



Künstliche Intelligenz
für Arbeit und Lernen

Notwendigkeit der Integration von ethischen, rechtlichen und sozialen Aspekten in die gängigen Vorgehensmodelle für IT-Projekte

Sascha Alpers¹

Der Artikel ist ein Preprint von: Alpers, Sascha (2022). Notwendigkeit der Integration von ethischen, rechtlichen und sozialen Aspekten in die gängigen Vorgehensmodelle für IT-Projekte. In Fazal-Baqae et al. (Hrsg.), Projektmanagement und Vorgehensmodelle, S. 171-180, Reihe: Oberweis et al.: Lecture Notes in Informatics(P-327), Bonn.

Abstract: Für die Steuerung von Softwareentwicklungsprojekten gibt es eine Vielzahl verschiedener Vorgehensmodelle. Diese betrachten auch die Erhebung und die Verwaltung von Anforderungen – mit ganz unterschiedlichen agilen beziehungsweise klassischen Ansätzen. Der insbesondere durch ethische Aspekte, rechtliche Randbedingungen und Fragen der sozialen Technikgestaltung (ELSA bzw. ELSI) vorgegebene Rahmen wird in den Vorgehensmodellen nicht explizit adressiert. Dafür gibt es eigene Ansätze wie beispielsweise den IEEE Standard Model Process for Addressing Ethical Concerns during System Design (IEEE7000-2021). Durch die fehlende explizite Integration dieser Fragen in gängige Vorgehensmodelle wie bspw. Scrum oder das V-Modell-XT ist jedoch nicht sichergestellt, dass in den Entwicklungsprojekten der notwendige Raum zur Reflexion von ELSA eröffnet wird. Der Beitrag diskutiert die Problemstellung und zeigt mögliche Lösungsoptionen für den weiteren Diskurs auch im Rahmen des Workshops auf.

Keywords: Vorgehensmodell; Ethik; Recht; soziale Technikgestaltung

¹ FZI Forschungszentrum Informatik, Forschungsbereich Software Engineering, Haid-und-Neu-Straße 10-14, 76131 Karlsruhe, alpers@fzi.de

Dieses Forschungs- und Entwicklungsprojekt wird durch das Bundesministerium für Bildung und Forschung (BMBF) im Programm „Zukunft der Wertschöpfung – Forschung zu Produktion, Dienstleistung und Arbeit“ (Förderkennzeichen: 02L19C250) gefördert und vom Projektträger Karlsruhe (PTKA) betreut. Die Verantwortung für den Inhalt dieser Veröffentlichung liegt bei der Autorin / beim Autor.

GEFÖRDERT VOM



Bundesministerium
für Bildung
und Forschung



1 Einleitung

Für die Steuerung von Softwareentwicklungsprojekten gibt es eine Vielzahl verschiedener Vorgehensmodelle wie bspw. das V-Modell XT [An06] oder Scrum [SS20]. Die Vorgehensmodelle betrachten auch die Erhebung und die Verwaltung von Anforderungen – mit ganz unterschiedlichen agilen beziehungsweise klassischen Ansätzen. Der insbesondere durch ethische Aspekte, rechtliche Randbedingungen und Fragen der sozialen Technikgestaltung (Ethical, Legal and Social Aspects, kurz: ELSA bzw. Ethical, Legal and Social Implications, kurz ELSI) vorgegebene Rahmen wird in den Vorgehensmodellen nicht explizit adressiert. Ethische Werte wie Menschenwürde, Achtung des Lebens, Solidarität und Gerechtigkeit, rechtliche Aspekte, wie sie sich aus den verschiedenen Rechtsgebieten wie bspw. Datenschutz-, Urheber- und Haftungsrecht ergeben, sowie Zielvorstellungen der sozialen Technikgestaltung, wie bspw. gute, d. h. nicht nur einkommenssichernde, sondern bspw. auch sinnstiftende Arbeit, werden bei der Durchführung von Projekten nach den bekannten Vorgehensmodellen nicht systematisch betrachtet (zumindest fordern es die Vorgehensmodelle nicht). Da Software teils auch für eine Vielzahl von künftigen, teils noch nicht konkret bekannten Anwendungsfällen entwickelt wird, nimmt die Herausforderung zu (so sind bspw. die Betroffenen noch nicht konkret bekannt und können nicht beteiligt werden; stattdessen sind zunächst andere Wege der Berücksichtigung ihrer Interessen zu wählen). Durch die fehlende explizite Integration von ELSA-Betrachtungen in gängige Vorgehensmodelle wie bspw. Scrum oder das V-Modell-XT ist nicht sichergestellt, dass in den Entwicklungsprojekten der notwendige Raum zur Reflexion von ELSA eröffnet wird. Dies steht im Widerspruch zur Bedeutung von ELSA-Aspekten, wie sie sich bspw. in den *Ethischen Leitlinien* der Gesellschaft für Informatik e. V. [Gl18] oder anderen Quellen wiederfindet, z. B. wissenschaftlich in [Go97],

gesellschaftlich in [Hi19, EK21] und sozialpartnerschaftlich in [Ve20]. Sowohl die wirtschaftliche Verantwortungszuweisung, wie sie sich bspw. in einer Umfrage einer großen Unternehmensberatung zeigt [We19, vgl. GV07], als auch die akademische Lehre ist diesbezüglich losgelöst von den typischen Vorgehensmodellen [vgl. Sp19, BH22]. Der Beitrag diskutiert die Problemstellung und zeigt mögliche Lösungsoptionen für den weiteren Diskurs auch im Rahmen des Workshops auf.

Hierzu wird im zweiten Abschnitt zunächst der Stand der Wissenschaft und Technik dargestellt. Hierzu werden zunächst Grundlagen zu Vorgehensmodellen zusammengefasst, bevor anschließend gezeigt wird, wie ELSA in bestehenden allgemeinen Vorgehensmodellen aktuell adressiert werden. Zusätzlich werden Vorgehensmodelle betrachtet, welche sich auf ELSA-Fragen spezialisiert haben.

Im dritten Abschnitt werden vier Thesen zur zukünftigen Betrachtung von ELSA in Softwareentwicklungsprojekten aufgestellt. Diese sollen die weitere Diskussion anregen und zur Weiterentwicklung einer systematischen Betrachtung von ELSA führen.

2 Stand der Wissenschaft und Technik

2.1 Vorgehensmodelle

Softwareentwicklungsprojekte werden in der Regel durch organisations- und ggf. projektspezifische Adaption vorhandener Vorgehensmodelle strukturiert. Wie Abbildung 1 zeigt, dienen Vorgehensmodelle dazu, den Weg von einem bestimmten Ausgangspunkt aus zu einem bestimmten Ziel zu gestalten. Nach [Br96] kombiniert dazu ein Vorgehensmodell (synonym: eine Vorgehensstrategie) verschiedene Methoden oder Methodenfragmente. Folgende Komponenten

definieren eine Methode [vgl. Wi03]: Der Prozess (teils synonym: Vorgehensmodell innerhalb der Methode) definiert, welche Aktivitäten in welcher Reihenfolge (ggf. mit zeitlicher Überschneidung) und unter welchen Bedingungen von welchen Rollen ausgeführt werden. Aktivitäten und Rollen sind in der Abbildung gesondert dargestellt. Die im Hintergrund liegenden weiteren Kästchen in der Abbildung verdeutlichen, dass es innerhalb einer Methode mehrere Aktivitäten bzw. Rollen gibt. Rollen, wie bspw. Modellierungsexperte oder Moderator, führen die Aktivitäten aus. Dabei können eine oder mehrere Rollen an der Durchführung einer Aktivität beteiligt sein.

Artefakte (Ergebnisse) werden von Aktivitäten bzw. im Rahmen von Aktivitäten erzeugt; einige Artefakte werden auch von wiederum anderen Aktivitäten als Eingabe benötigt [Wi03]. Dabei sind folgende Komponenten relevant:

- Sprache: Ergebnisse müssen in einer Sprache beschrieben werden. Dies kann eine Dokumentation der Ergebnisse in natürlicher Sprache sein oder auch eine Dokumentation der Ergebnisse in einem spezifizierten und ggf. spezialisierten Format. Die Sprachdefinitionen reichen von Satzschablonen [PR15, S. 57 ff.] über Tabellen und Diagramme bis hin zu formalisierten Modellen. Die Sprache sollte für eine stark formalisierte Anwendung nicht nur syntaktisch definiert sein, sondern es bedarf auch einer eindeutigen Festlegung ihrer Semantik [Br96].
- Technik: Technik bezeichnet „die jeweilige Vorschrift zur Erstellung (und damit zur Dokumentation) der Ergebnisse“ [Wi03, S. 88].
- Werkzeuge: In einer Methode können Werkzeuge bspw. zur Unterstützung bei der Technik dienen. Die Werkzeuge können unterschiedliche Aktivitäten einer Methode unterstützen, beispielsweise die werkzeuggestützte

Moderation einer Maßnahmenplanung (z. B. mithilfe von MiroBoard, <https://miro.com/de>), die Unternehmensmodellierung (bspw. mithilfe des Horus Business Modeler, <https://www.horus.biz>), die Prozessmodellierung (bspw. ebenfalls mit Horus oder dem Camunda Modeler, <https://camunda.com/de>) und die Formulierung von User Stories.

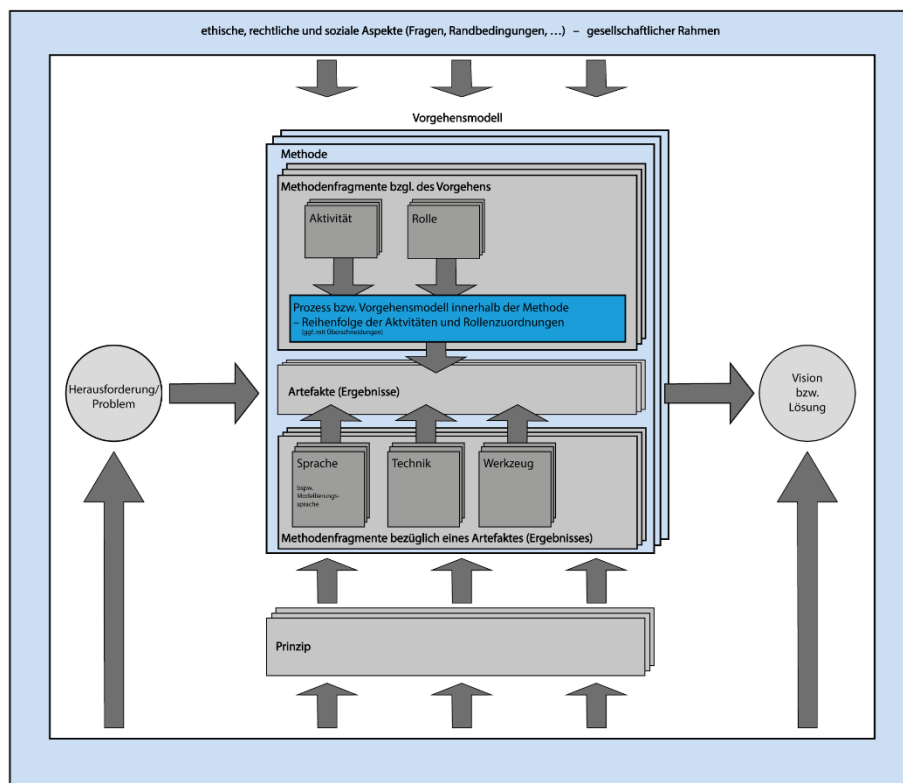


Abb. 1: Ethische, rechtliche und soziale Aspekte als Rahmen für Vorgehensmodelle
(basierend auf [AI21])

Für Softwareentwicklungsprojekte gibt es eine Vielzahl von Vorgehensmodellen, die – in unterschiedlicher Ausprägung und Intensität – diese Komponenten beinhalten. Dabei sind grundlegend die verschiedenen Philosophien *agil* und *klassisch* zu unterscheiden; von beiden sind unterschiedliche Vertreter in der Praxis im Einsatz [KL14].

GEFÖRDET VOM



Bundesministerium
für Bildung
und Forschung



Abbildung 1 stellt die ELSA als gesellschaftlichen Rahmen dar, was auch ihre Besonderheit ausmacht. Diese Aspekte unterscheiden sich von anderen Aspekten bzw. Anforderungen in Projekten, weil sie aus der Gesellschaft kommen bzw. durch die jeweilige Gesellschaft geprägt sind. Im Unterschied zu anderen Aspekten ist weder das Projekt- noch das Unternehmensinteresse leitend, sondern gesellschaftliche Zielvorstellungen. Diese lassen sich teils – im Falle von rechtlichen Aspekten – explizit durch Gesetze und Rechtsprechung als rechtlichen Rahmen beschreiben (d. h., in der Regel hat der gesellschaftliche Einigungs- und Entscheidungsprozess schon stattgefunden), teils sind sie – im Falle von ethischen und sozialen Aspekten – noch nicht unbedingt dokumentiert bzw. es hat bisweilen noch kein hinreichender gesellschaftlicher Meinungsbildungsprozess stattgefunden.

2.2 Ethische, rechtliche und soziale Aspekte (ELSA) in Vorgehensmodellen

Die Vorgehensmodelle für Softwareentwicklungsprojekte haben keine strukturelle Verankerung für ethische, rechtliche und soziale Aspekte. Eine solche Verankerung durch explizite Elemente (z. B. in Form von spezifischen Aktivitäten, Rollen oder Artefakten) existiert in typischen Vorgehensmodellen nicht. Ethische, rechtliche und soziale Aspekte werden – soweit es geschieht – typischerweise im Rahmen der Anforderungserhebung betrachtet und es werden (konkrete) nichtfunktionale Anforderungen erhoben, abgestimmt und dokumentiert. Die weitere Betrachtung im Projekt – insbesondere die Nachverfolgung – erfolgt dann, wie bei anderen nichtfunktionalen Anforderungen auch, im Rahmen des Anforderungsmanagements. Dieser Betrachtungsweg wird auch für andere spezifische Aspekte (wie bspw. IT-Sicherheit) genutzt.

In agilen Vorgehensmodellen gibt es – aufgrund der erst während des Projektverlaufs nach und nach entstehenden funktionalen Anforderungen – den Ansatz, als zusätzliche Qualitätsbedingung an eine einzelne (funktionale) Anforderung zu verlangen, dass diese ethisch, rechtlich und sozial abgestimmt ist. Im Vorgehensmodell Scrum wird die Qualitätsbedingung an eine Anforderung in der

„Definition of Ready“ [Da19] festgelegt. Diese „Definition of Ready“ kann dann auch das Erfordernis einer ELSA-Abstimmung für eine (funktionale) Anforderung enthalten. Dieser Betrachtungsweg in agil gesteuerten Vorhaben wird ebenfalls für andere spezifische Aspekte (wie bspw. IT-Sicherheit) genutzt.

Im Folgenden werden einige ausgewählte Vorgehensmodelle bzw. auch weitere Standards noch genauer betrachtet.

Das V-Modell-XT beschreibt in einem Abschnitt „Werte und Leitlinien“ Kerngedanken des Vorgehensmodells. Dabei heißt es unter anderem: „Das V-Modell ist anwenderfreundlich: Das V-Modell XT ist nicht nur Regelwerk, sondern Hilfsmittel“ [An06, S. 6]. Es gibt aber keine Elemente im Regelwerk oder in den Hilfsmitteln, die eine Betrachtung von ELSA systematisch unterstützen. Jedoch existieren bereits Rollen (wie bspw. der Lenkungsausschuss und das Change Control Board) und Aktivitätsgruppen (wie bspw. das Anforderungsmanagement), in die sich entsprechende ELSA integrieren lassen würden.

