



universität
wien

MASTERARBEIT / MASTER'S THESIS

Titel der Masterarbeit / Title of the Master's Thesis

„Human+“

Technikethische Analyse eines Sicherheitsforschungs-
projekts

verfasst von / submitted by

Christoph Hochwarter, BA BA MA

angestrebter akademischer Grad / in partial fulfilment of the requirements for the degree of
Master of Arts (MA)

Wien, 2022 / Vienna 2022

Studienkennzahl lt. Studienblatt /
degree programme code as it appears on
the student record sheet:

UA 066 641

Studienrichtung lt. Studienblatt /
degree programme as it appears on
the student record sheet:

Interdisziplinäres Masterstudium
Ethik für Schule und Beruf

Betreut von / Supervisor:

Dr. Martin Huth

Plagiatserklärung

Hiermit erkläre ich, die vorgelegte Arbeit selbständig verfasst und ausschließlich die angegebenen Quellen und Hilfsmittel benutzt zu haben. Alle wörtlich oder dem Sinn nach aus anderen Werken entnommenen Textpassagen und Gedankengänge sind durch genaue Angabe der Quelle in Form von Anmerkungen bzw. In-Text-Zitationen ausgewiesen. Dies gilt auch für Quellen aus dem Internet, bei denen zusätzlich URL und Zugriffsdatum angeführt sind.

Mir ist bekannt, dass jeder Fall von Plagiat zur Nicht-Bewertung der gesamten Lehrveranstaltung führt und der Studienprogrammleitung gemeldet werden muss. Ferner versichere ich, diese Arbeit nicht bereits andernorts zur Beurteilung vorgelegt zu haben.

Wien, am 13.4.2022

Christoph Hochwarter

Danksagung

Hiermit möchte ich mich bei meinem Betreuer Dr. Martin Huth für die sehr angenehme, konstruktive Zusammenarbeit und das minutiöse Feedback und die vielen Anregungen bedanken. Ein großes Dankeschön auch für den langen Atem, trotz der beachtlichen Erstellungsdauer der Arbeit – von ersten unverbindlichen Gesprächen zur Möglichkeit einer Betreuung (mit einer anschließenden, berufsbedingt längeren Funkstille meinerseits) über die Erstellungsphase bis hin zur Finalisierung – stets als Ansprechpartner zur Verfügung gestanden zu haben!

INHALTSVERZEICHNIS

1	Einleitung und Problemaufriss	7
2	Skizzierung der Human+ Technik	12
3	Technikfolgenabschätzung und Technikethik	16
3.1	Technikfolgenabschätzung und ihre Strömungen	16
3.2	Technikethik.....	21
3.3	Konvergenz von Technikfolgenabschätzung und Technikethik	24
4	Negativer Utilitarismus	29
4.1	Utilitarismus – Grundlagen und Kritik.....	29
4.2	Negativer Utilitarismus und die Asymmetrie von Leid und Glück	36
5	Ansätze und Frameworks in der Technikethik	41
5.1	Pluralität von Ansätzen und Methoden.....	41
5.2	Principlism.....	44
5.2.1	Begriffsbestimmung – Principlism und Prinzipien mittlerer Reichweite.....	44
5.2.2	Der Vier-Prinzipien-Ansatz von Beauchamp und Childress	46
5.2.3	Ethical Matrix.....	54
5.3	ETICA.....	59
5.4	eTA – Ethical Technology Assessment	62
5.5	Ethical Impact Assessment	64
5.6	Techno-ethische Zukunftsszenarien	68
5.7	ATE-Framework.....	71
5.8	Menlo Report.....	75
5.9	DESSI	79
5.10	Value Sensitive Design	82
5.11	5-Schritte-Schema ethischer Entscheidungsfindung	87
6	Bewertung von Human+	92
6.1	Einleitung	92
6.2	Harmonisierung des Bewertungsframeworks	93
6.3	Prinzipien mittlerer Reichweite des Human+ Bewertungsframeworks	95
6.4	Analyse des Ist-Zustandes (Schritt 1)	99
6.4.1	Harte Fakten und Kontextsensibilität.....	99
6.4.2	Stakeholder-Identifizierung.....	109
6.5	Hauptfrage 1 (Primärer Zweck) – Schritt 2 bis 4.....	111
6.5.1	Unmittelbarer Zweck – Argumente pro und kontra.....	111

