

# Clemens H. Cap, Michael Fellmann und Johann-Christian Pöder Muster für ein ethisches Design von Assistenzsystemen

---

## 1. Einleitung

In vielen Gebieten der Wissenschaft, Technik und Kunst wie beispielsweise der Architektur kennt man das Konzept von Mustern (Design Patterns).<sup>1</sup> 1994 wurde es unter Mitwirkung der Universität Zürich auch in die Informatik eingeführt.<sup>2</sup> Der Grundgedanke des Musters besteht darin, für wiederkehrende Probleme abstrakt beschriebene Lösungen bereitzustellen. Inhalte von Mustern sind oft Mechanismen, die sich bereits bei einer Vielzahl ähnlicher Probleme bewährt haben. Das Konzept des Musters wurde zunächst in den Bereich des Software Engineering übertragen<sup>3</sup> und ist dort auch als Analysemuster bekannt.<sup>4</sup> Darüber hinaus haben Muster auch in anderen Bereichen weite Verbreitung erlangt, wie etwa als Workflow

<sup>1</sup> Christopher Alexander: A Pattern Language. Towns, Buildings, Construction, New York 1977.

<sup>2</sup> Erich Gamma, Richard Helm, Ralph Johnson, John Vlissides: Design Patterns. Elements of Reusable Object-Oriented Software, Reading MA 1994.

<sup>3</sup> Martin Fowler: Analysis Patterns. Reusable Object Models, Reading MA 1997.

<sup>4</sup> Sandeep Purao, Veda C. Storey, Taedong Han: Improving Analysis Pattern Reuse in Conceptual Design. Augmenting Automated Processes with Supervised Learning, in: Information Systems Research 14/3 (2003) 269–290.

Patterns<sup>5</sup> zur Beschreibung von Abläufen oder als Muster zur modellbasierten Gestaltung, Konfiguration und Ausführung von Geschäftsprozessen.<sup>6</sup>

Im Folgenden stellen wir Design Patterns für ethische Probleme beim Entwurf von digitalen Assistenzsystemen vor. Unter einem Assistenzsystem verstehen wir ein System, das den Nutzer in bestimmten Situationen oder bei bestimmten Handlungen unterstützt.<sup>7</sup> Dazu muss das System gegebenenfalls die aktuelle Situation des Nutzers analysieren und Vorhersagen zur zukünftigen Situation treffen. Auch sollte das System sich harmonisch in die Abläufe des Nutzers einfügen. Es ist sehr wahrscheinlich, dass das Telefon oder Smartphone, das Sie besitzen, bereits mit einem Assistenten ausgestattet ist, dass Fahrzeuge, mit denen Sie kürzlich gefahren sind, über eine Vielzahl an Assistenzsystemen verfügen und Sie vielleicht auf Webseiten oder in Ihrer Wohnung bereits mit einem Assistenten interagiert haben. Folglich entwickeln sich digitale Assistenzsysteme zunehmend zu Gefährten des Menschen.<sup>8</sup> Gerade diese direkte Auswirkung auf unser Leben lässt die Frage eines ethischen Designs zunehmend dringlich erscheinen. Die Beschreibung von typischen Konflikten und Rahmenbedingungen sowie die Bereitstellung von Mustern soll die ethische Gestaltung erleichtern. Während die Beschreibung von Konflikten und Rahmenbedingungen Problemwissen sammelt, enthalten die Muster Lösungswissen, das in Form von Prinzipien und Mechanismen bereitgestellt wird, um den in diesem Bereich immer wieder gleichen, grundsätzlichen Fragen wirkungsvoll

<sup>5</sup> Wil M. van Der Aalst, Ter Hofstede, Bartek Kiepuszewski, Alistair P. Barros: Workflow Patterns, in: Distributed and Parallel Databases 14/1 (2003) 5–51.

<sup>6</sup> Michael Fellmann, Agnes Koschmider, Ralf Laue, Andreas Schoknecht, Arthur Vetter: Business Process Model Patterns. State-of-the-Art, Research Classification and Taxonomy, in: Business Process Management Journal 25/5 (2019) 972–994: <https://doi.org/10.1108/BPMJ-01-2018-0021>

<sup>7</sup> Vgl. <https://dbis.informatik.uni-rostock.de/forschung/schwerpunkte/assistenzsysteme/>

<sup>8</sup> Jasmin Niess, Pawel W. Wozniak: Embracing Companion Technologies (2020) arXiv:2004.07198 [cs]. arXiv: 2004.07198.

zu begegnen. Konkrete Lösungen oder die Option für spezifische Wertesysteme werden wir dabei jedoch bewusst vermeiden, da die Letztentscheidung für das Wertesystem der entsprechenden Zielgruppe angepasst sein soll. Der Vorteil dieses Ansatzes liegt bei den vorgedachten Mechanismen, die im ethischen Diskurs aktiviert werden können, sobald sie situativ sinnvoll erscheinen.

Die Nutzer dieser Muster können dabei keine widerspruchsfreien Leitlinien erwarten, wie das Beispiel der Türschliessung einer S-Bahn zeigen soll: Einerseits darf niemand den Waggon ausserhalb der Station verlassen, andererseits muss bei Brand ein Notausstieg möglich sein. Der oft gewählte Lösungsansatz sieht eine mechanische Entriegelung der Türe vor, die aber nur im Stillstand möglich ist. Die Türe wird durch die Entriegelung nicht vollständig geöffnet, die Flügel der Türe müssen mit einem gewissen Kraftaufwand auseinandergeschoben werden.<sup>9</sup> Dies ist für Erwachsene im Allgemeinen wohl problemlos zu leisten, während etwa Kinder die Türe nicht öffnen können. Der mit diesem Design verbundene Kompromiss erfordert also eine multivalente Abwägung.

## 2. Zur Auswahl ethischer Prinzipien

Obwohl die klassischen ethischen Theorien (Utilitarismus, Pflichtethik, Tugendethik) mit Blick auf ethisches Design hilfreich sein können, etwa als Inspiration beim gemeinsamen Brainstorming eines Designteams,<sup>10</sup> ist die konkrete und strikte Anwendung solcher Theorien schwierig. Es entstehen Probleme bei der (letzt)begründeten Theoriwahl, bei der Ableitung konkreter Handlungsanweisungen

<sup>9</sup> Nottürentriegelung – Türen öffnen mit Muckis und Verstand. Hinweise zur Tür-Notfallentriegelung auf den Webseiten der Münchner S-Bahn-Betriebe:

<https://www.s-bahn-muenchen-magazin.de/artikel/Nottuerentriegelung>

<sup>10</sup> Jet Gispén: Normative Design Scheme (2017):

<http://www.ethicsfordesigners.com/normative-design-scheme>

(besucht am 24.3.2021).