Für die agilen Vorgehensmodelle kann das *Manifest für Agile Softwareentwicklung* [Be01] als Zusammenfassung der Kernphilosophie zur ersten Analyse herangezogen werden. Es wurde 2001 von 17 Unterzeichnern als kleinster gemeinsamer Nenner verschiedener agiler Vorgehensmodelle formuliert. In vier Werten und zwölf Prinzipien werden die Grundgedanken beschrieben. Das erste und dritte Wert formuliert: Die Unterzeichner des Manifests schätzen „Individuen und Interaktionen mehr als Prozesse und Werkzeuge“ und „Zusammenarbeit mit dem Kunden mehr als Vertragsverhandlung“ [Be01]. Die Orientierung an natürlichen Personen fördert den menschenzentrierten Ansatz, wie er auch in der sozialen Technikgestaltung existiert [vgl. Kl19]. Die Prinzipien des Manifests legen den Fokus auf Kunden und Fachexperten (oft Fachanwender) sowie Entwickler. Weiteren Betroffene der Technikgestaltung werden nicht explizit genannt.

Die siebte Version des „standard for project management and a guide to the project management body of knowledge“ (PMBOKv7) [PM21a] des Project Management Institute (PMI) beschreibt, dass die zwölf im PMBOKv7 beschriebenen Prinzipien des

Projektmanagements so gestaltet sind, dass sie zu dem „PMI Code of Ethics and Professional Conduct“ [PM06] passen. Zusätzlich gibt es seit 2011 das „PMI Ethical Decision-Making Framework“ (EDMF), das als Leitfaden zur ethischen Entscheidungsfindung dient [PM21b]. Der Leitfaden beschreibt fünf Schritte zur Entscheidungsfindung und bezieht neben ethischen Aspekten auch rechtliche Rahmenbedingungen ein. Dabei wird die ethische Verantwortung der Entscheidungsträger durch eine Art Fragenkatalog unterstützt, der die vier Werte des PMI Code of Ethics [PM06] fokussiert: Verantwortung, Respekt, Fairness und Ehrlichkeit [PM06, deutsche Übersetzung]. ELSA wird also vom Project Management Institute ebenfalls nicht strukturiert bspw. im Rahmen spezieller Aktivitäten direkt im Standard verankert, aber einerseits durch den gesonderten fünfschrittigen Fragenkatalog EDMF gefördert; andererseits wirkt der Ethikcode auf die zwölf Prinzipien des Standards ein. Die aktuelle ELSA-Betrachtung des PMI entspricht also einem Teil der Darstellung in Abb. 1: Aus ELSA – bzw. konkreter hier fokussiert aus ethischen Aspekten – werden Prinzipien abgeleitet (bzw. im Falle des PMI unterstützen die Prinzipien jedenfalls den Code of Ethics), die dann auf das Vorgehen einwirken.

In der „Individual Competence Baseline für Projektmanagement“ gibt es in der aktuellen Version 4.0 (deutsche Fassung, kurz IPMA ICB4 PM) im Anhang F sowohl den Ethikkodex der GPM Deutsche Gesellschaft für Projektmanagement e. V. als auch den der International Project Management Association. Dort heißt es unter anderem: „Jeder Projektmanager räumt dem Gemeinwohl sowie der Gesundheit und Sicherheit jedes einzelnen Menschen hohe Priorität ein. Er trachtet nach Verbesserung der Lebensverhältnisse und der Umweltqualität. Weltoffenheit und Toleranz gegenüber anderen Kulturen bestimmen seine Haltung“ [SV, S. 212]. Außerdem gibt es im Standard selbst die Perspektive „Kultur und Werte“, u. a. mit dem Kompetenzindikator „Das Projekt mit der formellen Kultur und den Werten der Organisation in Einklang bringen“. Hier wird auch die Sozialverantwortung der Organisation adressiert. Dabei werden sowohl der projektdurchführenden Organisation als auch den einzelnen Mitarbeitenden Verantwortungen zugewiesen [SV, S. 61–63]. Zudem gibt es die Perspektive „Persönliche Integrität und

Verlässlichkeit“. Dort ist unter anderem definiert: „Persönliche Integrität bedeutet, dass der Einzelne gemäß seiner eigenen moralischen und ethischen Werte und Prinzipien handelt“ [SV, S. 70] – einzelne Kompetenzindikatoren führen dies weiter aus. Die rechtlichen Aspekte werden aus der Perspektive „Compliance, Standards und Regularien“ betrachtet und in einzelnen Kompetenzindikatoren (u. a. „Die für das Projekt gültigen Rechtsvorschriften identifizieren und einhalten“ [SV, S. 54]) ausgeführt. Diese Beispiele zeigen, dass die IPMA ICB4 PM die Erwartungshaltung an entsprechende Kompetenzen von Projektmanagern ausdrückt; allerdings beschreibt IPMA ICB4 PM selbst kein die entsprechende Verantwortung strukturiert integrierendes Vorgehensmodell.

Die Möglichkeiten einer strukturierten Betrachtung von ELSA-Aspekten zeigen Spezialnormen auf. Ein Vorgehensmodell aus einer Spezialnorm lässt sich jedoch nicht unmittelbar in die typischen Vorgehensmodelle des Software Engineering integrieren. Es wurde als unabhängiges und eigenständiges Vorgehen geschaffen und kann, wenn es die Verantwortlichen wissen und möchten, in Projekten zusätzlich eingesetzt werden. Ein Beispiel für eine solche Spezialnorm ist der IEEE-Standard „IEEE Model Process for Addressing Ethical Concerns during System Design“ (IEEE 7000-2021). Er wurde erstmals im Jahr 2021 publiziert. IEEE 7000-2021 sieht zwei Phasen vor: In der ersten Phase „Konzepterforschung“ werden zunächst das Einsatzkonzept und der Kontext erschlossen, um anschließend ethische Werte zu ermitteln und zu priorisieren. Der nächste Schritt „Definition ethischer Anforderungen“ beginnt noch in der Konzepterforschungsphase, wird jedoch in der Entwicklungsphase fortgesetzt. Diese besteht zusätzlich aus einem ethisch reflektierten Designprozess.

Auch im Kontext der Wirtschaftsethik gibt es verschiedene Ansätze für Unternehmen bzw. ihre Verantwortlichen, um zu ethisch reflektierten Entscheidungen zu kommen. Diese beziehen sich jedoch in der Regel auf die Ebene der Geschäftsmodelle und nicht auf die der Technikgestaltung. Gleichzeitig bieten auch diese Ansätze eine Möglichkeit für die ethische Betrachtung von (Informations-)Technik [UI98] allgemein bzw. spezifischer für die Integration von ELSA in Vorgehensmodelle des Software Engineering zu lernen. Ein

Beispiel ist die Gestaltung des Spannungsfeldes zwischen institutioneller und individueller Verantwortung [vgl. Ul16].

3 Thesen zur Weiterentwicklung

Im vorherigen Abschnitt wurde der aktuelle Stand der Wissenschaft und Technik beschrieben. Die Frage ist, wie dieser sich weiterentwickeln wird bzw. wie er aktiv weiterentwickelt werden kann. Die nachfolgenden Thesen verstehen sich als Diskussionsbeitrag hierzu und als Arbeitsimpuls für die Fachgruppen Vorgehensmodelle und Projektmanagement der GI bzw. allgemein für die Wissenschaft und Praxis.

- These 1: Es ist für eine effektive und effiziente Einbeziehung von ethischen, rechtlichen und sozialen Aspekten nicht ausreichend, diese allgemein als (gesellschaftlichen) Rahmen von Vorgehensmodellen bzw. Entwicklungsprojekten zu betrachten. Diese Sichtweise alleine ‚standardisiert‘ die Betrachtung von ELSA nicht und unterstützt Verantwortliche und durchführende Mitarbeitende nicht ausreichend.
- These 2: ELSA sind zu andersartig (bspw. verglichen mit anderen Anforderungen), um vollständig „nebenbei“ in dafür nicht dediziert ausgelegten Elementen von Vorgehensmodellen (z. B. des Requirements Engineering) beachtet zu werden (gesellschaftsrelevant, teils komplex, ...).
- These 3: Wir brauchen mehr strukturelle Verankerungen von ELSA in Standardvorgehensmodellen: spezifische Aktivitäten (z. B. Quality Gates mit Ethikcheck, Einbeziehung von „Betroffenen“ wie ggf. der Vertretung von Arbeitnehmer/-innen), spezifische Rollen (z. B. Ethikverantwortliche/-r), konkrete Verankerung der Aktivitäten und Rollen in einem Vorgehensmodell, spezifische Artefakte (z. B. Werteregister), ...; Dadurch wird – je nach Vorgehensmodell – die Perspektive von ELSA auch auf den gesamten Lebenszyklus von Systemen erweitert. So wird auch integriert, dass sich ELSA-Betrachtungen bzw. -Anforderungen während des Lebenszyklus ändern können und dann bspw. im Rahmen von „Wartung“ umgesetzt werden müssen.

- These 4: Es braucht Aktivitäten, die von einer Instanz (eines Vorgehensmodells) unabhängig sind. Hierzu gehört die Bildung und Pflege eines Wertekanons einer Organisation. Dieser kann dann in organisationseigenen Projekten verwendet werden und als eine Arbeitsgrundlage für die Abstimmung in organisationsübergreifenden Projekten dienen (wobei es unverhandelbare Werte geben wird, ohne deren Achtung eine Projektpartnerschaft mit anderen Organisationen nicht möglich ist). Andere Maßnahmen sind die weitere Sensibilisierung von bestimmten Berufsgruppen allgemein und Mitarbeitenden einer Organisation durch geeignete Aus- und Weiterbildung (vgl. GI-Ethikstandard). Im Allgemeinen wird die Umsetzung von These 4 zu einem kulturellen Wandel führen.

4 Fazit

Verantwortungsträger und Gestalter unterschiedlicher Handlungsebenen sind weiterhin gefordert, auf ihrer jeweiligen Handlungsebene die notwendige Verantwortung zu übernehmen bzw. Gestaltungsangebote für ethisch verantwortungsvolle, rechtlich zulässige und sozial gute Verfahren und Lösungen zu machen. Für eine systematische Betrachtung und Integration von ELSA in Softwareprojekten sind die dargestellten Themen ein Diskussions- und Arbeitsimpuls. Einen Diskurs hierzu können Gestalter von Vorgehensmodellen ebenso führen, wie sie prototypische, aber wohlüberlegte Anpassungen des Vorgehens im eigenen Kontext (Unternehmen, Projekt, ...) umsetzen bzw. anregen können. Diese Anpassungen können – bei entsprechender Bereitschaft – wissenschaftlich begleitet werden, um so aus den jeweiligen Kontexten allgemeine Erkenntnisse zu abstrahieren.

Für eine nachhaltige und kontextübergreifende Integration von ELSA in das Vorgehen in Softwareprojekten ist ein breiter Austausch über Integrationsoptionen notwendig. Dieser kann durch einzelne Diskussionsbeiträge, wie bspw. durch Projektberichte über ein angepasstes Vorgehen, weiter angeregt werden. Durch den Diskurs können dann angepasste, kontextunabhängige

Vorgehensmodelle entstehen. Darüber hinaus können Weiterbildungsangebote zur geeigneten Integration von ELSA in die bestehenden Vorgehensmodelle ein erster Schritt zur Anpassung von etablierten Vorgehensmodellen in der Praxis sein.

Literaturverzeichnis

- [AI21] Alpers, S; Karle, T; Schreiber, C.; Schönthaler, F.; Oberweis, A.: Process Mining bei hybriden Vorgehensmodellen zur Umsetzung von Unternehmenssoftware, In: Informatik Spektrum, 44, S.178-189. <https://doi.org/10.1007/s00287-021-01359-7>, 2021.
- [An06] Angermeier et al.: V-Modell XT – Das deutsche Referenzmodell für Systementwicklungsprojekte Version: 2.3. Verein zur Weiterentwicklung des V-Modell XT e. V., <http://ftp.tu-clausthal.de/pub/institute/informatik/v-modell-xt/Releases/2.3/V-Modell-XT-Gesamt.pdf>, 2006, zuletzt abgerufen am 04.07.2022.
- [Be01] Beck, K.; Beedle, M.; Bennekum, A.; Cockburn, A.; Cunningham, W.; Fowler, M.; Grenning, J.; Highsmith, J.; Hunt, A.; Jeffries, R.; Kern, J.; Marick, B.; Martin, R.; Mellow, S.; Schwaber, K.; Sutherland, J.; Thomas, D.: Manifesto for Agile Software Development, <https://agilemanifesto.org>, 2001. Deutsche Übersetzung Filho, R.; Preuss, I.; Bless, M.; Ehls, F.; Roock, S.; Eckstein, J.; Hoehn, S.; Schiffer, B.: Manifest für Agile Softwareentwicklung, zuletzt abgerufen am 04.07.2022.
- [BH22] Baase, S.; Henry, T.: A Gift of Fire – Social, Legal, and Ethical Issues for Computing Technology. 5th Edition. Pearson, New York. 2022.
- [Br96] Brinkkemper, S.: Method engineering: engineering of information systems development methods and tools. In: Information and Software Technology, 38/4,

- S. 275–280, [https://doi.org/10.1016/0950-5849\(95\)01059-9](https://doi.org/10.1016/0950-5849(95)01059-9), 1996.
- [Da19] Dalton, J.: Definition of Ready. In: Great Big Agile. Apress, Berkeley. https://doi.org/10.1007/978-1-4842-4206-3_26, 2019.
- [EK21] Evangelische Kirchen in Deutschland: Freiheit digital. Die Zehn Gebote in Zeiten des digitalen Wandels. Eine Denkschrift der Evangelischen Kirche in Deutschland. Evangelische Verlagsanstalt, Leipzig. 2021.
- [GC07] Grimm, P.; Capurro, R. (Hrsg.): Wirtschaftsethik in der Informationsgesellschaft. Schriftenreihe Medienethik. Franz Steiner Verlag. 2007.
- [GI18] Gesellschaft für Informatik e. V. (GI): Ethische Leitlinien, <https://gi.de/ethische-leitlinien>, zuletzt abgerufen am 07.07.2022, 2018.
- [Go97] Gotterbarn, D.; Miller, K.; Rogerson, S.: Software engineering code of ethics. In: Cohen, J. (Hrsg): Communications of the ACM, 40/11, New York, Association for Computing Machinery, S. 110–118, <https://doi.org/10.1145/265684.265699>, 1997.
- [Hi19] Himmer, N.: Computer und die Moral – Philosophische Nachhilfe für Nerds, Frankfurter Allgemeine Zeitung Online, <https://www.faz.net/-gyl-9ibi7>, 2019, zuletzt abgerufen am 04.07.2022.
- [KL14] Kuhrmann, M.; Linssen, O.: Welche Vorgehensmodelle nutzt Deutschland? In: Engstler, M.; Hanser, E.; Mikusz, M.; Herzwurm, G. (Hrsg.) Projektmanagement und Vorgehensmodelle. Bonn: Gesellschaft für Informatik e. V., S. 17–32, 2014.
- [KI19] Klose, E.; Ni, I; Schmidt, L.: Mit benutzerzentrierter Entwicklung zur Integration von sozialen Aspekten in die

- Projektarbeit. In INFORMATIK 2019: 50 Jahre Gesellschaft für Informatik – Informatik für Gesellschaft (Workshop-Beiträge). S. 447-463. https://dx.doi.org/10.18420/inf2019_ws49, 2019.
- [PM06] Project Management Institute: PMI Code of Ethics and Professional Conduct. <https://www.pmi.org/codeofethics>. 2006, zuletzt abgerufen am 04.07.2022
- [PM21a] Project Management Institute: The standard for project management and a guide to the project management body of knowledge (PMBOK guide). Seventh Edition. 2021.
- [PM21b] Project Management Institute: PMI Ethical Decision-Making Framework. <https://www.pmi.org/codeofethics>. 2021, zuletzt abgerufen am 04.07.2022.
- [PR15] Pohl, K.; Rupp, C.: Basiswissen Requirements Engineering (4. Auflage). dpunkt.verlag GmbH. 2015.
- [Sp19] Spiekermann, S.: Digitale Ethik – Ein Wertesystem für das 21. Jahrhundert. Droemer Verlag. 2019.
- [SS20] Schwaber, K.; Sutherland, J.: The Scrum Guide. <https://scrumguides.org/docs/scrum-guide/v2020/2020-Scrum-Guide-US.pdf>. 2020, zuletzt abgerufen am 04.07.2022.
- [SV17] Schoper, Y.; Viehbach, A. (Leitung Redaktion): Individual Competence Baseline für Projektmanagement, Version 4.0 (IPMA ICB4 PM), International Project Management Association, GPM Deutsche Gesellschaft für Projektmanagement e. V., 1. aktualisierte Auflage 2017.
- [UI98] Ulrich, P.: Integrative Wirtschaftsethik – eine Heuristik auch für die Technikethik? In: Lenk, H.; Maring, M.