6.5.2	Unmittelbarer Zweck – Evaluation und Entscheidung	112
6.5.3	Mittelbarer Zweck – Argumente pro und kontra	113
6.5.4	Mittelbarer Zweck – Evaluation und Entscheidung.....	122
6.5.5	Zwischenfazit	124
6.6	Hauptfrage 2 (ermöglichte außerplanmäßige Anwendungen) – Schritt 2 bis 4.....	125
6.6.1	Klärung der Möglichkeiten außerplanmäßiger Handlungen	125
6.6.2	Maßnahmen zur Verhinderung außerplanmäßiger Handlungen	128
6.7	Hauptfrage 3 (Folgen und Nebenfolgen der planmäßigen Anwendung) – Schritt 2 bis 4..	132
6.7.1	Argumente pro und kontra.....	132
6.7.2	Evaluation und Entscheidung	146
6.8	Empfehlungen (Schritt 5 Implementierung)	157
7	Schluss	160
8	Literaturverzeichnis.....	168
9	ANHANG	178
9.1	Abstract in deutscher Sprache	178
9.2	Abstract in English.....	178

Tabellenverzeichnis

Tabelle 1: EM Beispieltabelle 1	56
Tabelle 2: EM Beispieltabelle 2	56
Tabelle 3: Ethical Impact Assessment: Prinzipien, Kategorisierungen und Beispielfragen.....	66
Tabelle 4: Anticipatory Technology Ethics (ATE): Checkliste	74
Tabelle 5: Prinzipien und Beispielfragen im Menlo-Report	78
Tabelle 6: Bewertungsdimensionen im DESSI-Framework zu „Fundamental Rights and Ethics“	81
Tabelle 7: Value Sensitive Design - Wertetabelle.....	83
Tabelle 8: Schritte im Schema ethischer Entscheidungsfindung nach Bleisch und Huppenbauer	88
Tabelle 9: Human+ Bewertungsframework: verwendete Prinzipien mittlerer Reichweite	98
Tabelle 10: Verwendete Begriffe im Themenfeld Flucht und Migration	101
Tabelle 11: Stakeholder der Human+ Technik	110
Tabelle 12: Überblick über außerplanmäßige Handlungen	128
Tabelle 13: Ethical Matrix zu Folgen und Nebenfolgen der planmäßigen Technikverwendung	134

Abbildungsverzeichnis

Abbildung 1: ETICA: “Analytical Grid for Discourse Analysis” (Quelle: Stahl 2011: 143)	60
Abbildung 2: Human+ Bewertungsframework: Ablaufschema.....	95

Abkürzungsverzeichnis

AAO	Allgemeine Aufbauorganisation
BAO	Besondere Aufbauorganisation
BOS	Behörden und Organisationen mit Sicherheitsaufgaben
COC	Code-of-Conduct
ELSI	Ethische, legistische, soziale/soziologische Implikationen
EM	Ethical Matrix
GSK	Geistes-, sozial- und Kulturwissenschaftlich
IFES	Institut für empirische Sozialforschung
NU	negativer Utilitarismus
PU	positiver Utilitarismus
TA	Technikfolgenabschätzung bzw. technology assessment
TB	Technikbewertung
TE	Technikethik
TRL	Technology Readiness Level
UAV	unmanned-aerial-vehicle ("Drohne")
VSD	Value Sensitive Design
VOST	Virtual Operations Support Team

1 EINLEITUNG UND PROBLEMAUFRISS

Die vorliegende Masterarbeit, verfasst im interdisziplinären Masterstudium „Ethik für Schule und Beruf“, verfolgt zwei Ziele: ein inhaltliches und ein methodisches Ziel. Das **inhaltliche Ziel** ist die Bewertung des KIRAS-Sicherheitsforschungsprojekts Human+ aus technikethischer Perspektive bzw. der Perspektive der Technikfolgenabschätzung. Das zweite, **methodische Ziel**, ist die Erarbeitung eines Bewertungsframeworks, welches zum einen für die ethische Beurteilung von Human+ herangezogen werden kann, zum zweiten aber auch hinreichend generisch ist, um für andere Sicherheitsforschungsprojekte, mit denen der Verfasser der vorliegenden Arbeit in seiner beruflichen Praxis konfrontiert ist, geeignet zu sein. Um diese Ziele verständlicher zu machen, ist es notwendig, bereits an dieser Stelle in aller Kürze Human+ zu charakterisieren.