(«Abstraktheit» der Theorien) sowie mit Blick auf die Neuartigkeit heutiger Probleme.<sup>11</sup> Bei einem gemeinsamen Ethik-Brainstorming mithilfe ethischer Theorien kann man daher unter anderem fragen:

- Trägt dieses digitale Assistenzsystem dazu bei, den Nutzen zu maximieren oder möglichst viele Präferenzen zu erfüllen? (Utilitarismus)
- Würde das Assistenzsystem jemanden instrumentalisieren oder diskriminieren, Autonomie und Menschenwürde verletzen? (Pflichtethik)
- Würde ein tugendhafter Mensch ein solches Assistenzsystem bauen oder benutzen? Ist die Benutzung dieses Assistenzsystems mit Blick auf ein «gutes Leben» ratsam? (Tugendethik)

Angesichts der erwähnten Schwierigkeiten hat man sich im Bereich der angewandten Ethik in den letzten Jahrzehnten oft den sogenannten «mittleren Prinzipien» zugewandt, die zwischen situationsbezogenen Low-level-Urteilen und ethischen High-level-Theorien angesiedelt sind. Das bekannteste Beispiel sind die vier medizinethischen Prinzipien von Tom Beauchamp und James Childress:<sup>12</sup> Respekt vor Autonomie, Nichtschaden, Wohltun und Gerechtigkeit.

Bei einem solchen Ansatz gibt es zwei zentrale Problemzusammenhänge: die richtige oder plausible Auswahl von Prinzipien und die richtige Anwendung von Prinzipien in konkreten Situationen. Bei der Auswahl versucht man in der Regel, eine anspruchsvolle philosophische Letztbegründung zu vermeiden. Es scheint ratsam, im Bereich des ethischen Technikdesigns pragmatisch von einem

<sup>11</sup> Sven Ove Hansson: Theories and Methods for the Ethics of Technology, in: Sven Ove Hansson (ed.): The Ethics of Technology. Methods and Approaches, Lanham MD 2017, 1–14, hier 4–9. Einen guten Überblick über die Schwierigkeiten einer schlichten Anwendung ethischer Theorien im Bereich des technischen Designs bietet auch Jeroen van den Hoven: The Use of Normative Theories in Computer Ethics, in: Luciano Floridi (ed.): The Cambridge Handbook of Information and Computer Ethics, Cambridge 2010, 60–76, hier 65ff.

<sup>12</sup> Tom L. Beauchamp, James F. Childress: Principles of Biomedical Ethics, Oxford <sup>7</sup>2013.

«robusten» Wertpluralismus auszugehen, was jedoch nicht unbedingt auf einen moralischen Relativismus oder Skeptizismus hinauslaufen muss.<sup>13</sup> So sind Beauchamp und Childress der Ansicht, dass die vier genannten Prinzipien zum Kernbestand unserer Alltagsmoral (common morality) gehören und sich rekonstruktiv einsichtig machen lassen. In der Tat haben diese vier Prinzipien einen relativ grossen Konsens und Zustimmung gefunden. Beauchamp und Childress beobachteten bereits in den 1970er Jahren, dass in bioethischen Debatten immer wieder auf diese vier Prinzipien zurückgegriffen wurde. Auch ihre eigene Herausarbeitung dieser Prinzipien hat eine weite Verbreitung gefunden.<sup>14</sup> Im Bereich der Technikethik hat die Forschungsgruppe von Luciano Floridi sechs verschiedene Prinzipien-Reihen untersucht und 47 verschiedene Prinzipien identifiziert.<sup>15</sup> Diese wurden gruppiert und mit den Prinzipien von Beauchamp und Childress verglichen, wobei eine weitgehende Konvergenz festgestellt werden konnte. Floridi et al. haben jedoch vorgeschlagen, diese vier Prinzipien für den Bereich der künstlichen Intelligenz um ein fünftes Prinzip, die Erklärbarkeit (explicability) zu ergänzen. Damit ergeben sich die folgenden fünf Prinzipien:

- Beneficence: Promoting Well-Being, Preserving Dignity, and Sustaining the Planet
- Non-Maleficence: Privacy, Security and «Capability Caution»
- Autonomy: The Power to Decide (Whether to Decide)

<sup>13</sup> Jeroen van den Hoven: The Use of Normative Theories in Computer Ethics, in: Luciano Floridi (ed.): *The Cambridge Handbook of Information and Computer Ethics*, Cambridge 2010, 59–76, hier 62–64.

<sup>14</sup> Marcus Düwell: *Bioethik: Methoden, Theorien und Bereiche*, Stuttgart Weimar 2008, 89–90.

<sup>15</sup> Luciano Floridi, Josh Cowsls, Monica Beltrametti et al.: AI4People – An Ethical Framework for a Good AI Society. Opportunities, Risks, Principles, and Recommendations, in: *Minds and Machines* 28 (2018) 689–707, hier 695–696. – Vgl. auch den Abschnitt «Ethical Guidelines for AI» im Artikel «Ethics of Artificial Intelligence» von John-Stewart Gordon und Sven Nyholm in: *Internet Encyclopedia of Philosophy* (2021):

<https://iep.utm.edu/ethic-ai/>

- Justice: Promoting Prosperity and Preserving Solidarity
- Explicability: Enabling the Other Principles Through Intelligibility and Accountability

Als ein weiteres und auch für die Technikethik wichtiges Beispiel für «mittlere Prinzipien» kann Marta Nussbaums aristotelisch orientierte Liste von Fähigkeiten (capabilities) dienen.<sup>16</sup>

- Leben: die Fähigkeit, ein menschliches Leben normaler Dauer bis zum Ende zu leben
- körperliche Gesundheit: die Fähigkeit, bei guter Gesundheit zu sein (einschliesslich angemessener Ernährung und Unterkunft)
- körperliche Integrität: die Fähigkeit, sich frei zu bewegen und vor Gewalt sicher zu sein
- Sinneswahrnehmungen, Vorstellungsvermögen und Gedanken: die Fähigkeit, die Sinne zu benutzen, sich etwas vorzustellen und zu denken (einschliesslich der entsprechenden Erziehung und Ausbildung)
- Emotionen: die Fähigkeit, Bindungen zu Dingen und Personen aufzubauen (Liebe, Sorge, Sehnsucht, Dankbarkeit, berechtigter Zorn usw.)
- praktische Vernunft: die Fähigkeit, selbst eine Auffassung des Guten zu bilden und die eigene Lebensplanung kritisch zu reflektieren
- Zugehörigkeit: die Fähigkeit, mit anderen zu leben, an sozialer Interaktion teilzunehmen und mit Würde behandelt zu werden
- andere Spezies: die Fähigkeit, in Beziehung zu Tieren, Pflanzen und zur Natur zu leben
- Spiel: die Fähigkeit, zu lachen, zu spielen und sich zu erholen
- Kontrolle über die eigene Umwelt: die Fähigkeit, am politischen Leben teilzunehmen, zu arbeiten und Eigentum zu besitzen

Eine solche Liste mag im Vergleich zu Beauchamp und Childress gewagter und anspruchsvoller erscheinen, hat aber für viele Menschen eine intuitive, an der eigenen Lebenssituation verifizierbare Plausibilität. Crighton Nichols und Andy Dong heben hervor,