- (Hrsg.): Technikethik und Wirtschaftsethik. VS Verlag für Sozialwissenschaften, Wiesbaden. https://doi.org/10.1007/978-3-322-97402-0_3, 1998.
- [Ul16] Ulrich, P.: Integrative Wirtschaftsethik. Grundlagen einer lebensdienlichen Ökonomie, 5. Auflage, Paul Haupt Verlag, Bern/Stuttgart/Wien. 2016.
- [Ve20] ver.di Vereinte Dienstleistungsgewerkschaft: Ethische Leitlinien für die Entwicklung und den Einsatz von Künstlicher Intelligenz (KI): Gemeinwohl und Gute Arbeit by Design. https://innovation-gute-arbeit.verdi.de/++file++5e561a72452768ee1b1845cd/download/verdi_Ethische_Leitlinien_KI_170220.pdf. 2020, zuletzt abgerufen am 04.07.2022.
- [We19] Weyer, F. (Ansprechpartnerin): PricewaterhouseCoopers GmbH Wirtschaftsprüfungsgesellschaft. 2019. <https://www.pwc.de/de/managementberatung/berichtsband-digitale-ethik-vorabversion.pdf>, zuletzt abgerufen am 04.07.2022
- [Wi03] Winter, R.: Modelle, Techniken und Werkzeuge im Business Engineering. In Business Engineering. Springer, Berlin, Heidelberg, S. 87-118. https://doi.org/10.1007/978-3-642-19003-2_5, 2003.

Warum Ethik in KARL?

Ohne Zweifel steht die Menschheit vor großen globalen Herausforderungen wie Klimawandel mit seinen verheerenden Folgen weltweit, Pandemien und sozialen Verwerfungen in nahezu alle Gesellschaften. Daher stellt sich nicht mehr die Frage, ob ein gesellschaftlicher Wandel im Hinblick auf eine nachhaltige Entwicklung stattfindet, sondern wie und nach welchen sozialen und ethischen Prämissen dieser Wandel gestaltet werden kann und soll. Hierbei ist deutlich geworden, dass vor allem Technik, Wirtschaft und gesellschaftliche Entwicklungen zusammengeführt werden müssen und nicht mehr als unterschiedliche Bereiche getrennt behandelt werden können. Das betrifft in hohem Maße die technologische Entwicklung, die noch immer als Motor für wirtschaftliche Konkurrenzfähigkeit hochindustrialisierter Länder und somit als Garant für Wirtschaftswachstum und Wohlstand in eben diesen Ländern gilt. Unabhängig von den kontinuierlich diskursiv hergestellten Zusammenhängen zwischen neuen Technologien und Wirtschaftswachstum, wird seit Jahren auch die Frage intensiv diskutiert, inwiefern technologische Innovationen einen Beitrag für die Umsetzung einer nachhaltigen Entwicklung leisten können [1, 2]. Diese Frage wurde und wird intensiv im Rahmen der Digitalisierung sowie den mannigfaltigen Möglichkeiten der Künstlichen Intelligenz (KI) diskutiert. Die Fortschritte der KI in Arbeits- und Produktionsbereichen sowie im Alltag sind rasant und mit diesen technischen Möglichkeiten verbinden sich zum einen große Hoffnungen, aber auch große Befürchtungen. Neben diesen technischen Optionen bleiben jedoch die folgenden Fragen bestehen, die eine hohe Aktualität besitzen: welchen Beitrag leistet die KI im Rahmen einer ressourcenschonenden, sozialverträglichen und solidarischen Gesellschaft? Welche sozialen und gesellschaftlichen Problemstellungen können auf der Basis von KI

gelöst werden? Welche nicht? Vor welchen sozialen und ethischen Werten wird die KI entwickelt? Welche gesellschaftlichen Visionen werden durch die KI umgesetzt?

Das Projekt KARL ist in diesen umfassenden und anspruchsvollen Diskussionskontext eingebettet. So ist das ausgeflaggte Ziel von KARL, „menschzentrierte, transparente und lernförderliche KI-unterstützte Arbeits- und Lernsysteme“ (KI-ALS) zu konzipieren und in konkreten Praxisanwendungen anzuwenden.² Diese Praxisfelder beziehen sich auf sehr unterschiedliche gesellschaftliche Bereiche wie beispielsweise Dienstleistungen, produzierendes Gewerbe und Mobilität. Die KI gilt hierbei als Schlüsseltechnologie, deren Potentiale im Sinne zukunftsweisender und transparenter Arbeits- und Lernmodelle entfaltet werden sollen. Ethische, soziale und rechtliche Rahmenbedingungen sollen in dieser Gestaltung eine wichtige Rolle spielen.

Damit greift KARL relevante Rahmenbedingungen auf und hat sich zum Ziel gesetzt, diese Parameter in der Gestaltung der KI in den ausgewählten Handlungsfeldern zu berücksichtigen. Freilich wird die Frage, wie die KI als neue Form des technischen Fortschritts menschengerecht, transparent und im Sinne einer nachhaltigen Entwicklung eingesetzt werden kann, seit einigen Jahren intensiv diskutiert. „Die zunehmende Nutzung dieser immer komplexer werdenden Systeme hat weltweit Debatten über diskriminierende Effekte,

² Ziel des Kompetenzzentrums KARL – Künstliche Intelligenz für Arbeit und Lernen in der Region Karlsruhe – ist es, menschenzentrierte, transparente und lernförderliche KI-unterstützte Arbeits- und Lernsysteme (KI-ALS) zu konzipieren und in konkreten Praxisanwendungen vorzeigbar zu machen. Es werden Regeln, Handreichungen und vorzeigbare Konzepte von KI-ALS gestaltet und im Praxiseinsatz erprobt, welche die Vorteile menschlicher und künstlicher Intelligenz kombinieren. Das Vorhaben KARL fokussiert hierzu auf vier Anwendungsdomänen (AD), in denen die Region Karlsruhe ausgewiesene wirtschaftliche und technologische Stärken aufweist: (1) Mobilität und autonomes Fahren, (2) Wissensintensive Dienstleistungen und IKT-Systeme, (3) Produzierendes Gewerbe und (4) Bildungsbereich. In diesen ist die Anwendung von KI-Systemen bereits weit fortgeschritten bzw. es wird ein starker Einfluss auf die zukünftigen Arbeits- und Lernsysteme erwartet [3].

GEFÖRDERT VOM



Bundesministerium
für Bildung
und Forschung



intransparente Entscheidungs- und Optimierungsprozesse oder die Reproduktion gesellschaftlicher Ungleichheiten aufgeworfen“ [4, p. 5]. So gibt es inzwischen eine Reihe an nationalen und internationalen Publikationen, die sich – je nach Themenfeld – mit den Möglichkeiten und Risiken von KI auseinandersetzen [4]. Die ethische Dimension weist hierbei auf die gesellschaftlichen Werte und Normen, denen die Technikgestaltung zugrunde liegen und in der Regel bewusst oder unbewusst reproduziert werden. So werden seit einiger Zeit Fragen an technologische Innovationen als einem „menschenrechtlichen“ Fortschritt gestellt [5, p. 27], der eng an neue Vorstellungen des Begriffs der Wertschöpfung gekoppelt sind. So konstatiert beispielsweise die Ökonomin Sarah Spiekermann:

„Bewusst verwende ich den Begriff der „Wertschöpfung“, den ich etwas anspruchsvoller auslege, als die herkömmliche Wirtschaftswissenschaft es tut. [...]. Vielmehr geht es mir um die Frage, wie es gelingen kann, dass wir durch die Digitalisierung die gefühlten menschlichen und gesellschaftlichen Werte fördern, deren Erleben sich kaum in Geld umrechnen lässt. Zum Beispiel frage ich mich, wie man digitale Systeme so bauen und nutzen kann, dass sie das menschliche Wissen aufbauen helfen, Freundschaften unterstützen, menschliche Freiheit und Privatheit erhalten und zum gegenseitigen Respekt beitragen.“ [5, p. 27]

Dieses Zitat weist sehr schön auf die Bedeutungsinhalte hin, die alle digitalen Systeme in ihren jeweiligen sozialen und betrieblichen Kontexten vermitteln. Die (kritische) Reflexion dieser Werte könnte auch die Abkehr von traditionellen Werten und das Erarbeiten neuer Lebens- und Arbeitsvisionen auf der Basis von KI beinhalten. Die Reflexion der Werte, die der KI zugrunde liegen, wird zwar vielfach gemacht [6], häufig verbleiben diese Ansätze jedoch auf einer abstrakten Ebene und verweisen wenig auf die Operationalisierung in konkreten Anwendungsfeldern oder gar auf die Bedeutung eines Wertewandels hin.

Vor diesem Hintergrund steht in den folgenden Ausführungen das Feld der Ethik im weiteren Sinne für Reflexionsangebote aus der (Angewandten) Ethik [6] und des Konzepts der Nachhaltigkeit [7].³ Diese Reflexions- und Diskursangebote stammen aus sehr unterschiedlichen Forschungstraditionen und Wissensbezügen, stellen unterschiedliche Fragen und adressieren vielseitige Probleme [8]. Es zeigt sich vor allem durch die KI, dass sich die Fragestellungen verändert haben, denn angesichts der Komplexität der Problemstellungen wird „der moralische Kompass der Ethik“ zunehmend relevanter in den neuen Mensch-Maschine-Interaktionen [9, 10]. Themenfelder wie Verantwortungsethik, Handlungsethik und oder Technikethik werden zusammengeführt und mit unterschiedlichen Problemstellungen verknüpft. Das hat zur Folge, dass wissenschaftliches Wissen, Prozesswissen, aber auch Erfahrungs- und Alltagswissen der unterschiedlichen Akteure gleichermaßen und in gegenseitiger Anerkennung in die Technikgestaltung mit einbezogen werden [11, 12]. Die verwickelten und mitunter konfliktreichen Relationen zwischen Wissen und Werten treten so deutlicher hervor, wie dies u.a. die aktuelle pandemische Situation eindrücklich gezeigt hat. So wird die Berücksichtigung vielfältiger handlungsrelevanter Wissens Ebenen im Suchprozess nach verbindlichen Werten und Normen entscheidend, um zu verlässlichen Gestaltungsprozessen zu kommen. Dieser Prozess erfordert allerdings die Ausbildung eines „genuin praktischen, situationsbezogenen Gebrauchs von Vernunft“ [13], der sich einerseits aus dem Abwägen dieser Wissens Ebenen, andererseits aber auch aus den gesellschaftlichen Werten und Normen speist. Im Folgenden werden unterschiedliche Guidelines im Hinblick auf die ethische Dimension vorgestellt. So wird deutlich, dass es keine generalisierte Gestaltung von KI geben kann, sondern dass es

³ Im Antrag von KARL werden diese Angebote unter ELSI gefasst, was noch die sozialen und rechtlichen Rahmenbedingungen miteinbezieht.

lediglich einen konstruktiven Aushandlungsprozess über die Werte, die der Ausgestaltung der Technologie zugrunde gelegt wird, geben kann [8, 14]. Dies gilt insbesondere für den Bereich Arbeit, in der neue Technologien seit der Industrialisierung einen wichtigen Raum für die Gestaltung der Arbeitsprozesse und der Organisation der Arbeit einnehmen. Aber auch hier gilt: „[...] Vielmehr wird es immer dringlicher, den Verlockungen technisch-pragmatischer Lösungen – in der Theoriebildung ebenso wie im politischen Handeln – mit Skepsis zu begegnen [...]“ [15, p. 95]. Die Gestaltung von KI in Arbeitsprozessen und somit die Gestaltung zukünftiger Arbeit ist eine gesellschaftliche Aufgabe geworden, die sich nicht mehr auf „kurzatmige Reparaturversuche“ [15, p. 95] reduzieren lässt. Im Folgenden werden unterschiedliche Dimensionen im Dreiecksverhältnis KI-Arbeit-Ethik vorgestellt, um die Bandbreite dieses Verhältnisses zu verdeutlichen.

KI & Ethik

Angeichts der gesellschaftspolitischen wie wissenschaftlichen Brisanz der Einführung von KI in der Arbeitswelt verwundert die Vielzahl der inzwischen veröffentlichten (Ethik-)Guidelines kaum. Im Folgenden sollen, wie bereits angekündigt, drei dieser Guidelines vergleichend vorgestellt werden. Diese wurden von so verschiedenen Akteuren wie der Berlin-Brandenburgischen Akademie der Wissenschaft (BBAW), dem Institut für Ökologische Wirtschaftsforschung (IÖW) und der Europäischen Kommission veröffentlicht. Anschließend werden die Gemeinsamkeiten und Unterschiede festgehalten.

Die „Nachhaltigkeitskriterien für Künstliche Intelligenz“ [4] veröffentlicht vom IÖW, stellen einen ganzen Katalog mit 13 Kriterien sowie mit diesen verbundenen Indikatoren für eine nachhaltige KI dar. Das Paper richtet sich an Entwickler*innen der KI und zeichnet sich durch den Anspruch aus, einen ganzheitlichen Begriff von Nachhaltigkeit zu entwickeln. KI soll dementsprechend sowohl nachhaltig wirken – Stichwort „nachhaltige KI“ – als auch Nachhaltigkeit ermöglichen (KI der Nachhaltigkeit) (Ebenda: 19). Entlang der drei Nachhaltigkeitsdimensionen sozial, ökologisch und ökonomisch werden schrittweise sechs soziale, vier ökologische sowie drei ökonomische Kriterien für eine Nachhaltige KI entwickelt. Die

sozialen Kriterien umfassen unter anderem „Transparenz und Verantwortung“, „Nicht-Diskriminierung und Fairness“ sowie „Selbstbestimmung und Datenschutz“. Das Kriterium „Transparenz und Verantwortungsübernahme“ wird noch ergänzt durch zwei Indikatoren, welche den Inhalt des Kriteriums weiter konkretisieren. Der Faktor Transparenz wird durch den ersten Indikator noch ergänzt um „Erklärbarkeit und Prüfbarkeit des Modells“ und der zweite lautet „Informationsmöglichkeit zur Funktionsweise des Systems“ [4, p. 58]. Aus diesen Indikatoren werden schließlich weitere Subindikatoren abgeleitet. Diese Subindikatoren werden dann danach unterschieden, ob sie sich an die Entwickelnde oder die Einsetzende Organisation richten. Das IÖW adressiert in dieser Ethik-Guideline also gezielt bestimmte Akteure und zeigt damit deren Verantwortlichkeit auf.

