397.

Whittle, Jon; Ferrario, Maria Angela; Simm, Will; Hussain, Waqar (2021): A Case for Human Values in Software Engineering. In: *IEEE Softw.* 38 (1), S. 106–113. DOI: 10.1109/MS.2019.2956701.

Winkler, Till (2018): Human Values as the Basis for Sustainable Software Development. In: 2018 IEEE International Symposium on Technology and Society (ISTAS). 2018 IEEE International Symposium on Technology and Society (ISTAS). Washington DC, DC, USA, 13.11.2018 - 14.11.2018: IEEE, S. 37–42.

Winter, Emily; Forshaw, Stephen; Hunt, Lucy; Ferrario, Maria Angela (2019): Advancing the Study of Human Values in Software Engineering. In: 2019 IEEE/ACM 12th International Workshop on Cooperative and Human Aspects of Software Engineering (CHASE). 2019 IEEE/ACM 12th International Workshop on Cooperative and Human Aspects of Software Engineering (CHASE). Montreal, QC, Canada, 27.05.2019 - 27.05.2019: IEEE, S. 19–26.

Winter, Emily; Forshaw, Steve; Ferrario, Maria Angela (2018): Measuring human values in software engineering. In: Markku Oivo, Daniel Méndez und Audris Mockus (Hg.): Proceedings of the 12th ACM/IEEE International Symposium on Empirical Software Engineering and Measurement. ESEM '18: ACM / IEEE International Symposium on Empirical Software Engineering and Measurement. Oulu Finland, 11 10 2018 12 10 2018. New York, NY, USA: ACM, S. 1–4.

Zuber, Niina; Kacianka, Severin; Pretschner, Alexander; Nida-Rümelin, Julian (2021): Ethische Deliberation für agile Softwareprozesse: EDAP-Schema. bidt. München (bidt Working Paper, 1).

Zwingmann, Tobias; Gärtner, Tobias (2021): Vergessen wir mal die Roboter: So gelingt Künstliche-Intelligenz-Ethik in der Praxis. In: Reinhard Altenburger und René Schmidpeter (Hg.): CSR und Künstliche Intelligenz. Berlin, Heidelberg: Springer Berlin Heidelberg (Management-Reihe Corporate Social Responsibility), S. 105–119.

## Kontakt

Christian Reinboth, M.Sc.

Stabsstelle Forschung

Hochschule Harz

Friedrichstr. 57-59

38855 Wernigerode

Haus 6 | Raum 6.110a

Tel: 03943 659 896

Mobil: 0152 0900 6600

E-Mail: [creinboth@hs-harz.de](mailto:creinboth@hs-harz.de)