<sup>16</sup> Marta Nussbaum: Die Grenzen der Gerechtigkeit. Behinderung, Nationalität und Spezieszugehörigkeit, Berlin 2010, 112–114.

dass eine Anzahl von Nussbaums «Fähigkeiten» auf die Designpraxis verweisen.<sup>17</sup> Ein gutes Beispiel, wie man Nussbaums Ansatz für ethisches Design in digitaler Assistenz benutzen kann, findet man in Naomi Jacobs Ansatz des Capability Sensitive Design (CSD), der Nussbaums Theorie der Fähigkeiten mit dem Value Sensitive Design (VSD) verknüpft.<sup>18</sup>

### 3. Anwendungsregeln von Prinzipien

Das zweite Problem ist die richtige Anwendung von Prinzipien in konkreten Situationen. Für Beauchamp und Childress sind die zwei zentralen Methoden der Anwendung (3.1) Spezifizierung (specification) und (3.2) Abwägung (balancing).

#### 3.1 Spezifizierung

Bei der Spezifizierung geht es um die Verringerung des Anwendungsbereichs eines Prinzips, um die Anwendung konkreter und operationalisierbarer zu machen. Als Spezifizierung des Prinzips «Respekt vor Autonomie» können etwa die weiter unten erwähnten Muster «Abschalten» und «Abstufung» genannt werden. Ein konkretes Beispiel wäre etwa: Respektiere die Autonomie, indem du das digitale Assistenzsystem so baust, dass die Nutzer das Assistenzsystem nach Möglichkeit immer auch abschalten können. Oder noch weiter spezifiziert: ... indem die Nutzer des Autos das Warnsignal des Sicherheitsgurts abschalten können. Die letzte Spezifizierung kann jedoch mit einem anderen Prinzip in Konflikt geraten

<sup>17</sup> Crighton Nichols; Andy Dong: Re-conceptualizing Design Through the Capability Approach, in: Ilse Oosterlaken, Jeroen van den Hoven (eds.): *The Capability Approach, Technology and Design*, Dordrecht 2012, 189–201, 184.

<sup>18</sup> Naomi Jacobs: *Capability Sensitive Design for Health and Wellbeing Technologies*, in: *Science and Engineering Ethics* 26 (2020) 3363–3391.

(Nichtschaden, Wohltun oder Gerechtigkeit) und eventuell ethisch nicht zu verantworten sein. Auch Abstufung kann als Spezifizierung des Prinzips der Autonomie (aber auch anderer Prinzipien) angesehen werden. So kann «Respekt vor Autonomie» konkreter bedeuten, dass ein digitales Assistenzsystem verschiedene Stufen der Selbstbestimmung (Einstellungen) ermöglicht.

Ein weiteres wichtiges Muster der Spezifizierung ist die Transparenz. Mit ihrer Hilfe kann das Prinzip der Autonomie wie folgt spezifiziert werden: Respektiere die Autonomie, indem du die Entscheidungsarchitektur in deinem System möglichst transparent machst und damit informierte Entscheidungen über die Benutzung des digitalen Assistenzsystems ermöglichst. Das Gleiche gilt mit Blick auf die Rückkopplung: Rückkopplung kann etwa als Spezifizierung des Prinzips Wohltun, aber auch von Gerechtigkeit, Autonomie und Nichtschaden angesehen werden. Durch Rückkopplung kann ein Assistenzsystem die Autonomie der Nutzer erhöhen (siehe unten das Beispiel der unerwünschten Werbung unter 5.2 Transparenz), aber auch die Gerechtigkeit befördern (die Rückkopplung ermöglicht auch das kritische Feedback von Angehörigen einer Minderheit, deren Bedürfnisse beim Designprozess nicht hinreichend berücksichtigt wurden).

### 3.2 Abwägung

Bei der Abwägung geht es vor allem um Konfliktsituationen, in denen verschiedene relevante ethische Prinzipien miteinander in Konflikt stehen. Die Vorrangbeziehungen müssen manchmal schon im Designprozess bestimmt werden. Zum Beispiel wird die Abschaltung des Warnsignals für Sicherheitsgurte zwar generell nicht zugelassen, was zum Nachteil der Autonomie und zum Vorteil des Nichtschadens oder der Gerechtigkeit ist, jedoch wird als Kompromiss eine Warnmeldung erst ab einer bestimmten Geschwindigkeit von wenigen km/h ausgegeben, so dass der Nutzer etwa bei kurzen Rangierfahrten von wenigen Metern nicht zum Anlegen des



Gurts ermahnt wird. Um eine ausgewogene Berücksichtigung der verschiedenen Prinzipien in Abwägungsprozessen zu erreichen, werden in Bezug auf den Einsatz von Künstlicher Intelligenz Ethik-Kommissionen empfohlen, die sich aus Mitgliedern mit verschiedenen Hintergründen und Kompetenzen zusammensetzen.<sup>19</sup> In vielen Fällen wird das Assistenzsystem aber die Abwägung eher den Nutzern selbst freistellen. So überlässt beispielsweise ein Assistenzsystem des Streaming-Systems für Serien den Nutzern, wie viele Stunden man Binge Watching machen kann, also einen «Serienmarathon», indem man viele Folgen einer Serie hintereinander anschaut, obwohl man weiss, dass dies zu Suchterscheinungen und zur Minderung der Produktivität führen kann. Transparenz kann auch als eine wichtige Komponente der Abwägung angesehen werden. Um eine Abwägung von konfligierenden Prinzipien (etwa Gerechtigkeit vs. Wohltun) vorzunehmen, sollte man eine möglichst klare Einsicht in die Funktionsweise des Assistenzsystems haben. Das Gleiche gilt mit Blick auf Rückkopplung. Eine qualitative oder fehlende Rückkopplung kann bedeuten, dass die Abwägung von Prinzipien einen bestimmten Lauf nimmt oder nicht. Sind beispielsweise Feedbacks von Nutzererfahrungen effektiv möglich und wird die Gerechtigkeit aufgrund von Feedback auch tatsächlich erhöht, kann dies sensible Auswirkungen auf die Abwägung (Balance) etwa von Gerechtigkeit und Wohltun haben – die Benutzung einer Suchmaschine oder einer Crowdfunding-Plattform würde nicht mehr massiv das Prinzip der Gerechtigkeit verletzen und wäre im Sinne des Wohltuns zu rechtfertigen.