Das Heft „Verantwortungsvoller Einsatz von KI? Mit menschlicher Kompetenz!“ (BBAW 2021) entwickelt demgegenüber weder einen Kriterienkatalog, noch werden hier Handlungsanweisungen formuliert. Da es selbst nicht den Anspruch hat als unmittelbare Guideline zu dienen, bleibt es also im Vergleich zu den anderen hier vorgestellten Veröffentlichungen rein deskriptiv und stellt dementsprechend keine Forderungen auf. Nichtsdestotrotz ist diese Veröffentlichung insofern

relevant, als in ihr verschiedenste wissenschaftliche Perspektiven auf soziale und ethische Implikationen von KI auf die Gesellschaft zusammengebracht werden. Die Kapitel-Überschriften zeigen die gesetzten Schwerpunkte auf. Zu Beginn des Hefts wird generell auf das Thema Verantwortung im Kontext von KI eingegangen, anschließend zudem explizit auf die beiden Bereiche Entwicklung und Anwendung. Das erste Kapitel gliedert sich in eine philosophische, eine technische sowie zwei rechtswissenschaftliche Bestimmung des Verantwortungsbegriffs. Die Grundlage für die Diskussion des Verantwortungsbegriffs legt jedoch der philosophische Beitrag, der gegen den neurowissenschaftlichen Naturalismus die Entscheidungsfähigkeit der Menschen betont [16, p. 21].

Im zweiten Teil bezüglich der Entwicklung von KI wird betont, dass die Bewertung hinsichtlich einer möglichen Diskriminierung durch ein KI-System nicht nur bei den Entwickler*innen liegen darf. Stattdessen müssen für jeden Einzelfall alle Stakeholder in die Entscheidung eingebunden werden [17, p. 37]. Im letzten Abschnitt, welcher sich mit der Anwendung von KI-Systemen auseinandersetzt, fordern Hermann und Stock [18, p. 51] „sorgfältige Aus- und Weiterbildung“ sowohl für Nutzer*innen wie auch für Anwender*innen. So sollen diese in der Lage sein, kompetent mit diesen Systemen umgehen zu

können, da sie letztendlich nach wie vor die rechtliche Verantwortung für die Ergebnisse der KI tragen müssen. Das Heft der BBAW ist, wie eingangs bereits betont, keine Ethik-Guideline. Dennoch zeichnet es sich durch eine vielschichtige und angemessene Problembeschreibung aus. Darüber hinaus werden auch Handlungsanforderungen formuliert. Vor dem direkten Vergleich mit den beiden anderen Dokumenten, wird im Folgenden Abschnitt die dritte Veröffentlichung vorgestellt.

Die Europäische Kommission ist der Herausgeber des dritten Dokuments, welches abschließend vorgestellt wird. Die Kommission hat im Sommer 2018 eine Hochrangige Experten-Gruppe für Künstliche Intelligenz (HEG-KI) eingesetzt, die sich mit Leitlinien für eine vertrauenswürdige KI auseinandergesetzt hat. Der Hauptteil des Abschlussberichts ist gegliedert in die drei Kapitel: Fundamente einer vertrauenswürdigen KI, Verwirklichung einer vertrauenswürdigen KI und Bewertung einer vertrauenswürdigen KI. Vertrauenswürdige KI umfasst für die Verfasser des Papiers die drei Faktoren Rechtmäßige KI, Ethische KI und Robuste KI. Der Aspekt des Rechts wird allerdings in diesem Dokument vollkommen ausgespart. Um die Ethische KI weiter zu operationalisieren, werden stattdessen vier ethische Grundsätze formuliert. Diese umfassen die Achtung der menschlichen Würde, die Schadensverhütung,

Fairness sowie Erklärbarkeit [19, p. 14]. Diese sollen als „ethische Imperative“ von den maßgeblichen KI-Akteuren immer befolgt werden. Während sich die Achtung der Menschenwürde aus den Grundrechten ableiten, auf denen die EU gegründet wurde, zielt die Schadensverhütung neben der Achtung der Menschenwürde auch explizit auf die körperliche und geistige Unversehrtheit. Interessanter sind die beiden weiteren Grundsätze. Sowohl Fairness als auch Erklärbarkeit lassen sich nur durch die Nachvollziehbarkeit der Ergebnisse einer KI gewährleisten. Denn erst dann lassen sich etwaige ungerechte Ergebnisse erfassen und kritisieren. Der Grundsatz der Fairness bezieht sich dabei eher auf den Schutz einzelner Personen bzw. Gruppen sowie auf die Abwägung der Verhältnismäßigkeit zwischen Zweck und Mittel. Erklärbarkeit als letzter ethischer Grundsatz soll wie bereits erwähnt, die KI-Nutzenden in die Lage versetzen, die Ergebnisse der KI zu überprüfen. Dadurch, so die Hoffnung, kann sich Vertrauen in die Algorithmen etablieren. Aus den vier Grundsätzen werden anschließend sieben Anforderungen abgeleitet, welche die Vertrauenswürdigkeit einer KI sicherstellen sollen [19, pp. 17-18]. Da diese letztendlich lediglich die vier Grundsätze weiter aufschlüsseln und erläutern, werden diese hier nicht im Einzelnen vorgestellt. Nachdem bis zu dieser Stelle lediglich auf

theoretischer Ebene bestimmte Faktoren herausgearbeitet wurden, beginnt an dieser Stelle des Dokuments die praktische Handreichung der Guideline für die Entwicklung und Anwendung von KI. Im zweiten Teil des zweiten Abschnitts werden sowohl technische als auch nicht technische Verfahren vorgestellt, mit denen die Einhaltung der Grundsätze bzw. Anforderungen sichergestellt werden kann. Die technischen Verfahren beinhalten z.B. das Erproben und Testen der KI und die nicht technischen Verfahren unter anderem die Betonung der Wichtigkeit inklusiver Entwicklungsteams [19, p. 19]. Den Abschluss bildet eine Checkliste für Piloten einer Bewertungsliste für vertrauenswürdige KI. Im ersten Schritt werden dafür die unterschiedlichen Ebenen in den entwickelnden Organisationen identifiziert, welche durch die Bewertungsliste adressiert werden. Anschließend wird die Bewertungsliste dargestellt. Diese umfasst neuen Seiten und beinhaltet Fragen wie: „Haben Sie bspw. das Risiko drohender Arbeitsplatzverluste und Dequalifizierung von Arbeitnehmerinnen und Arbeitnehmer beurteilt?“ und weiter: „Welche Maßnahmen wurden ergriffen, um diesen Risiken entgegenzuwirken“ [19, p. 40]. Hier wird sowohl die Sinnhaftigkeit eines solchen Bewertungskatalogs als auch seine Grenzen deutlich. Denn selbstverständlich ist es wichtig, die Bedeutung dieser Frage zu betonen.

Dennoch bleibt die Wirksamkeit des Katalogs fraglich, wird doch Wertschöpfung in der Wirtschaft im Gegensatz zu der Definition von Spiekermann [5, p. 27] in erster Linie anhand von Geldwert beurteilt.

Wie deutlich geworden sein sollte, haben die drei Guidelines jeweils eigene Vorzüge. Allen drei gemeinsam ist das Verständnis für mögliche Probleme, welche im Verhältnis zwischen Mensch und intelligenter Maschine entstehen können. Zudem zeigen alle drei Veröffentlichungen durchaus sinnvolle Lösungen für die beschriebenen Probleme auf. Ihnen ist jedoch auch die Prämisse gemeinsam, dass KI per se zunächst als positiv und damit als sinnvoll zu betrachten ist. So fragt keine der hier vorgestellten Guidelines danach, ob die Anwendung von KI in einem bestimmten Bereich überhaupt sinnvoll ist oder ob es andere technische oder soziale Lösungen für ein Problem gibt. Erst wenn diese „radikalen“ Fragen gestellt werden, rücken auch Antworten in den Vordergrund, welche über das Spektrum an Lösungen im Bereich der KI hinausgehen. Auch sind Guidelines selbstverständlich kein Ersatz für gesetzliche Regulierungen. Indem sie die Entwicklungen im Bereich der KI jedoch kritischer begleiten als bislang könnten sie ihren Nutzen für die Gesellschaft jedoch deutlich erhöhen,

etwa in dem sie zu einer reflektierteren Gestaltung von KI beitragen.

Einige ethische Prämissen der Gestaltung von KI in der Arbeitswelt

Die Einführung von KI im Arbeitskontext wirft in diesem Sinne eine Reihe von Fragestellungen ethischer Relevanz auf. Diese ergeben sich einerseits allgemein durch die Relevanz des Einsatzes von KI für übergreifende gesellschaftliche Herausforderungen, etwa der Klimakrise. Andererseits können sich ethische Probleme aus dem spezifischen Anwendungskontext von KI in der Arbeitswelt bzw. der spezifischen Beschaffenheit von KI-Anwendungen ergeben. Zu denken wäre hier etwa an die populäre Angst vor negativen Beschäftigungseffekten der Automatisierung von Arbeit [20–23] oder die spezifische Risiken, die sich durch die Opazität lernender Systeme ergeben können [24].

Angeichts der sich rapide vertiefenden Klimakrise und einer fortschreitenden sozialen Polarisierung muss der technologische Wandel nicht zuletzt hinsichtlich seiner ökologischen und sozialen Implikationen befragt werden. Die konkrete Bewertung technologischer Artefakte erfordert dabei einen Ansatz, der weder der Versuchung einer „Ethik der Solution“ [25]

erliegt, die sich von technologischem Fortschritt die Lösung jedweder gesellschaftlichen Herausforderung erhofft und darüber die Notwendigkeit sozialer Innovationen verkennt, noch einer pauschalisierenden Technikfeindschaft verfällt. Stattdessen ist im Sinne der Brundtland-Definition⁴ Nachhaltigkeit als eine regulative Leitidee zu verstehen, die auf Gerechtigkeit abzielt – innerhalb der heutigen Menschheit wie auch gegenüber zukünftigen Generationen. Ein nachhaltiger Einsatz von KI lässt sich aus ethischer Perspektive entsprechend dreifach fassen: „Eine nachhaltige KI ist aus unserer Perspektive vorhanden, wenn Entwicklung und Einsatz dieser Systeme die planetaren Grenzen respektiert, keine problematischen ökonomischen Dynamiken verstärkt und den gesellschaftlichen Zusammenhalt nicht gefährdet.“ [4, p. 30]

In sozialer Hinsicht zielt Nachhaltigkeit als regulative Leitidee also in ihrer weiten ethischen Auslegung auch auf den Abbau sozialer Ungleichheit. Im Kontext von KI als Automationstechnologie wären hier zunächst mögliche Beschäftigungseffekte ihres Einsatzes abzuschätzen:

⁴ Nachhaltige Entwicklung ist nach ihr „eine Entwicklung, die die Bedürfnisse der Gegenwart befriedigt, ohne zu riskieren, daß künftige Generationen ihre eigenen Bedürfnisse nicht befriedigen können.“ [26, p. 46].

- bedroht dieser Einsatz bestehende – oder im Sinne intergenerationaler Gerechtigkeit auch zukünftige – Arbeitsplätze?
- Können die Arbeitenden ihre materielle Existenzsicherung gewährleisten auf der Basis von Erwerbsarbeit?
- Besteht die Möglichkeit einer Weiterqualifikation und eines Rollenwandels für Beschäftigte innerhalb des Unternehmens?
- Wie wird „gute“ Arbeit im Sinne einer sinnstiftenden Arbeit auf der Basis neuer Technologien gefördert?

Im Idealfall eines mitbestimmten Unternehmens können in diesem Kontext Fragen der Weiterbildung, Requalifikation und Beschäftigungssicherung in Form einer Betriebsvereinbarung geregelt werden [27]. Die Zentralität dieser Dimension liegt nicht zuletzt darin begründet, dass gesellschaftliche Integration weiterhin stark durch Lohnarbeit geleistet wird und diese den wesentlichen Mechanismus darstellt, der breiten Teilen der Bevölkerung (insbesondere materielle) Teilhabe ermöglicht.

Doch auch jenseits der – für viele Beschäftigten natürlich zentralen – Beschäftigungsproblematik, wirft der Einsatz von KI zahlreiche ethische Fragen auf. So kann es durch den Einsatz von KI zu Veränderungen von Arbeitsorganisation und -

Inhalten kommen. In diesem Kontext stellt sich die Frage, inwiefern die Würde der Beschäftigten gewahrt bleibt, beschränkt oder ausgebaut wird [15] – insbesondere in Form ihrer Autonomie, ihren Arbeitsprozess gestalten zu können. Die Unterwerfung von Beschäftigten unter Formen algorithmischen Managements [28], das sie zu Befehlsempfängern degradiert, die freie Entfaltung ihrer Befähigungen einschränkt [29] und ständige Leistungskontrolle zumindest technisch ermöglicht, ist hierbei insbesondere zu problematisieren. Auch sind Diskriminierungsrisiken in Betracht zu ziehen [30].⁵

Eine umfassende Bewertung der Folgen von KI in der Arbeitswelt aus ethischer Perspektive kann dabei auf die etablierte Methode des „Gute Arbeit“-Indexes zurückgreifen, die im Auftrag des DGB durch das Internationale Institut für Empirische Sozialökonomie entwickelt wurde [31]. Diese erhebt mittels elf Kriterien, die ihrerseits mit einem Katalog von einigen Dutzend Fragen verbunden sind, die Qualität von Arbeit. Der Einfluss der Einführung von KI auf diese Kriterien kann dementsprechend als Maß dienen, inwiefern sie auf das normative Leitbild Guter Arbeit „einzahlt“. In einer Aneignung der „Gute Arbeit“-Erhebungsmethode wäre somit unter anderem

⁵ Zu gegenwärtigen Ansätzen, entsprechende Risiken etwa durch Bestimmungen im Datenschutz- und Arbeitsrecht zu regulieren, vgl. den entsprechenden Beitrag von Manuela Wagner und Ann-Katrin Dreher in dieser Publikation.

danach zu fragen, welchen Einfluss die Einführung von KI auf die Einfluss- und Gestaltungsmöglichkeiten der Beschäftigten hat – einerseits was ihren Arbeitsprozess im allgemeinen anbelangt,⁶ aber auch den Prozess der Technikeinführung und die Form der Nutzung andererseits⁷.

Weitere Kriterien umfassen den Themenbereich der (Re-)Qualifikation und Entwicklungsmöglichkeiten (s.o.), Veränderungen in der Führungskultur, die Wahrnehmung der Sinnhaftigkeit der eigenen Arbeit, Arbeitszeitautonomie, soziale, körperliche und emotionale Anforderungen, Arbeitsintensität, Frustrationen die sich durch erhöhte Komplexität des Arbeitsprozesses ergeben können, Einkommen und Beschäftigungssicherheit (ebenfalls s.o.) [31, 17ff.]. Die Beantwortung dieser Fragen ist dabei allerdings nicht auf den Anwendungskontext der KI zu beschränken, sondern auch auf die Arbeitsbedingungen in der KI-Entwicklung selbst zu beziehen, insbesondere wenn es zum Outsourcing einzelner Arbeitsschritte kommt. Gleichzeitig können Überwachungs-, Manipulations- und Diskriminierungsrisiken nicht nur Beschäftigte in der

⁶ Kann die Arbeit selbstständig geplant und eingeteilt werden? Haben die Beschäftigten Einfluss auf die Arbeitsmenge? Wie können sie ihre Arbeitszeit gestalten? Ist klar erkennlich, wer wann für was Verantwortung trägt?