Human+ (vgl. den Projekteintrag auf der Website der österreichischen Forschungsförderungsgesellschaft FFG 2018) ist ein deutsch-österreichisches Forschungsprojekt (Laufzeit: 2018-2020), das auf österreichischer Seite im Rahmen des KIRAS Forschungsförderungsprogrammes finanziert wurde. Auf deutscher Seite erhielt das Projekt Fördermittel durch das Bundesministerium für Bildung und Forschung (BMBF) im Rahmen des Programms „Forschung für die zivile Sicherheit“ (vgl. DHPol 2021). Die Programmverantwortung für das (österreichische) KIRAS-Programm lag seit dem Antritt der Regierung Kurz II beim Bundesministerium für Landwirtschaft, Regionen und Tourismus (BMLRT), davor beim Bundesministerium für Verkehr, Infrastruktur und Technologie (BMVIT). Human+ befasste sich mit der Entwicklung von digitalen Techniken und Services vor dem Hintergrund der Erfahrungen von Organisationen und Behörden mit Sicherheitsaufgaben (BOS) während der „Flüchtlingskrise von 2015/2016“.¹ Die wissenschaftlich-technischen Ziele des Human+ Projekts lassen sich gemäß FFG Projekteintrag folgendermaßen umreißen:

- Detektion und Vorhersage von Migrationsbewegungen, Identifikation von Migrationsschwerpunkten, umfassende Betrachtung von ELSI-Aspekten
- Lagebilderstellung: Erstellung eines echtzeitnahen Lagebildes, echtzeitnahe Datenanalyse (soziale Medien und Bilddaten), Erstellung kurzfristiger Prognosen

¹ Eine Vorbemerkung: Streng genommen muss bereits die Art der Bezeichnung der Personen, deren Migrationsbewegungen in 2015/2016 den Anlass für das Human+ Projekt gab, Teil der ethischen Analyse sein. Vorausgreifend kann schon jetzt gesagt werden, dass in der vorliegenden Arbeit statt „Flüchtling“ überwiegend der häufig als „neutraler“ wahrgenommene Begriff „Geflüchtete/r“ verwendet wird. „Flüchtling“ wird dann genutzt, wenn der Begriff Teil eines Terminus Technicus ist, z.B. hat sich während des Projekts gezeigt, dass auch innerhalb der bei der Bewältigung der Einsatzlage von 2015/2016 beteiligten BOS häufig von „Flüchtlingslage“ die Rede war. Ähnlich wird verfahren mit „Flüchtlingslager“, „Flüchtlingsbewegungen“, „Flüchtlingskrise“. Für eine Betrachtung der möglichen Terminologie vgl. auch Abschnitt 6.4.1.1 ab S. 99.

- Information und Entscheidungsunterstützung: nutzergruppenabhängige, interaktive Informationsvisualisierung, (semi-) automatisierte Qualitätskontrollen, Sicherstellung der Interoperabilität mit Hilfe eines labornahen Versuchsaufbaus und einem Übungsszenario

Das übergeordnete, dezidierte Ziel des Projekts gemäß Projektantrag liegt darin, zur Gewährleistung von humanitärer Sicherheit und einer besseren Vorbereitung auf Migrationswellen beizutragen (vgl. Human+ 2016: 1).

Die wissenschaftlich-technischen Ziele von Human+ berühren zahlreiche ethische und rechtlich heikle Themenbereiche. Die vorliegende Masterarbeit befasst sich folglich mit der Untersuchung der ethischen Zulässigkeit des oben nur kurz skizzierten Projekts. Sie versteht sich als Beitrag zur umfassenden Betrachtung von ELSI-Aspekten. Der Verfasser der vorliegenden Masterarbeit hat während der gesamten Projektlaufzeit von März 2018 bis Juni 2020 als wissenschaftlicher Mitarbeiter am Institut für empirische Sozialforschung (IFES), welche als geistes-, sozial- und kulturwissenschaftliche (GSK) Partnerinstitution im Human+ Projektkonsortium teilgenommen hat, mitgewirkt und war dabei vornehmlich mit der Erarbeitung von soziologischen und ethischen Implikationen der fraglichen Techniken befasst. Die vorliegende Masterarbeit baut somit auf den vom Verfasser bereits während der Projektlaufzeit durchgeführten Vorarbeiten auf, und dient dazu, diese analytisch zu vertiefen und auf eine noch stärker reflexionswissenschaftliche Basis zu stellen.

Die **übergeordnete Forschungsfrage**, die sich auf das **inhaltliche Ziel** der Arbeit bezieht, lautet somit: Inwiefern sind die angestrebte Human+ Technik bzw. ihre Teilkomponenten ethisch zulässig bzw. nicht zulässig?