<sup>19</sup> Marcus Schmitz: Auf dem Weg zu einer ethisch verantwortungsvollen Datennutzung, in: Informatik Spektrum 43 (2020) 366–373: <https://doi.org/10.1007/s00287-020-01298-9>

#### 4. Bezug zum Value Sensitive Design (VSD)

Die geschilderten Prinzipien und Anwendungsregeln haben eine Nähe zu dem von Batya Friedman und ihren Mitarbeitern entwickelten Ansatz des VSD.<sup>20</sup> Dieser Ansatz basiert auf der Einsicht, dass technische Artefakte nie gänzlich neutral sind, sondern immer mehr oder weniger wertebeladen. Das VSD versucht, die Reflexion und Berücksichtigung von moralischen Werten proaktiv und praktisch zum integrativen Teil des gesamten Designprozesses zu machen. Auch wenn VSD statt von Prinzipien von Werten spricht, kann man inhaltlich von wichtigen Konvergenzen ausgehen. Wie aber findet die Auswahl von relevanten Werten statt? Das VSD scheint mit einer doppelten und etwas spannungsvollen Strategie zu operieren. Einerseits geht es darum, die Werte der relevanten Stakeholder empirisch oder partizipativ zu ermitteln und im Designprozess zu berücksichtigen. Andererseits operiert das VSD aber durchaus auch mit universellen menschlichen Werten, die es immer zu berücksichtigen gilt. Batya Friedman und Peter Kahn schlagen eine Liste von zwölf ethischen Werten (human values with ethical import) vor:<sup>21</sup>

1. Human welfare
2. Ownership and property
3. Privacy
4. Freedom from bias
5. Universal usability
6. Trust
7. Autonomy

<sup>20</sup> Batya Friedman, Peter H. Kahn, Alan Borning: Value Sensitive Design and Information Systems, in: Ping Zhang; Dennis Galletta (eds.): Human-Computer Interaction in Management Information Systems, New York 2006, 348–372.

<sup>21</sup> Batya Friedman; Peter H. Kahn: Human Values, Ethics, and Design, in: Julie A. Jacko, Andrew Sears (eds.): The Human-Computer Interaction Handbook, Mahwah NJ (2003) 1177–1201, hier 1187–1193. – Vgl. Batya Friedman, David Hendry: Value Sensitive Design: Shaping Technology with Moral Imagination, Cambridge MA 2019.

8. Informed consent
9. Accountability
10. Identity
11. Calmness
12. Environmental sustainability

Diese Liste soll jedoch offen für Verbesserungen sein und die jeweilige Kontextualisierung der Werte berücksichtigen. Eine solche Liste hat nach Friedman und Kahn eine heuristische Funktion – sie kann den Designprozess von Anfang an produktiv unterstützen, indem sie wichtige Werte hervorhebt und davor schützt, dass diese unter Zeitdruck übersehen werden. Was die Anwendung von Werten angeht, schlägt das VSD eine dreigliedrige Methodik vor, die aus konzeptionellen, empirischen und technischen Untersuchungen besteht. Konzeptionelle Untersuchungen nehmen eine Analyse ethisch relevanter Werte vor und reflektieren mögliche Abwägungen zwischen den Werten. Empirische Untersuchungen analysieren die Werte und Bedürfnisse der Stakeholder, mögliche Nutzungsweisen des Designprodukts sowie mögliche Wirkungen des Geräts auf die Nutzer. Technische Untersuchungen fokussieren darauf, welche technischen Eigenschaften des Systems humane Werte fördern oder behindern, und wie ein digitales Assistenzsystem technisch aussehen sollte, um einen wichtigen Wert proaktiv unterstützen zu können (beispielsweise Privacy). Die Spezifizierung eines Wertes, zum Beispiel mithilfe von Abschaltung, Abstufung, Transparenz oder Rückkopplung, kann im Kontext des VSD durch eine produktive Zusammenarbeit von allen drei Arten von Untersuchungen stattfinden. Die konzeptionelle Untersuchung kann zum Beispiel allgemein sagen, dass mit Blick auf den Wert Autonomie die zu entwickelnde Social-Media-Plattform abgestufte Einstellungen erlauben sollte. Die technische Untersuchung kann für die Abstufung unterschiedliche technische Lösungen vorschlagen, zum Beispiel verschiedene Einstellungen. Die empirische Untersuchung kann feststellen, ob die Nutzer auch wirklich von diesen Funktionen Gebrauch machen würden oder diese Einstellungen als eine befriedigende Lösung ansehen würden.

Insgesamt ist die bisher beschriebene ethische Gestaltung, Implementierung und Nutzung von Assistenzsystemen unter Verwendung von Mustern also durch den in Abbildung 1 illustrierten Prozess gekennzeichnet:

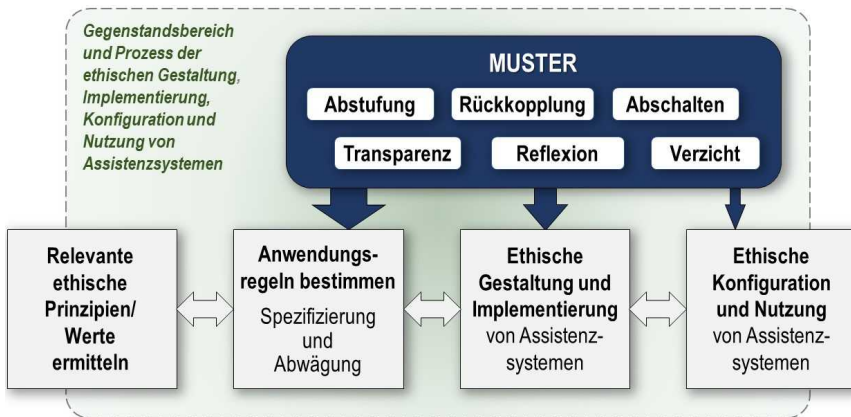


Abbildung 1: Prozess des ethischen Designs

Zunächst wird ermittelt, welche ethischen Prinzipien und Werte als relevant erachtet werden. Dies bildet die Grundlage für die nächste Phase, in der die Anwendungsregeln dieser Werte bestimmt werden – etwa die Konkretisierung und Anwendung der Werte und Prinzipien durch deren Spezifizierung und Abwägung. Hierbei stellen die im nächsten Abschnitt beschriebenen Muster eine Grundlage bereit, um die Spezifizierung und Abwägung im Einzelfall zu erleichtern, etwa durch das Bereitstellen von Beispielen zur Umsetzung der Prinzipien oder von zu erwartenden Argumenten in Konfliktsituationen. Nach der Spezifizierung und Abwägung erfolgt die Phase der Gestaltung und Implementierung des Assistenzsystems. In dieser Phase stellen die Muster abstrakt beschriebene Lösungen bereit, die bei häufigen Gestaltungsproblemen im Sinne einer Vorlage zur Lösung verwendet werden können. Abschliessend wird das Assistenzsystem konfiguriert und genutzt, auch hier können Muster wie etwa das Abschalten durch eine Systemeinstellung oder der situative oder sogar permanente Verzicht zum Einsatz kommen.