⁷ In welcher Form werden die Bedürfnisse der Beschäftigten bei der Entwicklung beachtet? Wie können sie selbst Einfluss nehmen? Kommt es zu einer Subsumption unter das technische System, die beispielsweise eine Steigerung der Monotonie der Arbeit zur Folge hat?

Arbeitswelt sondern beispielsweise auch die Endkunden betreffen, die entsprechende technische Systeme nutzen bzw. von deren Nutzung betroffen sind.

Aber auch in ökologischer Hinsicht lassen sich eine Vielzahl von Bewertungsdimensionen für den nachhaltigen Einsatz von KI identifizieren [vgl. 4] und deren Relevanz über den unmittelbaren Arbeitskontext hinausgeht. Diese sollen an dieser Stelle allerdings nur schlagwortartig angerissen werden:

- Spielt der Energieverbrauch bei der Auswahl des Modells eine Rolle? Werden entsprechende eventuell weniger komplexe Systeme gegenüber möglicherweise avancierteren, aber energieintensiver arbeitenden Alternativen? Werden Schritte zur Messung und Verbesserung der Energieeffizienz unternommen?
- Wie hoch ist die Emissionsintensität der benutzenden Stromquelle? Wird auf erneuerbare Energien gesetzt?
- Dient die Anwendung womöglich explizit nachhaltigen Anwendungszwecken, wie z.B. Ressourceneinsparungen? Falls ja: in welchem Verhältnis stehen diese zu dem ökologischen Impact der Soft- und Hardware?

- Auf welche Hardware wird zurückgegriffen? Gibt es Überlegungen, die ökologischen Belastungen über den gesamten Nutzungszyklus der Hardware zu senken?

Die Herausforderung im weiteren Projektverlauf wird sein, einerseits weiter für die ethischen Implikationen der Entwicklung und Nutzung von KI in der Arbeitswelt im Kontext des Projektverbundes zu sensibilisieren, die ja wenigstens zum Teil auch rechtlichen Niederschlag gefunden haben (vgl. den Beitrag von Manuela Wagner und Ann-Katrin Dreher) – andererseits aber auch die Möglichkeit auszuloten, grundsätzliche ethische Erwägungen zu „operationalisieren“, etwa auf der Ebene des betrieblichen Innovationsregimes (vgl. den Beitrag von Welf Schröter in dieser Publikation) und durch eine entsprechende Weiterentwicklung bestehender Vorgehensmodelle in der IT-Entwicklung (vgl. den Beitrag von Sascha Alpers).

References

- [1] A. Grunwald, *Technikfolgenabschätzung - eine Einführung: Zweite, grundlegend überarbeitete und wesentlich erweiterte Auflage*, 1st ed. Baden-Baden: Nomos Verlagsgesellschaft mbH & Co. KG, 2010. [Online]. Available: <http://nbn-resolving.org/urn:nbn:de:bsz:31-epflicht-1201792>
- [2] S. Böschen, A. Grunwald, B.-J. Krings, and C. Rösch, Eds., *Technikfolgenabschätzung: Handbuch für Wissenschaft und Praxis*, 1st ed. Baden-Baden: Nomos Verlagsgesellschaft mbH & Co. KG, 2021. [Online]. Available: <http://nbn-resolving.org/urn:nbn:de:bsz:31-epflicht-1861819>
- [3] KARL. [Online]. Available: <https://www.h-ka.de/ilin/projekte/karl> (accessed: Jun. 7 2022).
- [4] F. Rohde et al., "Nachhaltigkeitskriterien für künstliche Intelligenz: Entwicklung eines Kriterien- und Indikatorensets für die Nachhaltigkeitsbewertung von KI-Systemen entlang des Lebenszyklus," Schriftenreihe des IÖW 220/21, 2021.
- [5] S. Spiekermann, *Digitale Ethik: Ein Wertesystem für das 21. Jahrhundert*. München: Droemer, 2019. [Online]. Available: <https://www.perlentaucher.de/buch/sarah-spiekermann/digitale-ethik.html>
- [6] M. Hengstschläger, Ed., *Digitaler Wandel und Ethik*, 1st ed. Salzburg, München: Ecwin, 2020.
- [7] A. Grunwald and R. Hillerbrand, Eds., *Handbuch Technikethik*, 2nd ed. Berlin, Heidelberg: J. B. Metzler Verlag, 2021. [Online]. Available: <http://www.springer.com/>
- [8] J. Nida-Rümelin and N. Weidenfeld, *Digitaler Humanismus: Eine Ethik für das Zeitalter der künstlichen Intelligenz*. München: Piper, 2018.

- [9] M. Göpel, *Unsere Welt neu denken: Eine Einladung*, 16th ed. Berlin: Ullstein, 2021.
- [10] P. G. Kirchschläger, *Digital transformation and ethics: Ethical considerations on the robotization and automation of society and the economy and the use of artificial intelligence*, 1st ed. Baden-Baden: Nomos, 2021.
- [11] A. Bogner, "Politisierung, Demokratisierung, Pragmatisierung. Paradigmen der Technikfolgenabschätzung im Wandel der Zeit," in *Technikfolgenabschätzung: Handbuch für Wissenschaft und Praxis*, S. Bösch, A. Grunwald, B.-J. Krings, and C. Rösch, Eds., 1st ed., Baden-Baden: Nomos Verlagsgesellschaft mbH & Co. KG, 2021, pp. 43–58.
- [12] P. Frey, C. Schneider, and C. Wadehul, "Demokratisierung von Technik ohne Wirtschaftsdemokratie?," *TATuP*, vol. 29, no. 3, pp. 30–35, 2020, doi: 10.14512/tatup.29.3.30.
- [13] O. Mitscherlich-Schönherr, "Plädoyer für eine praktische Vernunft," *Frankfurter Rundschau*, 27./28. November, pp. 38–39, 27./28. November.
- [14] B.-J. Krings, "Technik und Arbeit," in *Handbuch Technikethik*, A. Grunwald and R. Hillerbrand, Eds., 2nd ed., Berlin, Heidelberg: J. B. Metzler Verlag, 2021, pp. 273–278.
- [15] O. Negt, *Arbeit und menschliche Würde*. Göttingen: Steidl, 2001.
- [16] J. Nida-Rümelin, "Zu menschlichen Handlungs- und Verantwortungsbegriff," in *#VerantwortungKI - Künstliche Intelligenz und gesellschaftliche Folgen*, 4/2021, *Verantwortungsvoller Einsatz von KI? Mit menschlicher Kompetenz!*, Berlin-Brandenburgische Akademie der Wissenschaften, Ed., Berlin: Berlin-Brandenburgische Akademie der Wissenschaften, 2021, pp. 21–22.

- [17] J. Peters, "Fairness und Diskriminierung in daten-basierten Entscheidungssystemen," in *#VerantwortungKI - Künstliche Intelligenz und gesellschaftliche Folgen*, 4/2021, *Verantwortungsvoller Einsatz von KI? Mit menschlicher Kompetenz!*, Berlin-Brandenburgische Akademie der Wissenschaften, Ed., Berlin: Berlin-Brandenburgische Akademie der Wissenschaften, 2021, pp. 37-41.
- [18] I. Hermann and G. Stock, "KI und Herausforderungen für menschliche Kompetenz," in *#VerantwortungKI - Künstliche Intelligenz und gesellschaftliche Folgen*, 4/2021, *Verantwortungsvoller Einsatz von KI? Mit menschlicher Kompetenz!*, Berlin-Brandenburgische Akademie der Wissenschaften, Ed., Berlin: Berlin-Brandenburgische Akademie der Wissenschaften, 2021, pp. 51-55.
- [19] European Commission. Directorate General for Communications Networks, Content and Technology. and Hochrangige Expertengruppe für künstliche Intelligenz., *Ethikleitlinien für eine vertrauenswürdige KI*: Publications Office, 2019.
- [20] M. I. Wolter et al., "Economy 4.0 and its labour market and economic impacts: Scenario calculations in line with the BIBB-IAB qualification and occupational field projections," Institut für Arbeitsmarkt-und Berufsforschung (IAB), Nürnberg, IAB research report 13/2016, 2016.
- [21] C. Frey and M. A. Osborne, "The future of employment: How susceptible are jobs to computerisation?," *Technological Forecasting and Social Change*, vol. 114, pp. 254-280, 2017, doi: 10.1016/j.techfore.2016.08.019.
- [22] P. Frey, "Automatisierung - (Alb-)Traum der emanzipatorischen Linken?: Zu den gesellschaftspolitischen Implikationen avancierter Produktivkräfte," in *Kybernetik*,

- Kapitalismus, Revolutionen: Emanzipatorische Perspektiven im technologischen Wandel*, P. Buckermann, A. Koppenburger, and S. Schaupp, Eds., 1st ed., Münster: Unrast, 2017, pp. 109–123.
- [23] P. Frey, "Visions of Automation: A Comparative Discussion of Two Approaches," *Societies*, vol. 11, no. 2, 2021, doi: 10.3390/soc11020063.
- [24] S. Larsson and F. Heintz, "Transparency in artificial intelligence," *Internet Policy Review*, vol. 9, no. 2, 2020, doi: 10.14763/2020.2.1469.
- [25] O. Nachtwey and T. Seidl, *Die Ethik der Solution und der Geist des digitalen Kapitalismus*: Institut für Sozialforschung an der Johann Wolfgang Goethe-Universität, 2017.
- [26] V. Hauff, Ed., *Unsere gemeinsame Zukunft*. Greven: Eggkamp, 1987.
- [27] T. Harbecke and G. Möhge, "Digitalisierungsstrategien im Portrait: Wie gestaltet die betriebliche Mitbestimmung digitale Transformationsprozesse im Unternehmen?," *Mitbestimmungspraxis* 34, 2020.
- [28] S. Schaupp, *Technopolitik von unten: Algorithmische Arbeitssteuerung und kybernetische Proletarisierung*. Berlin: Matthes & Seitz, 2021. [Online]. Available: <https://www.deutschlandfunk.de/simon-schaupp-technopolitik-von-unten-100.html>
- [29] M. C. Nussbaum and A. Sen, Eds., *The quality of life: A study prepared for the World Institute for Development Economics Research (WIDER) of the United Nations University ; [papers derive from a conference that took place at the WIDER in Helsinki in July 1988]*. Oxford: Clarendon Press, 2009.

- [30] A. Kolleck and C. Orwat, "Mögliche Diskriminierung durch algorithmische Entscheidungssysteme und maschinelles Lernen – ein Überblick," 2020.
- [31] M. Holler, Krüger, Thomas, P. Kulemann, F. Mußmann, and P. Sopp, "Methodenbericht zur Weiterentwicklung des DGB-Index Gute Arbeit in der Erhebungsperiode 2011/2012," Stadtbergen, 2013.

KARL – Rechtliche Aspekte

Sowohl im Rahmen der Erforschung und Entwicklung eines KI-Systems als auch bei der Einführung eines solchen im Betrieb sollten rechtliche Aspekte nicht vernachlässigt werden. Welche Rechtsfragen tangiert sind, entscheidet sich vornehmlich nach dem Einsatzkontext sowie betroffenen Daten und potentiellen Folgen und Risiken.

Anknüpfungspunkt	Rechtsgebiet / Gesetze	Zuständige Rollen (sofern vorhanden)
Verarbeitung personenbezogener Daten	Datenschutzrecht (DSGVO, BDSG, LDSG, etc.)	Datenschutzbeauftragte/r, ggf. Betriebsrat
Telekommunikation / Telemedien	TTDSG, TKG	Datenschutzbeauftragte/r
Bestimmte Angelegenheiten Arbeitsorganisation	Arbeitsrecht (BetrVG)	Betriebsrat
Geistiges Eigentum (Texte, Bilder, Datenbanken,...)	Urheberrecht (UrhG)	Rechtsabteilung / Compliance-Beauftragter
Geschäftsgeheimnisse	GeschGehG	Geheimnisbeauftragte/r
Materielle & immaterielle Schäden durch System	Haftungsrecht (BGB, ProdHaftG, etc.)	Rechtsabteilung / Compliance-Beauftragte/r

KI-System

Künftig: KI-Verordnung (Entwurf der EU-Kommission)

Für die Entwicklung von **standardisierten Vorgehensmodellen** für Software-Entwicklungsprozesse, wie sie in KARL diskutiert werden, sollte danach differenziert werden, welche Rollen sowie ggf. intern vorhandene oder extern hinzuzuziehende Expertisen in welchen Stadien und zu welchen Arbeitsschritten beratend hinzugezogen werden sollten oder zwingend eingebunden werden müssen.

Datenschutzrecht

Die datenschutzkonforme Verarbeitung personenbezogener Daten bei der Planung, Entwicklung, Umsetzung und im Betrieb eines KI-Systems bedarf der umsichtigen und vorausschauenden Organisation, indem datenschutzrechtliche Anforderungen bei der Systementwicklung von Anfang an mitgedacht werden und relevante Fachverantwortliche, wie der/die Datenschutzbeauftragte und ggf. der Betriebsrat, von Beginn an beteiligt werden.

Datenschutzrechtliche Pflichten greifen bei der Verarbeitung **personenbezogener Daten** [1]. Eine erste Weichenstellung bei der Entwicklung KI-basierter Systeme ist die Feststellung des Personenbezugs durch Analyse der Art der Daten und der geplanten Datenflüsse sowie die Entscheidung, ob dieser Personenbezug für die Zielerreichung erforderlich bzw. unvermeidlich ist oder ob über eine **Anonymisierung** die „Flucht aus dem Datenschutzrecht“ angetreten werden soll. Gerade bei der Verarbeitung großer Datenmengen aus unterschiedlichen

Quellen sind die Risiken einer Identifizierbarkeit regelmäßig zu (re-)evaluieren, da mit zunehmenden technischen Möglichkeiten auch die Zuordnung von Daten zu einer identifizierbaren Person steigen können [2].

Im nächsten Schritt steht die Identifizierung des anwendbaren Rechts. Mit der Datenschutz-Grundverordnung (DSGVO) wurde das Datenschutzrecht weitgehend EU-weit harmonisiert, allerdings sind aufgrund von Öffnungsklauseln und Übergangsregeln in bestimmten Fällen Vorschriften des Bundesdatenschutzgesetzes (BDSG), der Landesdatenschutzgesetze oder des Telekommunikations- und Telemedien-Datenschutzgesetzes (TTDSG) anwendbar [3], [4]. Bei Letzterem gilt zu bedenken, dass dieses auch nicht-personenbezogene Daten einbezieht. Datenschutzrechtliche Vorschriften können sich auch aus dem jeweils einschlägigen Fachrecht ergeben, wie bspw. dem Messstellenbetriebsgesetz (MsbG) im Bereich der Energieversorgung.

Für eine strukturierte Umsetzung der vielschichtigen Vorgaben bietet sich die Systematisierung anhand der zentralen **Datenschutzgrundsätze** an: Rechtmäßigkeit, Transparenz, Zweckbindung, Datenminimierung, Richtigkeit, Speicherbegrenzung, Integrität und Vertraulichkeit sowie Rechenschaftspflicht. Das Fundament bildet dabei die Festlegung eines Verarbeitungszwecks, an dem sich die Erforderlichkeit der Datenverarbeitung messen muss [5]. Die Beschränkung auf erforderliche Daten ergibt sich schon aus dem Datenminimierungsgrundsatz. Bereits bei der Gestaltung von technischen Systemen sollte die Verarbeitung von Daten mit Personenbezug, soweit technisch möglich und wirtschaftlich zumutbar begrenzt oder ganz vermieden werden.