Für die Beantwortung dieser sehr umfassenden Frage ist es notwendig, die Themenstellung in kleinere Fragestellungen aufzubrechen. Angelehnt an Fenner (2010: 228) soll die Bewertung der ethischen Zulässigkeit sich hauptsächlich an drei Leitfragen (im Folgenden als „**Hauptfragen**“ bezeichnet) orientieren. Die Beantwortung dieser Leitfragen wird anhand eines noch zu entwerfenden Bewertungsframeworks durchgeführt, welches sich, soviel sei vorausgeschickt, stark an einem Ansatz von Prinzipien mittlerer Reichweite orientieren wird. Die zu beantwortenden Hauptfragen sind folgende:

Hauptfrage 1: Sind die Handlungsmöglichkeiten, die durch diese Technik eröffnet oder verbessert werden, an sich ethisch legitim? (primärer Zweck)

Hauptfrage 2: Werden ausreichend Maßnahmen ergriffen, um alle moralisch bedenklichen außerplanmäßigen Handlungsweisen auszuschließen, die sich damit abseits vom primären Zweck durchführen lassen? (Dies setzt allerdings die Beantwortung von 2 Unterfragen voraus, nämlich Unterfrage 2a: Welche außerplanmäßigen Handlungsweisen abseits vom primären Zweck lassen sich durchführen? Sowie Unterfrage 2b: Welche davon sind moralisch bedenklich?)

Hauptfrage 3: Gefährden die Folgen und Nebenwirkungen der planmäßigen Techniknutzung die Lebensbedingungen oder die Lebensqualität gegenwärtiger oder zukünftiger Menschen?

Erschwert wird eine ethische Bewertung im obigen Sinne dadurch, dass es nicht die „eine, allgemein anerkannte Moral“ gibt, sondern nur „[...] recht verschiedenartige moralische Orientierungskomplexe und Werthaltungen, die teils beziehungslos nebeneinander, teils aber auch in offenem Konflikt miteinander existieren“ (Ropohl 2016: 133). Kurz gesagt: Keine moralphilosophische Konzeption hat einen tatsächlichen breiten Konsens, obwohl es Ansätze gibt, die eine Art „ethischen Minimalkonsens“ skizzieren (vgl. Ropohl 2016: 316) oder aber ethische Prinzipien postulieren, deren Gültigkeit mit ganz unterschiedlichen moralphilosophischen Systemen kompatibel sei (zumindest nach Ansicht ihrer ProponentInnen), und die in einer Art gemeinsamen Moral verankert liegen (dies unternimmt etwa der Principlism von Beauchamp und Childress [2009], welcher eine zentrale Rolle in der vorliegenden Arbeit spielen wird). Manche normative Zukunftsvorstellungen, die für eine Technikbewertung nötig sind (etwa Vorstellungen von einer wünschenswerten zukünftigen Gesellschaft, an deren Zielerreichungsbeitrag fragliche Techniken gemessen werden können) scheinen allerdings diskursbehaftet zu bleiben. Wenngleich das Fehlen eines allgemein verbindlichen ethischen Fundaments gemäß Grunwald (1999: 173) auch im Gegenteil erst als Voraussetzung für ethische Überlegungen gelten kann (denn sonst gäbe es keinen Bedarf an einer ethischen Reflexionsdisziplin, die sich mit sonst ja nicht existierenden konfligierenden moralischen Einstellungen befasst), bedeutet dies dennoch, dass es kein bloß anzuwendendes „Patentrezept“ für obiges Vorhaben gibt.

Es braucht daher zur Beantwortung dieser obigen Fragen einen erst zu schaffenden klaren und transparenten reflexionstheoretischen Rahmen, um die Bewertungen und die jenen zu Grunde liegenden moralphilosophischen Ausgangspunkte sowohl nachvollziehbar wie auch kritisierbar zu machen – das schon erwähnte **methodische Ziel** der vorliegenden Arbeit. Bestandteile dieses Bewertungsframeworks werden, um einen kleinen Ausblick zu geben, einerseits der schon erwähnte „Principlism“ darstellen (vgl. für einen konzisen Überblick über historische Wurzeln, Kernelemente und übliche Kritikpunkte des Principlism: Rauprich 2012), daneben auch eine besondere Spielart des Utilitarismus (nämlich der negative Utilitarismus, vgl. für einen konzisen Überblick Akerma 2000). Zudem werden zahlreiche weitere, v.a. im Bereich der IKT-bezogenen Angewandten Ethik geläufigen Bewertungsframeworks herangezogen, vorgestellt, und für die vorliegende Fragestellung relevante methodische und inhaltliche Aspekte übernommen und in ein harmonisiertes Framework zusammengeführt.