Alle Muster sind grundsätzlich für alle Phasen mit Ausnahme der ersten Phase (Selektion ethischer Prinzipien/Werte) relevant, auch wenn eine umfassende und tiefgreifende Auseinandersetzung mit den Mustern eher in der Phase der Anwendungsregelbestimmung und der Gestaltung und Implementierung zu erwarten ist. Die Betonung dieser beiden Phasen rührt auch daher, dass einige Muster bei der Konfiguration und Anwendung des Assistenzsystems nur dann relevant sind, wenn diese auch zuvor bei der Gestaltung und Implementierung berücksichtigt worden sind. So lässt sich etwa das Prinzip der Abstufung oder Transparenz bei der Nutzung des Assistenzsystems nur dann anwenden, wenn das System überhaupt eine Abstufung implementiert und Einblick in seine Funktionsweise gewährt. Der Prozess der ethischen Gestaltung, Implementierung, Konfiguration und Nutzung von Assistenzsystemen ist keine einfache, sequentielle Abfolge von Phasen, sondern kann auch beliebige Rücksprünge beinhalten. So könnte sich erst während der Nutzung herausstellen, dass die Anwender des Systems von einer wichtigen Konfigurationseinstellung keinen Gebrauch machen, womit in die Phase der Gestaltung und Implementierung zurückgesprungen werden könnte und eine andere Grundeinstellung gesetzt werden könnte (beispielsweise Vor-einstellung einer Konfigurationsoption im Sinne eines Opt-in statt Opt-out).

## 5. Muster

Im Folgenden stellen wir einige Muster für wiederkehrende ethische Probleme beim Design von Assistenzsystemen vor, die uns bei der Analyse eines Katalogs von Beispielen als wegweisend erschienen; wir erheben dabei keinen Anspruch auf Vollständigkeit. Jedes Muster wird zunächst erklärt, dann mit einem Beispiel versehen und schliesslich einer weitergehenden Analyse unterzogen.

## 5.1 Abstufung

*Erklärung:* Lenkende Eingriffe eines Assistenzsystems sollten in abgestufter, an die Situation angepasster Form geschehen.

*Beispiel: Ausstiegswarner beim Pkw.* – Ein Auto- oder Mitfahrer sollte eine Türe nicht plötzlich aufstossen, denn dadurch können andere Verkehrsteilnehmer wie etwa Radfahrer zu Schaden kommen. Ein Assistenzsystem kann dieses Ziel unterstützen und versuchen, diese auch als Dooring bezeichneten Unfälle zu verhindern. Dabei sind unterschiedliche Abstufungen denkbar. Gegenwärtig bereits vereinzelt implementierte Systeme warnen beim Öffnen der Türe audiovisuell durch einen Warnton und Blinken im Aussenspiegel (Mercedes) oder verzögern mechanisch die Öffnung der Türe um eine knappe Sekunde (Audi), um die Insassen haptisch zu warnen.<sup>22</sup> Allerdings ist auch hier die Art des Eingriffs sorgfältig zu prüfen und in ihrer praktischen Auswirkung zu beobachten (siehe auch die Muster 5.3 Rückkoppelung und 5.5 Reflexion). So gibt der ADAC diesbezüglich zu bedenken: «Das kurze Blockieren der Tür, die haptische Warnung, ist sehr effektiv. Allerdings besteht die Gefahr, dass die Türe nach der knappen Sekunde Verzögerung doch aufgestossen wird. Die Situation wäre dann sogar noch kritischer als vorher, da der Radfahrer keine Chance mehr zur Reaktion hätte. Die ADAC Technik-Experten empfehlen, die Öffnungsverzögerung solange aufrechtzuhalten, bis die kritische Situation vorbei ist. Hier müssen weitere Untersuchungen angestellt werden, ob eine längere Verzögerung bei Kunden Akzeptanz finden würde. Kommt eine haptische Warnung (verzögerte Öffnung) zu spät, sollte zumindest durch Licht und Ton gewarnt werden.»<sup>23</sup>

*Analyse:* Das Muster ist in Verbindung mit dem Abschalten (siehe unten 5.4) zu sehen und kann als eine Weiterentwicklung dieses Musters verstanden werden.

<sup>22</sup> Ausstiegswarner: So schützen sie vor Dooring-Unfällen. Bericht der ADAC Redaktion vom 21.5.2021: <https://www.adac.de/rund-ums-fahrzeug/tests/assistenzsysteme/ausstiegswarner/>

<sup>23</sup> Ebd.

## 5.2 Transparenz

*Erklärung:* Transparenz bedeutet, dass der Designer eines Systems über die wesentlichen Aspekte, welche die Entscheidungsarchitektur in seinem System beeinflussen, vollständig informiert.

*Beispiel: Wert eines Profils.* – Der Käufer eines Gebrauchtwagens findet mannigfaltige Unterstützung bei der Aushandlung des Preises. Er kennt den Listenpreis, kann sich über die Eurotax-Liste Expertenmeinungen einholen oder das Fahrzeug bei einem Auto-Club wie dem ADAC begutachten lassen. Schliesslich kann er den Preis in Relation zur eigenen Tätigkeit und zum eigenen Stundenlohn setzen. Die Nutzer eines Profils bei Google, Twitter oder Facebook stehen regelmässig vor der Frage, welche Informationen sie diesen Unternehmen anvertrauen, ob sie sich vor einer Suchanfrage bei ihrem Account anmelden oder abmelden und welche Mechanismen sie nutzen, um diesen Unternehmen Daten zu übermitteln, vor ihnen zu verbergen oder sie aktiv zu löschen. Anders als der Gebrauchtwagenkäufer kann der Nutzer aber keinerlei Abschätzung über den ökonomischen Wert oder Schaden seines Verhaltens treffen. Wird er nun mit mehr und mehr Werbung bombardiert? Wird er deshalb Dinge kaufen, die er eigentlich gar nicht braucht? Wird er Produkte zu erhöhten Preisen angeboten bekommen, weil Werbepartner der sogenannten sozialen Netzwerke um seine Zahlungsfähigkeit oder sein besonderes Interesse an bestimmten Produkten wissen? Der einzige Anhaltspunkt könnte allenfalls die Marktkapitalisierung des Unternehmens geteilt durch die Anzahl seiner Nutzer sein – doch auch dieser Eckwert hat nur beschränkte Aussagekraft. Bereits bei Google versagt der Ansatz, da der Konzern in einer Vielzahl unterschiedlichster Aktivitäten engagiert ist; bei Facebook oder Pinterest hingegen ist der Umsatz im Wesentlichen noch mit sozialen Profilen alleine verbunden.

*Analyse:* Transparenz ist für Aussenstehende sehr schwer zu auditieren und zu prüfen. Sie ist für ein Unternehmen, das neue Big-Data-Anwendungen auf den Markt bringen möchte, aus Gründen des Schutzes der neuen Idee unter Umständen sehr problematisch.

Transparenz wird durch die Datenschutzgrundverordnung (DSGVO) teilweise gefordert. Dennoch geht die dort verlangte Information über den Zweck einer Datenerfassung nicht weit genug, da sie den Zusammenhang mit den Interessen des Anwenders nicht ausreichend herausarbeitet.