Aufgrund des der DSGVO zugrundeliegenden risikobasierten Ansatzes, insbesondere im Hinblick auf die Konzepte Privacy by Design und by Default [6], Datensicherheit sowie der vom Risiko abhängigen Notwendigkeit eine Datenschutz-Folgenabschätzung [7] durchzuführen, kann die datenschutzrechtliche Bewertung bereits im Rahmen der Entwicklung Anpassungen des KI-Systems im Hinblick auf **technische und organisatorische Maßnahmen** erforderlich machen. Insofern bietet sich ein iterativer Prozess der Konkretisierung rechtlicher Anforderungen als nicht-funktionale Anforderungen, sowie die regelmäßige Re-Evaluierung der Maßnahmen an.

Im Forschungskontext können Forschungsprojekte wie KARL von einigen Privilegien profitieren: so bestehen gewisse Spielräume u.a. im Rahmen der Einwilligung, Zweckbindung, Transparenz und Speicherbegrenzung. Allerdings ist die Reichweite dieser Regelungen teils noch umstritten [8]-[10].

Arbeitsrecht

Vom Arbeitgeber sind neben der Einhaltung der speziellen Datenschutzvorschriften im Beschäftigungskontext auch die **Mitbestimmungsrechte** nach Betriebsverfassungsgesetz (BetrVG) zu beachten. Hierbei kann es sich anbieten, eine **Betriebsvereinbarung** nach Art. 88 Abs. 2 DSGVO iVm § 26 Abs. 4 BDSG zu schließen, um eine Rechtsgrundlage für die Verarbeitung personenbezogener Daten zu schaffen [11]. Der Betriebsrat hat gemäß § 80 Abs. 1 Nr. 1 BetrVG zu überwachen, dass die Vorschriften der DSGVO und des BDSG eingehalten werden, woraus sich für den Betriebsrat aus § 80 Abs. 2 BetrVG ein allgemeiner **Informationsanspruch** ergibt.

Betrifft ein geplanter KI-Einsatz unterschiedliche Aspekte der Arbeitsorganisation, können **Unterrichtungs-, Beratungs- oder Mitbestimmungsrechte** des Betriebsrats eröffnet sein. Mitbestimmungspflichtige Aspekte sind u.a. in § 87 BetrVG gelistet: (1) Fragen der Ordnung des Betriebs und des Verhaltens der Arbeitnehmer im Betrieb; (2) Arbeitszeit / Verteilung der Arbeitszeit; (3) vorübergehende Verkürzung oder Verlängerung der betriebsüblichen Arbeitszeit; (4) Zeit, Ort und Art der Auszahlung der Arbeitsentgelte; (5) allgemeine Urlaubsgrundsätze / Urlaubsplans; (6) technische Einrichtungen, die objektiv geeignet sind, das Verhalten oder die Leistung der Beschäftigten zu überwachen; (7) Verhütung von Arbeitsunfällen und Berufskrankheiten / Gesundheitsschutz; (8) Sozialeinrichtungen; (9) Betriebswohnungen; (10) Lohngestaltung; (11) Akkord- und Prämiensätze / leistungsbezogener Entgelte; (12) betriebliche Vorschlagswesen; (13) Gruppenarbeit; (14) mobile Arbeit. Der Aspekt der Künstlichen Intelligenz wurde jüngst klarstellend bezüglich der Unterrichts- und Beratungsrechte des Betriebsrats über die Planung von Arbeitsverfahren und Arbeitsabläufen in § 90 BetrVG sowie bezüglich der Mitbestimmung bei Auswahlrichtlinien in § 95 BetrVG aufgenommen. Somit wird sichergestellt, dass die Pflichten der Arbeitgeber und Rechte des Betriebsrats weiterhin gelten, auch wenn KI zum Einsatz kommt [12], [13]. Weiter besteht ein Mitbestimmungsrecht des Betriebsrates gemäß § 94 BetrVG bei Personalfragebögen und Beurteilungsgrundsätzen in Verbindung mit einem KI-System. [13] Unterrichts- und Beratungsrechte bestehen zudem bei Einführung grundlegend neuer Arbeitsmethoden (§ 111 BetrVG).

Damit der Betriebsrat seine Beratungs- und Mitbestimmungsrechte tatsächlich wahrnehmen kann, sollte die Einbindung rechtzeitig vor einem geplanten KI-Einsatz erfolgen sowie

Unterlagen zur Funktionsweise umfassen [13]. Muss der Betriebsrat zur Durchführung seiner Aufgaben die Einführung oder Anwendung von Künstlicher Intelligenz beurteilen, gilt insoweit die Hinzuziehung eines Sachverständigen als erforderlich. Eine Definition des Begriffs „Künstliche Intelligenz“ enthält das BetrVG allerdings nicht.[12], [13]

Sollen KARL-Sachverhalte bereits in realen Umgebungen erprobt werden oder Leitlinien und Hilfestellungen für einen späteren Praxiseinsatz bieten, sollten Abstimmungsprozesse von Beginn an mitgedacht werden. Wie diese praktisch umgesetzt werden könnten, wird im Abschnitt zum „**Moderierten Spezifikationsdialog**“ als Instrument antizipierender Technikeinführung beschrieben.

Urheberrecht

Je nachdem welche Datenquellen zum Training von KI-Systemen herangezogen werden sollen, können Urheber- und Leistungsschutzrechte tangiert sein. Im Folgenden werden typische Beispiele beschrieben:

- **Textdokumente** sind sofern es sich um persönlich geistige Schöpfungen handelt urheberrechtlich geschützt, ausgenommen sind amtliche Werke, wie Gesetze, Verordnungen und Bekanntmachungen. Dem/der Schöpfer:in des Werkes stehen bestimmte Verwertungsrechte zu, an denen Nutzungsrechte eingeräumt werden können. Das Urheberrecht erlischt siebenzig Jahre nach dem Tode des/der Urheber:in.
- **Bilder und Videos** können als geistige Schöpfung wie Textwerke urheberrechtlichen Schutz genießen.

Daneben sind sie als Lichtbilder⁸ und Erzeugnisse, die ähnlich wie Lichtbilder hergestellt werden, für 50 Jahre geschützt, wobei das Recht dem/der Lichtbildner:in zukommt.

- **Datenbanken** werden bei individueller Schöpfungshöhe bezüglich Auswahl oder Anordnung der Elemente als Datenbankwerke und damit selbstständiges Werk geschützt. Der/die Datenbankhersteller:in ist darüber hinaus geschützt, wenn die Elemente der Datenbank systematisch oder methodisch angeordnet und einzeln mit Hilfe elektronischer Mittel oder auf andere Weise zugänglich sind und deren Beschaffung, Überprüfung oder Darstellung eine nach Art oder Umfang wesentliche Investition erfordert.
- **Computerprogramme** genießen bei individuell-geistiger Leistung ebenfalls Schutz. Dies gilt allerdings nicht für dem Programm zugrunde liegende Ideen und Grundsätze. Die zustimmungsbedürftigen Handlungen und Ausnahmen sind gesondert, abweichend von den anderen Werkkategorien geregelt.

Bestimmte Nutzungsarten werden gesetzlich erlaubt, hierzu zählen Nutzung für Wissenschaft und Lehre wie auch kommerzielle Nutzungen in gesetzlich definierten Fällen (Schranken des Urheberrechts). Im KARL-Kontext relevant ist die gesetzlich erlaubte Nutzung zum **Text und Data Mining (TDM)**, wobei zwischen Forschungszwecken und sonstigen Zwecken unterschieden wird. TDM ist die automatisierte Analyse von einzelnen oder mehreren digitalen oder digitalisierten

⁸ Unter Lichtbildern werden Abbildungen verstanden, die „dadurch entstehen, dass strahlungsempfindliche Schichten chemisch oder physikalisch durch Strahlung eine Veränderung erfahren“ [14]§ 72 Rn. 11

Werken, um daraus Informationen insbesondere über Muster, Trends und Korrelationen zu gewinnen. [15]

Text und Data Mining für Zwecke der wissenschaftlichen Forschung: die Norm berechtigt zunächst Forschungsorganisationen Werke auch ohne Zustimmung zu vervielfältigen, sofern sie nicht kommerzielle Zwecke verfolgen, sämtliche Gewinne in die Forschung reinvestieren oder im Rahmen eines staatlich anerkannten Auftrags im öffentlichen Interesse tätig sind. Für die Kooperation mit Dritten, wie privaten Unternehmen, dürfen in bestimmten Fällen Werke zugänglich gemacht werden. Zudem sollen diese sicher verwahrt werden.[16]

Text und Data Mining: in sonstigen Fällen sind Vervielfältigungen im Rahmen des TDM zulässig bei rechtmäßig zugänglichen Werken. Der/die Rechtsinhaber:in kann sich allerdings vorbehalten, dass diese Nutzungen ausgeschlossen sind. Bei online zugänglichen Werken muss dieser Nutzungsvorbehalt in maschinenlesbarer Form erfolgen.

Bei Open-Source-Projekten bzw. als Open Data bereitgestellten Daten werden zumeist über die Lizenzen Rechte zur Vervielfältigung und Bearbeitung gewährt. Diese können allerdings mit unterschiedlichen Bedingungen versehen sein, insbesondere die unter Nutzung der Lizenzen entstandenen Werke ebenfalls zur freien Nutzung zur Verfügung zu stellen (sog. Copy-Left).

Schutz von Geschäftsgeheimnissen

Daten können neben als auch ohne Personenbezug auch bei Unternehmensbezug rechtlichen Schutz genießen. Dies ist der Fall, wenn diese Daten nicht allgemein bekannt oder nicht ohne Weiteres zugänglich sind und daher von

wirtschaftlichem Wert sind, die Gegenstand von den Umständen nach angemessenen Geheimhaltungsmaßnahmen durch ihren rechtmäßigen Inhaber sind und bei der ein berechtigtes Interesse an der Geheimhaltung besteht [17], [18]. Das Gesetz unterscheidet zwischen explizit erlaubten Handlungen, Handlungsverboten und Ausnahmen. Zu den erlaubten Handlungen zählt mittlerweile das Reverse Engineering. Ausnahmen gelten insbesondere für Whistleblower [19].

Zu Überlappungen mit dem Datenschutzrecht kommt es, wenn Geschäftsgeheimnisse ebenfalls über Personenbezug verfügen. Organisatorische und technische Schutzmaßnahmen erfüllen dann eine Doppelfunktion, da Geschäftsgeheimnisse per Definition **angemessene Geheimhaltungsmaßnahmen** erfordern [20]. Während das Datenschutzrecht zwingende Vorgaben enthält, erfolgt der Schutz von Geschäftsgeheimnissen zumeist im Eigeninteresse, um wertvolles Know-How des Unternehmens nicht nach außen zu offenbaren. Daneben können vertragliche Verpflichtungen den Schutz der Geschäftsgeheimnisse Dritter erforderlich machen.

Sonstige Haftungsfragen

Sollen KI-Systeme in der Praxis eingesetzt werden bzw. für den Praxiseinsatz vertrieben werden, sollten sich Entwickler:innen und Anbieter:innen stets potentiellen Haftungsrisiken bewusst sein. Zudem wurden mit der Novelle des Verbraucherschutzrechts Herstellerpflichten zur **Aktualisierung**, d.h. Bereitstellung von Software-Updates, etabliert [21], [22]. Die folgende Tabelle zeigt, in welchen Situationen bestimmte Haftungsfragen näher geprüft werden sollten.

Bereich	Voraussetzungen	Haftungsrisiken
Vertragliche Haftung (B2B)	Verträge über die Bereitstellung von Software können insbes. unter das Kauf-, Miet-, Werk- oder Dienstvertragsrecht fallen - oftmals liegen typen- gemischte Verträge vor.	Haftung für Mängel beim Kaufvertrag: Abweichung von subjektiven oder objektiven Anforderungen: - Vereinbarte Beschaffenheit über Art, Menge, Qualität, Funktionalität, Kompatibilität, Interoperabilität - Eignung nach der im Vertrag vorausgesetzten Verwendung - Erwartbares / vereinbartes Zubehör / Anleitung vorhanden & korrekt - Eignung für gewöhnliche Verwendung - Übliche Beschaffenheit, insbes. Haltbarkeit, Funktionalität, Kompatibilität und Sicherheit - Entsprechung Muster / Testversion
Verbraucher- verträge (B2C): Verbraucher sind natürliche Personen, die Rechtsgeschäfte zu privaten Zwecken abschließen	Digitale Produkte: - digitale Inhalte: Daten, die in digitaler Form erstellt und bereitgestellt werden. - digitale Dienstleistungen: die dem Verbraucher Verarbeitung von Daten in digitaler Form oder den Zugang zu solchen Daten ermöglichen, oder die gemeinsame Nutzung in digitaler Form	- Haftung für Mängel bei Abweichung von objektiven oder subjektiven Anforderungen (s.o.) - Aktualisierungspflichten: für einen festgelegten Bereitstellungszeitraum oder vernünftigerweise erwartbaren Zeitraum (Ausnahme: Verbraucher unterlässt Installation)

	hochgeladenen / erstellten Daten oder sonstige In- teraktionen mit diesen Daten er- möglichen.	
	Waren mit digitalen Elementen: Sachen, die ein digitales Produkt enthalten ohne welches sie ihre Funktion nicht erfüllen	Haftung für Mängel bei Abweichung von objektiven oder subjektiven Anforderungen (s.o.) sowie Nichtbereitstellung Aktualisierungen (s.o.)
Produkthaftung	Hersteller haftet verschuldensunabhängig für von Produktfehlern verursachte Schäden (nicht Schäden am Produkt selbst)	Jemand wird durch den Fehler eines Produkts getötet, sein Körper oder seine Gesundheit verletzt oder eine Sache beschädigt, die für den privaten Ge- oder Verbrauch bestimmt & genutzt wird
Produzentenhaftung	Hersteller haftet für Konstruktions-, Fabrikations- und Instruktionsfehler sowie die Verletzung der Produktbeobachtungspflicht , es sei denn es gelingt der Nachweis fehlenden Verschuldens	Haftung für vorsätzliche oder fahrlässige Verletzung von Leben, Körper, Gesundheit, Freiheit, Eigentum oder ein sonstiges Recht eines anderen. Zu sonstigen Rechten zählt u.a. das allgemeine Persönlichkeitsrecht.

In bestimmten, besonderen Konstellationen können sich Haftungsfragen aus weiteren Regelungen ergeben, wie bspw. dem Straßenverkehrsrecht bei KI-Systemen im Rahmen des automatisierten Fahrens (Halterhaftung, Fahrerhaftung). Einige Normen, bspw. aus dem Strafrecht, wirken darüber

hinaus als Schutzgesetze, deren Verletzung zivilrechtlich verfolgt werden kann (bspw. § 303a StGB Datenveränderung).