### 5.3 Rückkopplung

*Erklärung:* Rückkopplung bedeutet, dass die Subjekte eines Systems seinem Designer Rückmeldungen geben über die Auswirkungen sowie die erwünschten und unerwünschten Nebenwirkungen des Systemdesigns.

*Beispiel: Spezifische Werbung.* – Ein Mailanbieter kann durch Analyse der E-Mails seiner Kunden Informationen über den Beziehungsstatus seiner Nutzer erhalten. Für die Betreiber sozialer Netzwerke sind diese Informationen meist direkt aus Profilangaben ersichtlich. Beides kann nun dazu führen, dass Personen, die keine Beziehung suchen oder die vielleicht gerade den Tod eines Partners oder eine Trennung betrauern, Werbung für Single-Plattformen erhalten. Wenn Alice für ihre Tante Carol auf Amazon ein Gartenwerkzeug bestellt, kann sie die nächsten Monate von Amazon immer wieder auf Gartenprodukte aufmerksam gemacht werden, obwohl sie das Thema nicht interessiert. Bei spezifischeren Produkten, etwa besonderem Sexspielzeug, können solche Werbungen zu weiteren seltsamen Effekten führen. Rückmeldung würde dem Nutzer das Signal erlauben, dass eine spezifische Lenkungsaktivität des Systems aus verschiedenen Gründen besonders unerwünscht ist.

*Analyse:* Rückkopplung kann die Beziehung zwischen dem Nutzer und dem Entwickler eines Systems verbessern. Rückkopplung kann auch als Alibi-Mechanismus benutzt werden in dem Sinn: «Lieber Anwender, ich frage zwar deine Meinung zum System ab und du kannst sie mir auch schreiben und dadurch zu einer emotionalen Entlastung deiner Unzufriedenheit gelangen. Die Rückmeldung werde ich aber nicht ernsthaft in Betracht ziehen».



Rückkopplung kann missbraucht werden, wenn der Entwickler eines Systems dadurch zusätzliche Einblicke in die Wirkmechanismen seiner Architektur bekommt und Informationen gegen das Interesse des Anwenders verwendet. So kann die Rückmeldung «Immer, wenn ich ein Produkt mehrfach kaufe, wird es teurer» dazu führen, dass der Anbieter gerade dadurch auf die Idee kommt, ein Produkt für Mehrfachkäufer ganz besonders teuer zu machen.

#### 5.4 Abschalten

*Erklärung:* Abschalten bedeutet, dass der Benutzer spezifische Funktionen eines Assistenzsystems oder das System als Ganzes für bestimmte Zeiträume ausser Funktion setzen kann. Dieses Muster kann man sowohl als eine Spezifizierung der Privatheit im Sinne eines «right to be left alone»<sup>24</sup> als auch als eine Spezifizierung des Prinzips der Autonomie verstehen. Man könnte es als Recht kodifizieren, das dem mündigen Bürger zugesteht, Eingriffe eines technischen Systems in sein Leben zu unterbinden, jedenfalls in allen Situationen, in denen Dritte nicht zu Schaden kommen können. Ebenso bietet sich eine Assoziation zum zweiten Asimovschen Gesetz an, dem zufolge ein Roboter den von einem Menschen gegebenen Befehlen gehorchen muss, was auch Befehle zur Untätigkeit umfasst. Bereits bei Asimov findet sich eine wichtige Schranke für dieses Muster im Nachsatz «es sei denn, ein solcher Befehl würde mit Regel eins kollidieren», welche fordert, dass ein Roboter kein menschliches Wesen wissentlich verletzen darf oder ihm durch Untätigkeit wissentlich Schaden zulassen darf.<sup>25</sup>

*Beispiel 1: Berganfahrhilfen und Autopiloten.* – Manche Autos haben eine Berganfahrhilfe. Beim Lösen der Bremse in der Steigung hält das Fahrzeug für einen kurzen Moment die Bremsen aktiv, bis der

<sup>24</sup> Samuel D. Warren, Louis D. Brandeis: The Right to Privacy, in: Harvard Law Review 4/5 (December 15, 1890) 193–220:

<https://www.cs.cornell.edu/~shmat/courses/cs5436/warren-brandeis.pdf>

<sup>25</sup> Isaac Asimov: Meine Freunde, die Roboter, München 1997.

Fahrer aus dieser Situation heraus das Fahrzeug beschleunigt. Sinn des Systems ist die Unterstützung des Fahrers, der ansonsten Bremse, Gaspedal und Handbremse gleichzeitig koordiniert bedienen muss. Weil bei manchen Assistenten allerdings die Bremse an den Hinterrädern wirksam ist und erst dann gelöst wird, wenn das Fahrzeug eine bestimmte Zugkraft nach vorne über die angetriebenen Vorderräder aufgebaut hat, ist das Anfahren auf ansteigenden, glatten Flächen zuweilen nicht mehr möglich, weil die Vorderräder zu wenig Bodenhaftung haben, um die erforderliche Kraft aufzubauen. Ähnliche Situationen gibt es bei Flugzeugen, bei denen Autopiloten den Betrieb des Flugzeugs in Parameterbereichen verhindern, die zum Absturz führen würden. In der Luftfahrt wurden tödliche Unfälle verursacht, weil Piloten die genaue Funktion eines Assistenzsystems nicht kannten (siehe 5.2 Transparenz) oder es nicht (auf einfache Weise) abschalten konnten (Beispiel: Boeing 737 MAX 8).

*Beispiel 2: Atemalkoholkontrolle.* – Alice, die hochschwanger ist, und Bob sind auf einer Party. Bob hat etwas zu viel getrunken. Eigentlich möchte er nicht mehr mit dem Auto fahren, obwohl er sich im Notfall noch ein sicheres Fahren zutraut. Daher bittet er Alice für die Rückfahrt ans Steuer. Alice bekommt jedoch plötzlich verfrühte Wehen. Nun möchte Bob das System seines Autos abschalten, das bei einem Fahrer eine Atemalkoholkontrolle vornimmt, bevor es den Motor startet. Das Beispiel lässt sich mit leichter Variation aber auch in ein Gegenbeispiel verändern: Hier nimmt Bob höchst leichtsinnig und trotz massiver Alkoholisierung in emotional schwierigem Zustand sein Auto in Betrieb, missbraucht dazu das Abschaltssystem und verursacht einen schweren Unfall.

*Gegenbeispiel: Elektrische Sicherung.* – Eine elektrische Sicherung verhindert die Überlastung eines Stromkreises und die damit verbundene Brandgefahr in Gebäuden. Es ist vermutlich nicht sonderlich sinnvoll, einem Stromkunden die Möglichkeit zu geben, die Abschaltfunktion einer Sicherung oder eines Fehlerstromschalters zu überbrücken oder diese ausser Betrieb zu setzen. Dessen ungeachtet sind Situationen denkbar, in denen eine solche Überbrückung

sinnvoll sein könnte. Allerdings bleibt fraglich, ob solche Situationen so häufig sind, dass eine solche Überbrückung als grundsätzliche Funktion sinnvoll erscheint.