Künftige KI-Regulierung

Künstliche Intelligenz soll nach Ansicht der EU-Kommission künftig auf EU-Ebene rechtlich reguliert werden. Allerdings betreffen nicht alle Anforderungen sämtliche KI-Systeme gleichermaßen, sondern entscheidend ist das damit verbundene Risiko für die Grundrechte potentiell betroffener Personengruppen. Insofern werden KI-Systeme im aktuellen Entwurf⁹ eingeteilt in:

- **Verbotene Praktiken:** KI-Systeme mit Schadenspotential und unbewusst manipulierendem Charakter oder Ausnutzung der Schutzbedürftigkeit von Personen, mithin einem unannehmbaren Risiko sowie biometrische Echtzeit-Fernidentifizierungssysteme
- **Hochrisiko KI-Systeme:** Systeme, die unter bestimmte Rechtsvorschriften der EU fallende Produkte oder Sicherheitskomponenten in solchen Produkten sind und eine Konformitätsbewertung erfordern oder im Bereich biometrische Identifizierung / Kategorisierung, Kritische Infrastrukturen, allgemeine / berufliche Bildung, Beschäftigung, Personalmanagement / Selbstständigkeit, Zugang zu privaten / öffentlichen Diensten / Leistungen, Strafverfolgung, Migration, Asyl und Grenzkontrolle,

⁹ Vorschlag für eine Verordnung des Europäischen Parlaments und des Rates zur Festlegung harmonisierter Vorschriften für Künstliche Intelligenz (Gesetz über Künstliche Intelligenz) und zur Änderung bestimmter Rechtsakte der Union, COM/2021/206 final vom 21.4.2021.

Rechtspflege und demokratische Prozesse eingesetzt werden.

- **KI-Systeme:** werden im Entwurf der EU zur KI-Verordnung definiert als: „eine Software, die mit einer oder mehreren der in Anhang I [der KI-Verordnung] aufgeführten Techniken und Konzepte entwickelt worden ist und im Hinblick auf eine Reihe von Zielen, die vom Menschen festgelegt werden, Ergebnisse wie Inhalte, Vorhersagen, Empfehlungen oder Entscheidungen hervorbringen kann, die das Umfeld beeinflussen, mit dem sie interagieren“ Anhang I des Entwurfs nennt wiederum zahlreiche Techniken und Konzepte, wie die des maschinellen Lernens, Logik- und wissensgestützte Konzepte als auch statistische Ansätze. Insbesondere die Weite des gewählten Rahmens unter Verweis auf einen Anhang hat vielfache Kritik provoziert [23]–[25].

Für Hochrisiko-KI-Systeme sind zahlreiche Schutzmechanismen vorgesehen, wie ein Risikomanagementsystem, Qualitätsmanagementsystem, menschliche Aufsicht, Transparenz, technische Dokumentation und Aufzeichnungspflichten, Daten & Daten Governance, technische und organisatorische Maßnahmen zu Genauigkeit, Robustheit & Cybersicherheit sowie Konformitätserklärung /-bescheinigung & -kennzeichnung und Registrierungen. Für gewöhnliche KI-Systeme gelten lediglich Transparenzregeln, sodass ersichtlich sein muss, dass es sich um eine KI und keinen Menschen handelt (bspw. bei Chatbots, Deepfakes, etc.).

Gegenstand andauernder Diskussionen ist die Frage, ob mit diesem Entwurf die im Abschnitt „**Warum Ethik in KARL**“ aufgeworfenen sozialen, ethischen und gesellschaftspolitischen

Herausforderungen einer ausgewogenen rechtlichen Regulierung unterworfen werden, um rechtliche Impulse hin zur Entwicklung menschenzentrierter, transparenter, lernförderlicher *und* nachhaltigen KI-Systeme zu setzen. Wichtige Kritikpunkte sind die geringe Adressierung der von den Risiken betroffenen Personen sowie die Verwendung zahlreicher unbestimmter Begriffe, die einen weiten Interpretationsspielraum belassen und damit mit der Befürchtung von Rechtsunsicherheit einhergehen [23]–[27]. So wird deutlich, dass es keine generalisierte Gestaltung von KI geben kann, sondern stets die Dimensionen des Einsatzkontextes, Bandbreiten der Einsatzzwecke, ökologische Implikationen und Beziehungen der betroffenen Personenkreise in die Betrachtung einbezogen werden sollten.

Quellen

- [1] Artikel-29-Datenschutzgruppe, „Stellungnahme 4/2007 zum Begriff ‚personenbezogene Daten‘ - WP 136“, Artikel-29-Datenschutzgruppe, Brüssel, Juni 2007. Zugriffen: 31. August 2018. [Online]. Verfügbar unter: https://www.lida.bayern.de/media/wp136_de.pdf
- [2] Artikel-29-Datenschutzgruppe, „Stellungnahme 5/2014 zu Anonymisierungstechniken - WP 216“, Brüssel, Apr. 2014. Zugriffen: 10. November 2018. [Online]. Verfügbar unter: https://ec.europa.eu/justice/article-29/documentation/opinion-recommendation/files/2014/wp216_de.pdf
- [3] J. Kühling u. a., *Die Datenschutz-Grundverordnung und das nationale Recht: erste Überlegungen zum innerstaatlichen Regelungsbedarf*. Münster: Verlagshaus Monsenstein und Vannerdat, 2016.

- [4] DSK - Datenschutzkonferenz, „Orientierungshilfe der Aufsichtsbehörden für Anbieter:innen von Telemedien ab dem 1. Dezember 2021 (OH Telemedien 2021)“, Dez. 2021.
- [5] Artikel-29-Datenschutzgruppe, „Opinion 03/2013 on Purpose Limitation - WP 203“, Brüssel, Apr. 2013. Zugriffen: 19. Oktober 2018. [Online]. Verfügbar unter: http://ec.europa.eu/justice/data-protection/article-29/documentation/opinion-recommendation/files/2013/wp203_en.pdf
- [6] European Data Protection Board, „Guidelines 4/2019 on Article 25 Data Protection by Design and by Default, Version 2.0“, Brüssel, Okt. 2020.
- [7] DSK - Datenschutzkonferenz, „Kurzpapier Nr. 5 Datenschutz-Folgenabschätzung nach Art. 35 DS-GVO“, Dez. 2018. Zugriffen: 20. Mai 2020. [Online]. Verfügbar unter: https://www.datenschutzkonferenz-online.de/media/kp/dsk_kpnr_5.pdf
- [8] F. Molnár-Gábor, „Die Regelung der wissenschaftlichen Forschung in der DSGVO“, *DSR/ITB*, S. 159-173, 2018.
- [9] A. Roßnagel, „Datenschutz in der Forschung“, *ZD*, S. 157-164, 2019.
- [10] T. Weichert, „Die Forschungsprivilegierung in der DSGVO“, *ZD*, S. 18-24, 2020.
- [11] J. Holthausen, „Big Data, People Analytics, KI und Gestaltung von Betriebsvereinbarungen - Grund-, arbeits- und datenschutzrechtliche An- und Herausforderungen“, *RdA*, S. 19-32, 2021.
- [12] O. Reinartz, „Das Betriebsrätemodernisierungsgesetz“, *NZA-RR*, S. 457-470, 2021.

- [13] J. Frank und M. Heine, „Künstliche Intelligenz im Betriebsverfassungsrecht“, *NZA*, S. 1448-1452, 2021.
- [14] H. Ahlberg, H.-P. Götting, und A. Lauber-Rönsberg, Hrsg., *Beck'scher Online-Kommentar zum Urheberrecht*, Bd. 34. Edition. München: C.H. Beck, 2022. Zugriffen: 11. Juni 2016. [Online]. Verfügbar unter: https://beck-online.beck.de/?vpath=bib-data/komm/BeckOK_UrhR_10/cont/BeckOK.UrhR.htm
- [15] B. Raue, „Das Urheberrecht der digitalen Wissen(schaft)s-gesellschaft“, *GRUR*, S. 14, 2017.
- [16] F. Kleinkopf, J. Jacke, und M. Gärtner, „Text- und Data-Mining“, *MMR*, S. 196-200, 2021.
- [17] M. Dann und J. Markgraf, „Das neue Gesetz zum Schutz von Geschäftsgeheimnissen“, *NJW*, S. 1774-1779, 2019.
- [18] A. Ohly, „Das neue Geschäftsgeheimnisgesetz im Überblick“, *GRUR*, S. 441-451, 2019.
- [19] C. Alexander, „Geheimnisschutz nach dem GeschGehGE und investigativer Journalismus“, *AfP*, S. 1-11, 2019.
- [20] P. Gola, „Das Geschäftsgeheimnisgesetz und die Datenschutz-Grundverordnung: Parallele Regelungen mit neuen Verpflichtungen und Aufgaben für Datenschutzbeauftragte?“, *DuD*, Bd. 43, Nr. 9, S. 569-574, Sep. 2019, doi: 10.1007/s11623-019-1165-8.
- [21] C. Felsch, J. Kremer, und F. Jacoby, „Handhabung der neuen Aktualisierungspflicht bei digitalen Produkten“, *MMR*, S. 18-23, 2022.
- [22] J. Kühner und C. Piltz, „Die Updatepflicht für Unternehmen in Umsetzung der Digitale Inhalte Richtlinie – Der

Regelungsmechanismus im Referentenentwurf des BMJV v. 3.11.2020 zur Umsetzung der Richtlinie 2019/770/EU“, *CR*, Bd. 37, Nr. 1, S. 1-7, Jan. 2021, doi: 10.9785/cr-2021-370106.

[23] A. Ebert und I. Spiecker gen. Döhmann, „Der Kommissionsentwurf für eine KI-Verordnung der EU“, *NVwZ*, Nr. 16, S. 1188-1193, 2021.

[24] M. Ebers, V. R. S. Hoch, F. Rosenkranz, H. Ruschemeier, und B. Steinrötter, „Der Entwurf für eine EU-KI-Verordnung: Richtige Richtung mit Optimierungsbedarf“, *RDi*, Bd. 3, Nr. 11, S. 528, Nov. 2021.

[25] Europäischer Wirtschafts- und Sozialausschuss und Christa Schweng, „Stellungnahme des Europäischen Wirtschafts- und Sozialausschusses zum ‚Vorschlag für eine Verordnung des Europäischen Parlaments und des Rates zur Festlegung harmonisierter Vorschriften für künstliche Intelligenz (Gesetz über künstliche Intelligenz) und zur Änderung bestimmter Rechtsakte der Union‘ (COM(2021) 206 final – 2021/106(COD))“. 22. Dezember 2021.

[26] Europäischer Datenschutzausschuss und Europäischer Datenschutzbeauftragter, „EDSA-EDSB Gemeinsame Stellungnahme 5/2021 zum Vorschlag für eine Verordnung des Europäischen Parlaments und des Rates zur Festlegung harmonisierter Vorschriften für künstliche Intelligenz (Gesetz über künstliche Intelligenz) und zur Änderung bestimmter Rechtsakte der Union“. 18. Juni 2021. Zugriffen: 30. Mai 2022. [Online]. Verfügbar unter: https://edpb.europa.eu/our-work-tools/our-documents/edpbedps-joint-opinion/edpb-edps-joint-opinion-52021-proposal_de

[27] A. Sasing und A. Tschech, „AGG und KI-VO-Entwurf beim Einsatz von Künstlicher Intelligenz“, *MMR*, S. 24-30, 2022.

Das Erfahrungswissen der Mitarbeitenden als wesentlicher Baustein für den digitalen Wandel im Unternehmen

„Moderierter Spezifikationsdialog“ als Instrument antizipierender Technikeinführung

Die Praxis der Einführung technischer Neuerungen zugunsten von betrieblichen Arbeitsprozessen hat in den zurückliegenden mehr als drei Jahrzehnten immer wieder gezeigt, dass „harte“ Technikfaktoren nur dann zu erfolgreichen Lösungen geformt werden können, wenn sie frühest möglich mit „weichen“ Faktoren kombiniert sind. Mit anderen Worten: Technische Innovationen bedürfen der sozialen Innovationen. Wer nur technikzentriert voran geht, erzeugt zumeist innovatives Strohfeuer, aber keine nachhaltige Verbesserung. Nicht der alleinige Blick auf die pure Mensch-Technik-Interaktion eröffnet ausreichend Chancen, sondern vor allem das „Lernen-in-Prozessen-zu-Denken“ schafft das Fundament des Erfolgs.

Nach mehr als dreißig Jahren Praxiserfahrung mit der Digitalisierung von Arbeitswelten hebt das Netzwerk „Forum Soziale Technikgestaltung“ (FST) mit seinen mehr als 4.700 Frauen und Männern aus Beschäftigtenvertretungen und Belegschaften die hohe Bedeutung des Erfahrungswissens der Mitarbeiterinnen und Mitarbeiter im Vorgang des betrieblichen Umbaus („digitale Transformation“) mit Nachdruck hervor. Entlang dieser Linie von FST-Impulsen[1] erfolgt seit

dreißig Jahren auch die Sammlung von Erfahrungen im Umgang mit der sogenannten „Künstlichen Intelligenz“.^[10]

Während Ende der achtziger Jahre noch umfänglich von „starker“ und „schwacher“ „KI“ gesprochen wurde, haben sich die Wahrnehmungen, die Potenziale und Anwendungsbreiten von „algorithmischen Steuerungs- und Entscheidungssystemen“ (FST) deutlich ausdifferenziert. Hinsichtlich der sogenannten „KI“ sprechen wir nicht nur einerseits von traditionell „regelbasierten“ und andererseits von neuen „datengetriebenen“ Software-Instrumenten, sondern vor allem unterscheiden wir Software-Systeme, die sich durch Anwendung verändern und solchen, die im Kern konstant bleiben. Software-Systeme, die sich durch Anwendung – also durch mathematische Datenaufnahme und Datenverarbeitung – nicht verändern, lassen sich nach bekannten Mustern sozialer Technikgestaltung und partizipativ-mitbestimmter Vereinbarung (inklusive von Experimentierräumen) einführen. „Datengetriebene“ Software-Systeme müssen jedoch zum Teil nach einem neuen Muster gestaltet und angepasst werden, denn dieser Teil der algorithmischen Werkzeuge lässt sich nach der Implementierung und nach eingehender Nutzung nicht mehr oder nur mit enormem Aufwand (auch Kostenaufwand) gestalten. Führende IT-Entwickelnde räumen ein, dass diese „sich selbstverändernden Software-Systeme“ (FST) auch von den IT-Teams nicht mehr vollständig durchschaut werden können. Die

¹⁰ Schon früh wurde unter Betriebs- und Personalräten das Thema „Künstliche Intelligenz“ diskutiert. Bei der gemeinsamen Forschungskonferenz des DGB Baden-Württemberg und der Hans-Böckler-Stiftung am 1. Juli 1989 in der „Wissenschaftsstadt Ulm“ stellte der Autor des vorliegenden Beitrages die Fragen: „Wie verändert die ‚Künstliche Intelligenz‘ unsere Zukunft? Wie können wir die ‚Künstliche Intelligenz‘ gestalten?“ Er empfahl damals: Die Konferenz soll „über die neuesten Forschungen auf dem Gebiet der ‚Künstlichen Intelligenz‘ informieren und Kriterien für eine gesellschaftlich verantwortliche Gestaltbarkeit entwickeln. Wie wird der Einsatz von KI die Zukunft der Fabrik umwälzen? Wie rationalisieren die ‚Expertensysteme‘ die Arbeit in Büro und Verwaltung? Welche Chancen und Risiken liegen in der Perspektive von KI? Dabei sollen neben Formen der maschinellen Sprachverarbeitung, der Mensch-Maschine-Kommunikation vor allem auch die Ergebnisse der modernen Gehirnforschung angesprochen werden. Die sogenannte ‚starke KI-These‘ will die Zukunftscomputer der Informationsverarbeitung des Gehirns nachbauen. ‚Neurocomputer‘ spiegeln die Nervenbahnen, die ‚neuronalen‘ Wege des Menschen.“ (Welf Schröter im Konferenzprogramm 1989)

Dokumentierbarkeit und technische Nachvollziehbarkeit der einzelnen Entscheidungsschritte ist kaum mehr gegeben.

Notwendigkeit „antizipierender präventiver Arbeitsgestaltung“

Daraus zieht das „Forum Soziale Technikgestaltung“ den Schluss, dass die soziale, ethische und rechtliche Gestaltung dieses Typs „datengetriebener“ Software vor(!) ihrer Implementierung realisiert werden muss.[2] Das FST spricht deshalb von der Notwendigkeit „antizipierender präventiver Arbeitsgestaltung“. Um diesen neuen, zusätzlichen Weg der Gestaltung operativ umsetzbar zu machen, entwickelte das FST das neue Format des „Moderierten Spezifikationsdialoges“ zwischen Beschäftigten und Beschäftigtenvertretungen auf der einen Seite und den IT-Teams auf der anderen Seite.

Die zentrale Aufgabe des „Moderierten Spezifikationsdialoges“ besteht darin, das hohe Gut des Erfahrungs- und Prozesswissens der Beschäftigten auf gleicher Augenhöhe mit dem Fach- und Sachwissen der IT-Teams in einen Spezifizierungsprozess zu transformieren. Es geht also nicht primär um technische „Requirements“ sondern um die Einbettung des erfahrungsgeleiteten „Lernens-in-Prozessen-zu-denken“[3] und um die „Über-Setzung“ von Erfahrungskennntnissen in die Entwicklungswelt.

Dabei ist es nicht die Aufgabe der „Erfahrungs“-Seite, sich in Programmiersprachen und IT-Termini auszudrücken. Das Ziel des Dialoges ist es, dass die IT-Seite die erfahrungsgeleiteten Anforderungen aus der Arbeitsprozessperspektive heraus verstehen lernt und in die IT-Sprachen übersetzt. Die Aufgabe der „Erfahrungs“-Seite besteht vor allem darin, ihre Anforderungen so präzise wie möglich zu artikulieren. Die bloße

Formel einer „sozialen KI“ ist wenig hilfreich. Wenn zum Beispiel eine Abteilung den Arbeitsschutzgedanken unmittelbar im algorithmischen Steuerungs- und Entscheidungssystem softwaretechnisch verankern will, muss die Abteilung ihre arbeitsplatz- und arbeitsprozessbezogenen Anforderung ganz präzise zur Sprache bringen, um deren Berücksichtigung zu ermöglichen.

Soziale Technikgestaltung erhält hierbei eine ergänzende Dimension, denn die Anforderungen werden nicht nur in Betriebs- oder Dienstvereinbarungen festgelegt, sondern zudem direkt im algorithmischen System verankert. IT-Teams, die diesen Denkansatz verstanden haben, reagieren sehr erfreut, da sie zwar über brillante Kenntnisse in der Technikentwicklung aber selten über Prozesskenntnisse bei den späteren Anwendenden verfügen.

Das „Forum Soziale Technikgestaltung“ fasst die dialogische Anforderungsumsetzung innerhalb der algorithmischen Steuerungs- und Entscheidungssysteme (Algorithmen, Datenzufuhr, Datenverarbeitung etc.) seit mehreren Jahren verkürzt unter das Sprach-Logo „Der mitbestimmte Algorithmus“.[4]

Die späteren Anwendenden im Betrieb bzw. die späteren Kundinnen und Kunden bestimmen zum frühest möglichen Zeitpunkt mit. Es macht dabei Sinn, dass die Moderation von externen, unabhängigen Personen durchgeführt wird. Wichtig sind dabei eine aktive Kommunikationsbereitschaft und eine offene Kommunikationskultur, um diesen Weg einer ergänzenden Interessenaushandlung zu gewährleisten.

Überlegungen innerhalb von KARL zum Thema „Moderierter Spezifikationsdialog“

Das vorgeschlagene Format von „Moderierten Spezifikationsdialogen“, deren Anwendungen über Betriebs- oder Dienstvereinbarungen gestützt sind, hat im Projekt KARL eine interessierte Diskussion ausgelöst. Gesprächsteilnehmende betonten unter anderem die Relevanz folgender Eckpunkte:

- Gegenseitige Verständlichkeit zwischen IT-Experten und Laien sicherstellen (wegen unterschiedlicher Auffassung von Fachbegriffen);
- Wertvolle Erfahrungen der Mitarbeitenden über den Arbeitsprozess wertschätzen und mitnehmen;
- Starke instrumentelle Begründung des Wertes von Spezifikationsdialogen (Probleme vorab vermeiden, Erfahrungswissen nutzen);
- Kommunikation zwischen IT-Entwicklung, Geschäftsleitung und Belegschaft stärken;
- Partizipation von Mitarbeitenden im Design über User Centred Design-Ansätzen unmittelbar am Anfang ermöglichen;
- Stärkere Verbindlichkeit von Partizipation;
- Rechtliche Bedeutung der Mitbestimmung;
- Stakeholderanalyse als wichtiges Instrument der Mitbestimmung;
- Mitbestimmung heißt Verbindlichkeit versus bloßer Beteiligung;
- Chancen der Vermeidung unnötiger Kosten;
- Notwendigkeit kontinuierlicher Reviews.

Grundsätzlich wurde das Format „Moderierter Spezifikationsdialog“ im passenden KARL-Workshop am 7. April 2022 begrüßt mit den Worten: „Ein wirklich sehr überzeugendes Konzept.“ Nun gilt es, diesen Ansatz zu einem im Betrieb operativ

einsetzbares Modell KARL-spezifisch zu verfeinern. Es bedarf eines generisch strukturierten Aufbaus, um eine skalierbare Lösung für KMUs anzubieten.

Im Workshop wurde darauf hingewiesen, dass es Sinn macht, im Prozess der Software-Entwicklung zu unterscheiden zwischen Standardentwicklung sich selbst nicht verändernder Software einerseits und komplexen algorithmischen Steuerungs- und Entscheidungssystemen andererseits, die sich durch Datenaufnahme ständig weiter wandeln können.

Aus betrieblichen Erfahrungen ist zu entnehmen, dass nicht selten sowohl Geschäftsleitungen wie auch Beschäftigtenvertretungen beim Umgang mit komplexen sich selbst verändernden Software-Werkzeugen überfordert sind. „Moderierte Spezifikationsdialoge“ können für beide Seiten bei der Kompetenzentwicklung sehr hilfreich sein.[5]

Für die Organisation von diesem Typ „Spezifikationsdialog“ sollten verschiedene Intensivierungsvarianten berücksichtigt werden:

- Dialog Beschäftigtenseite/Beschäftigtenvertretung mit IT-Seite;
- Dialog Beschäftigtenseite/Beschäftigtenvertretung mit IT-Seite und mit Geschäftsleitung;
- Dialog Beschäftigtenseite/Beschäftigtenvertretung mit IT-Seite und mit Sachverständigen; (Datenschutz, Arbeitsschutz, Gesundheitsschutz, Gefährdungsbeurteilung, Schwerbehindertenvertretung, Weiterbildung, ...);
- Gemeinsame Austausch- und Kommunikationsräume schaffen;
- Prinzipien der Agilität (z.B. SCRUM) als Methode nutzen,
- Jeweils mit externer, „neutraler“ Moderation.

Die Spezifikationsdialoge sollten dazu beitragen, in einem interessengruppenübergreifenden Aushandlungsvorgang eine gemeinsam vereinbarte Liste von Gestaltungs- und Zulassungskriterien zu erarbeiten, nach denen komplexe Software-Systeme gehandhabt werden können. Das „Forum Soziale Technikgestaltung“ hat dazu dreißig generische Kriterien[6] vorbereitet, veröffentlicht und in Diskussionsprozesse eingebracht.

Diese Kriterienliste kann als einer der Ausgangspunkte für weitere Beratungen im Projekt KARL dienen. Die Sammlung von Gestaltungs- und Zulassungskriterien konnten aus den Erfahrungen von Belegschaften und Beschäftigtenvertretungen[7] abgeleitet werden. Diese Sammlung gilt es fortzuentwickeln.

Differenzierung von Handlungsebenen

Im Diskurs zu ethischen, sozialen und rechtlichen Aspekten der Technikgestaltung und Technikeinführung im Kontext algorithmischer Steuerungs- und Entscheidungssysteme (sogenannte „Künstliche Intelligenz“) ist es erforderlich, die Kriterien nach unterschiedlichen Handlungsebenen zu modifizieren, aufzuteilen und diesen zuzuordnen. Die ethische Anforderung nach Souveränität des Individuums offenbart sich zum Beispiel auf der Handlungsebene der unmittelbaren Mensch-Technik-Interaktion anders als auf der gesellschaftspolitischen Handlungsebene. Ebenso differenziert sie sich zwischen der betriebsübergreifend arbeitsprozessorientierten Ebene entlang von Wertschöpfungsketten und den anders geregelten innerbetrieblichen Handlungsebenen.

Die jeweiligen Subjektrollen und Subjektzuweisungen – zum Beispiel gemäß des Betriebsverfassungsgesetzes – zeigt für

Betriebsräte andere – nämlich erweiterte ethische wie rechtliche – Rahmensetzungen auf, als sie für individuelle Verbraucherinnen und Verbraucher, Kundinnen und Kunden, Bürgerinnen und Bürger gegeben sind. Eine Betriebsrätin bzw. ein Betriebsrat tritt nicht nur für seinen Individualanspruch der Souveränität ein sondern sieht den kollegialen Anspruch der gesamten Belegschaft.

Vor allem diese vier Handlungsebenen[8] dominieren die Anwendungswelt der sogenannten „Künstlichen Intelligenz“:



Ziele

Es soll das Ziel gesetzt werden, für die Unternehmen und Betriebe skalierbare praktische Handreichungen für größere oder kleinere KMUs bzw. Handwerksfirmen zu erstellen, die einerseits die Einführung sogenannter „KI“-Systeme beschleunigen und andererseits berücksichtigen, dass die

Implementierung solcher Systeme grundsätzlich einen mitbestimmungspflichtigen Vorgang darstellt.

Der Einsatz algorithmischer Steuerungs- und Entscheidungssysteme (sogenannte „Künstliche Intelligenz“) beginnt in der Regel innerhalb der betrieblichen Abläufe (vertikale Wertschöpfungskette).

Letztlich einflussreich werden diese Systeme durch ihre Steuerungsfunktion in den zwischenbetrieblichen Geschäftsabläufen (Prozesse) (horizontale Wertschöpfungskette).

Algorithmische Steuerungs- und Entscheidungssysteme werden zunehmend die innerbetrieblichen Abläufe von außen beeinflussen und bestimmen.
Wir müssen unsere betriebszentrierte Denkweise öffnen.

Es gilt, bei den betrieblichen Partnerinnen und Partnern Sensibilität und Awareness zu erzeugen, Kenntnisse sowie Kompetenzen[9] zu stärken. Verständliche Handlungshilfen und Praxisbeispiele kommen dabei Geschäftsleitungen, Belegschaften, Beschäftigtenvertretungen und IT-Teams entgegen.

Der Impuls für „Moderierte Spezifikationsdialoge“ versteht sich als ein Baustein für ein gemeinsames Anwendungskonzept, in dem verbindliche, ethische, legale und soziale Anforderungen (ELSA) aus der Arbeit der Partnerinnen und Partner des Gestaltungsfeldes 5 (FZI, ITAS, FST) im BMBF-Projekt KARL zusammenwachsen.

Welf Schröter
schroeter@talheimer.de
21. Juni 2022

- [1] www.blog-zukunft-der-arbeit.de (Zugriff am: 7. Juni 2022).
- [2] W. Schröter, „Der mitbestimmte Algorithmus“ – Ein erweiternder Ansatz zur Gestaltung der sogenannten „Künstlichen Intelligenz“. Beitrag für das Online-Magazin „DENK-doch-MAL“ (2022). Der Text ist frei zugänglich und verfügbar unter: <http://denk-doch-mal.de/wp/welf-schroeter-der-mitbestimmte-algorithmus-ein-erweiternder-ansatz-zur-gestaltung-der-sogenannten-kuenstlichen-intelligenz/> (Zugriff am: 7. Juni 2022).
- [3] Projekt DigiGAAB „Digital unterstützter Gesundheits-und Arbeitsschutz im Arbeitsprozess“. Verfügbar unter: <https://digi-gaab.stuck-komzet.de/> (Zugriff am: 7. Juni 2022).
- [4] W. Schröter, Hg., *Der mitbestimmte Algorithmus. Gestaltungskompetenz für den Wandel der Arbeit*. Mössingen: Talheimer Verlag, 2019. – Siehe dazu auch: W. Schröter: Selbstbestimmung zwischen „nachholender Digitalisierung und „autonomen Software-Systemen“. Wenn Betriebsräte „vorausschauende Arbeitsgestaltung“ erproben. In: W. Schröter (Hg.): *Autonomie des Menschen – Autonomie der Systeme. Humanisierungspotenziale und Grenzen moderner Technologien*. Mössingen: Talheimer Verlag, 2017, S. 187–256.
- [5] Im Zusammenhang mit dem Jubiläum „Dreißig Jahre Forum Soziale Technikgestaltung“ (FST) fanden eine Reihe von Fachdialogen mit Fachexpertinnen und Fachexperten aus Wirtschaft, Arbeitswelt, Wissenschaft und Gesellschaft statt. Dazu gehörten die „Allianz Industrie 4.0 Baden-Württemberg“, „Cyber Valley“, ITAS, Universität FAU, Universität Hohenheim, IAT Uni Stuttgart, Fraunhofer IPA, Fraunhofer IAO, GPAI (Global Partnership on Artificial Intelligence), Hans-Böckler-Stiftung Labor.A, Fritz-Erler-Stiftung, Heinrich-Böll-Stiftung, Evang. Akademie Bad Boll, Arbeitskreis Evangelischer Unternehmer in Deutschland, Baden-Württemberg: connected, ...).

- [6] W. Schröter, „Der mitbestimmte Algorithmus. Arbeitsweltliche Kriterien zur sozialen Gestaltung von Algorithmen und algorithmischen Entscheidungssystemen“ in W. Schröter, Hg., *Der mitbestimmte Algorithmus. Gestaltungskompetenz für den Wandel der Arbeit*. Mössingen: Talheimer Verlag, 2019, S. 101-150.
- [7] Publikationen aus der Arbeit des FST: <http://www.blog-zukunft-der-arbeit.de/publikationen/>
- [8] W. Schröter, *Vertiefung des Ansatzes „Der mitbestimmte Algorithmus“ - Überprüfung der Handlungsebenen und Gestaltungskriterien aus der Perspektive der Mitbestimmung*, Vortrag auf dem PROTIS-BIT-Online-Workshop *Wie gestalten wir die sogenannte „Künstliche Intelligenz“?* des Forum Soziale Technikgestaltung am 2. Juni 2022 (Link zum Video im FST-Youtube-Kanal (Zugriff 21. Juni 2022): <https://www.youtube.com/watch?v=6HHakS0IUXw>).
- [9] W. Schröter, Gestaltungskompetenz für Betriebsräte im Spannungsverhältnis von Assistenz- und Delegationstechnik. Erfahrungen aus dreißig Jahren „Forum Soziale Technikgestaltung“, in E. Gruber, J. Schindler, Hg., *Digitalisierung und Erwachsenenbildung. Reflexionen zu Innovation und Kritik*. Magazin erwachsenenbildung.at – Das Fachmedium für Forschung, Praxis und Diskurs, Ausgabe 44-45, 2022, Wien 2022, S. 15-0 bis 15-8. Verfügbar unter: https://erwachsenenbildung.at/magazin/22-44u45/15_schroeter.pdf (Zugriff am: 7. Juni 2022).