*Analyse:* Dieses Muster setzt das der Transparenz (siehe oben 5.2) voraus. Ein Benutzer kann eine spezifische Funktion eines Assistenzsystems nur dann abschalten, wenn er überhaupt weiss, dass es sie gibt. Das Muster ist nicht ohne Risiko und erfordert Aufklärung des Benutzers. Es gibt einen Grund, weshalb bestimmte technische Systeme zum Schutz des Anwenders auf bestimmte Weise reagieren; dieser Grund muss aber nicht immer zutreffen und die Abwägung des Entwicklers kann fehlerhaft sein. Manche Hersteller bieten eine Unterstützung des Abschaltens an, wie etwa Apple mit der Anzeige der Bildschirmzeiten am Mobiltelefon.

## 5.5 Reflexion

*Erklärung:* Die Entwicklung digitaler Lenkungsmechanismen sollte durch ein besonderes Mass an Reflexion begleitet werden. Insbesondere sollte darauf geachtet werden, inwieweit vorgeschlagene Unterstützungsmechanismen Probleme tatsächlich lösen oder sie durch ihr Wirken Probleme oder Bedarfe, dann meist anderer Natur, erst erzeugen. Optimierung ist als Mechanismus zunächst einmal gut. Ein Übermass an Optimierung kann jedoch ausser Kontrolle geraten und dadurch selbst zum Problem werden. Dabei sind gerade solche Mechanismen, die Paul Watzlawick Probleme zweiter Ordnung nennt,<sup>26</sup> bedeutsam. Bei diesen führt eine zunächst als vermeintliche Lösung angesehene Massnahme zu einem Problem – nämlich einem Problem zweiter Ordnung –, da mittlerweile die Beteiligten ihre Erwartungshaltungen so verändert haben, dass sie nicht in der Lage sind, die Lösung des ersten Problems als Ursache des neuen Problems zu erkennen. Besonders effiziente technische

<sup>26</sup> Paul Watzlawick, John Weakland, Richard Fisch: Lösungen. Zur Theorie und Praxis menschlichen Wandels, Bern, Stuttgart, Wien 1979.

Lösungen stehen in erhöhter Gefahr für solche Entwicklungen. Wir verweisen zur Erläuterung insbesondere auf zwei Beispiele.

*Beispiel 1: Optimierung der Kommunikation.* – E-Mail beschleunigt die Kommunikation und soll sie verbessern. Für manche Zwecke ist E-Mail aber zu langsam, zu formalisiert und zu träge. Sie wird daher durch eine rasche Abfolge immer effizienterer Formen ersetzt: Short Message, Messenger, Chats, Tweets, Slack-Beiträge und Timelines lösen die E-Mail ab. Schliesslich werden Status-Mechanismen erforderlich, in denen ein Teilnehmer angeben kann, ob er gerade «anwesend» ist (also innerhalb der nächsten Sekunden auf Kontaktaufnahmen reagiert), oder ob er gerade «beschäftigt» ist (der Partner also keine schnelle Reaktion erwarten soll). Am Ende dieser Entwicklung steht eine Gesellschaft, in der junge Menschen zunehmend an Aufmerksamkeitsstörungen und Hyperaktivität leiden, entsprechende Krankheitsbilder in die Diagnose-Klassifikationen aufgenommen werden und die Verschreibung von Ritalin explosionsartig ansteigt. Ins Bild passt hier auch eine Studie der Ohio State University, nach welcher die Anwesenheit eines Like-Buttons die Lesedauer und die eigene inhaltliche Auseinandersetzung mit Texten verkürzt. Die Bestätigung der eigenen Meinung wird wichtiger als die Auseinandersetzung mit Inhalten. Bemerkenswerterweise tritt ein entsprechender Effekt aber auch in der Agenturnachricht auf,<sup>27</sup> der wir diese Information entnehmen. Diese verweist nämlich nur auf die Homepage der Ohio State University statt auf die dort vermutlich auffindbare Studie; wesentlich erscheint also wiederum nicht die Auseinandersetzung mit den Inhalten, die (rasche?) Fertigung einer weiteren Pressenachricht steht im Vordergrund und für die journalistische Sorgfalts- und wissenschaftliche Zitierpflicht wird ein beliebiger, wenig hilfreicher Link als ausreichend angesehen. Die Lösung eines Problems wird selbst zu einem Problem.<sup>28</sup>

<sup>27</sup> Georg Haas: Like-Button verkürzt Lesezeit bei Online-Artikeln (2020): <https://www.presetext.com/news/like-button-verkuerzt-lesezeit-bei-online-artikeln.html> (besucht am 24.3.2021).

<sup>28</sup> Vgl. auch P. Watzlawick, J. Weakland, R. Fisch: Lösungen; Byung-Chul Han: Müdigkeitsgesellschaft, Berlin 2010.

*Beispiel 2: Eigenleben der Assistenzsysteme.* – In den Wissenschaften können wir – vereinfacht dargestellt – zwei Arten von Assistenzsystemen für unser Gehirn benutzen: Bleistift und Papier oder digitale Systeme. Bleistift und Papier entwickeln kein Eigenleben, sie nehmen Gedanken auf. Digitale Systeme aber wollen konfiguriert und angepasst werden, benötigen Updates, liefern Benachrichtigungen und drängen sich auf vielfältige Weise zwischen die Person und ihre ursprüngliche Aufgabe. Ähnlich verhält es sich bei Kommunikationssystemen: Telefone und Videokonferenzen erleichtern das soziale Leben. Ich muss nicht mehr nach Tokyo fliegen, um meine japanischen Freunde zu sehen. Mit Facebook und Twitter aber schieben sich neue Aspekte in den Vordergrund. Anstatt Beziehungen zu pflegen, gestalten wir Profile und Timelines aus einem Selbstzweck heraus und sammeln Likes. Der ursprüngliche Aspekt der Begegnung wird durch einen neuen, oft auch unerwünschten oder technischen Selbstzweck verdrängt.

*Analyse:* Das reflektierende Nachdenken über Probleme höherer Ordnung ist eine Sache für Philosophen und nicht für Praktiker. Der Aufruf zu Reflexion kann daher in das Gegenteil umschlagen, weil er als weltfremd und innovationsfeindlich wahrgenommen wird. Das aber charakterisiert gerade das Problem zweiter Ordnung: Man nimmt es nicht als Problem wahr, weil man die innewohnenden Mechanismen nicht erkennt. Somit benötigt es besonders lange und tiefgreifende Schädigungen, bis die Gesellschaft bereit ist, sich mit der Problematik überhaupt zu beschäftigen, dann aber möglicherweise überzogen und falsch reagiert.

## 5.6 Verzicht

*Erklärung:* Verzicht bedeutet die bewusste Entscheidung gegen ein Assistenzsystem. Während das Abschalten die selektive Nichtnutzung eines grundsätzlich verfügbaren Assistenzsystems bedeutet, meint der Verzicht die vollständige Absage an die Nutzung oder Entwicklung eines Assistenzsystems für einen bestimmten

Lebensbereich. Erfordert das Abschalten meistens noch die Unterstützung durch den Hersteller, so ist der Verzicht eine persönliche Entscheidung oder gesellschaftliche Regelung gegen Assistenzsysteme.

*Beispiel 1: Digital Detox.* – Im Internet finden sich Selbstberichte und zunehmend entstehen auch Studien zum Phänomen eines Digital Detox. Das Schlagwort steht für ein Beenden oder für den bewussten, vorübergehenden Verzicht auf die Nutzung sozialer Medien wie Twitter oder Facebook. So wurden in Studien bereits positive Effekte von Digital-Detox-Apps auf problematische, suchtähnliche Nutzungsmuster festgestellt.<sup>29</sup> Auch wird insbesondere von Kindern und Jugendlichen selbst eine Verbesserung des allgemeinen Lebensgefühls berichtet, wenn sie beispielsweise im Ferienlager vorübergehend ganz auf ihr Smartphone verzichten müssen.<sup>30</sup>

*Beispiel 2: Google Glass.* – 2013 führte Google mit Google Glass einen tragbaren Computer in Form einer Brille mit eingebauter Kamera ein. Das System erlaubt eine Vielzahl spannender und innovativer Anwendungen. Die Markteinführung war jedoch von einer breiten Welle an Kritik begleitet, die sich unter anderem an der Sorge von verdeckten Aufnahmen und einer Weiterverbreitung von Videoaufnahmen privater Situationen entzündete. Schliesslich verboten viele Kinos, Bars und Restaurants den Zutritt mit diesem Assistenzsystem und ihre Träger wurden in einem Wortspiel als «(gl)assholes» bezeichnet. Das Gerät wird von Google zwar weiterentwickelt, ist aber nur mehr für Unternehmensanwendungen gedacht.

*Analyse:* Der Verzicht auf digitale Systeme ist zunächst mit einem Verlust der von ihnen gewährleisteten Vorteile verbunden. Weitere Effekte ergeben sich durch erworbene Gewohnheiten, aber auch

<sup>29</sup> Desirée Schmuck: Does Digital Detox Work? Exploring the Role of Digital Detox Applications for Problematic Smartphone Use and Well-Being of Young Adults Using Multigroup Analysis, in: *Cyberpsychology, Behavior, and Social Networking*, 23/8 (2020) 526–532: <http://doi.org/10.1089/cyber.2019.0578>

<sup>30</sup> Sommerlager bieten Kindern digitalen Entzug. Weniger Stress und Mobbing – Jugendliche sehen Erfahrung ohne Smartphone sehr positiv (2019): <https://www.presetext.com/news/20190628001>

durch Abwehrreaktionen der Hersteller, die von einer Systemnutzung profitieren. Ein weiterer Aspekt im Rahmen zunehmender Digitalisierung ist, dass auf etliche Dienste kaum mehr wirklich verzichtet werden kann und eine Trennung in gewollte und ungewollte Nutzungsformen schwierig ist.

## 6. Zusammenfassung

Digitale Assistenzsysteme durchdringen zunehmend alle Lebensbereiche. Sie werden oft in der guten Absicht entwickelt, das Leben zu verbessern, indem Schwächen der Anwender ausgeglichen (Kompensation) oder ihre Fähigkeiten mithilfe der Assistenz erweitert werden (Augmentation). Jedoch wirft die Entwicklung dieser Systeme oft ethische Fragen auf, deren Beantwortung sich bei genauerer Betrachtung herausfordernd gestaltet. Die Schwierigkeit besteht hierbei darin, dass umfangreiche ethische und philosophische Diskurse und die daraus resultierenden abstrakten Grundsätze und Prinzipien in konkrete Merkmale eines zu entwickelnden Systems übersetzt werden müssen. Erschwerend kommt hinzu, dass für diese Übersetzungsleistung in der Praxis oft nur eine begrenzte Zeit und Aufmerksamkeit zur Verfügung steht, weil bereits immer komplexer werdende Technologien und kürzere Entwicklungszyklen diese knappen Ressourcen binden oder einschränken. Um die genannte Herausforderung des ethischen Designs von digitalen Assistenzsystemen leichter zu bewältigen, haben wir daher sechs Muster vorgestellt: Abstufung, Transparenz, Rückkoppelung, Abschalten, Reflexion und Verzicht. Während die ersten vier Muster direkt auf funktionale Merkmale von Assistenzsystemen abzielen, betreffen die beiden letztgenannten Muster eher die Phase der Entwicklung und der Anwendung von Assistenzsystemen. Die vorgestellten Muster erheben keinen Anspruch auf Vollständigkeit. Vielmehr ist es unser Anliegen, aufzuzeigen, dass Muster ein probates Instrument sind, um abstrakte ethische Prinzipien und Grundsätze in leicht handhabbare und praxisrelevante Anforderungen zur

Systemgestaltung zu übersetzen. Sie bilden damit gleichsam einen Nexus, an dem sich ethische Grundsätze und deren Anwendungsregeln (Spezifizierung und Abwägung) mit dem Wissen um konkrete informationstechnische Gestaltungsmöglichkeiten verbinden. Wir hoffen, dass die vorgestellten Muster somit zu einer informierten Gestaltung von Assistenz beitragen, indem sie erwartbare Probleme und Argumente beim ethischen Design vorstrukturieren, aber auch Lösungswissen und Beispiele in kompakter Form offerieren. Bedarf für die zukünftige Forschung liegt einerseits in einer weiteren theoretischen oder empirischen Begründung der Muster, andererseits in der Sammlung weiterer Muster und der Aufbereitung der Sammlung etwa in Form eines strukturierten Katalogs.

conexus 4 (2021) 131–154

© 2021 Clemens H. Cap, Michael Fellmann und Johann-Christian Pöder. Dieser Beitrag darf im Rahmen der Lizenz CC BY-NC-ND 4.0 – Creative Commons: Namensnennung/nicht kommerziell/keine Bearbeitungen – weiterverbreitet werden.



<https://doi.org/10.24445/conexus.2021.04.009>

Prof. Dr. Clemens H. Cap, Universität Rostock, Lehrstuhl für Informations- und Kommunikationsdienste, Albert-Einstein-Strasse 22, 18059 Rostock, Deutschland

[clemens.cap@uni-rostock.de](mailto:clemens.cap@uni-rostock.de)

Prof. Dr. Michael Fellmann. Universität Rostock, Lehrstuhl für Wirtschaftsinformatik, Albert-Einstein-Strasse 22, 18059 Rostock, Deutschland

[michael.fellmann@uni-rostock.de](mailto:michael.fellmann@uni-rostock.de)

JProf. Dr. Johann-Christian Pöder, Universität Rostock, Theologische Fakultät, Universitätsplatz 1, 18055 Rostock, Deutschland

[johann-christian.poder@uni-rostock.de](mailto:johann-christian.poder@uni-rostock.de)