Die Arbeit gliedert sich in folgende weitere Abschnitte:

In Kapitel 2 wird zunächst eine kurze Skizzierung der Human+ Technik vorgenommen, um einen Eindruck von der zu analysierenden Technik zu bekommen. Im Kapitel 3 werden die Disziplinen der Technikfolgenabschätzung und der Technikethik erläutert, um die vorliegende Arbeit, so ein Vorgriff auf das Fazit, in einem Schnittmengenbereich beider Disziplinen zu verorten. Danach folgen Kapitel, die sich mit Konzepten beschäftigen, die in das Bewertungsframework einfließen werden. Zunächst wird in Kapitel 4 der Negative Utilitarismus (NU) beleuchtet, indem der klassische Utilitarismus skizziert wird, um dann in einer Auseinandersetzung mit dem sich davon absetzenden NU das sogenannte Metaprinzip des Negativen Utilitarismus zu formulieren. In Kapitel 5 werden weitere Ansätze und Frameworks, wie sie in der Technikethik geläufig sind, vorgestellt, um mögliche weitere Elemente für das zu erstellende Bewertungsframework zu sondieren. Begonnen wird dabei mit einer grundlegenden Beleuchtung des Principlism von Beauchamp und Childress, welcher sich auch bei zahlreichen anderen der in jenem Kapitel dargestellten Frameworks im Hintergrund finden wird. Im Anschluss an diese Erarbeitung der Literatur wird in Kapitel 6 die „eigentliche“ Bewertung von Human+ vorgenommen. Dabei wird zunächst das Bewertungsframework aus den in den vorherigen Kapiteln erläuterten Ansätzen harmonisiert und dargestellt, und anschließend zur Anwendung kommen. Dabei werden auch die schon in der Einleitung vorgestellten Hauptfragen beantwortet. Kapitel 7 wird zusammenfassend noch einmal sowohl die Ziele der Arbeit wie auch die Erreichung dieser Ziele reflektieren (und so auch zu einem differenzierten Fazit über die ethische Zulässigkeit der Human+ Technik gelangen), und auf methodische und inhaltliche Beschränkungen der Arbeit verweisen.

Da zahlreiche der verwendeten Texte in englischer Sprache abgefasst sind, wird häufig in eckigen Klammern zum besseren Verständnis der Originalbegriff hinter die deutsche Übersetzung gestellt.

Diese Einleitung beschließend sei noch auf das Selbstverständnis der vorliegenden Arbeit verwiesen. Zwar werden potentielle ethische Konfliktpunkte identifiziert und auch Handlungsempfehlungen gegeben, wie diesen Konflikten begegnet werden könnte, aber: „Ethik dient nicht dazu, dem Menschen zu sagen, was richtig oder falsch ist. Ethik ist die notwendige und methodische Reflexion auf moralische Normen und Vorstellungen. Es ist aber nicht die Aufgabe der Ethik, Entscheidungen vorwegzunehmen“, so Martin Huth im Zuge eines gemeinsamen Workshops zwischen der österreichischen UNESCO-Kommission und der Vetmeduni Wien (UNESCO 2019). Ähnlich sehen es viele VertreterInnen, die im Bereich der Technikfolgenabschätzung tätig sind: „Eine Technikfolgenabschätzung nimmt Entscheidungen keinesfalls vorweg sondern trägt – wie auch andere Instrumente der Zukunftsforschung [...] – relevante Informationen zusammen und schafft eine zum Entscheidungszeitpunkt optimal mögliche Informationsgrundlage“ (Zweck 2013: 145). Das bedeutet, dass ethische Untersuchungen zwar Orien-

tierungswissen für notwendige Entscheidungen bereitstellen können, den jeweiligen EntscheidungsträgerInnen (wie etwa im Human+ Fall diejenigen, die über eine Anwendung, Weiterentwicklung oder Anschaffung von Human+ Komponenten entscheiden) diese aber nicht abnehmen können und wollen.

2 SKIZZIERUNG DER HUMAN+ TECHNIK

Dieser Abschnitt umreißt in grundlegenden Zügen, worum es bei der Human+ Technik eigentlich geht. Die Kenntnis der einzelnen Komponenten, und die Vertrautheit der damit verbundenen Bezeichnungen (z.B. In-Situ-Daten, Social-Media-Daten) ist essentiell für das Verständnis der ethischen Analysen in den unteren Bereichen dieser Arbeit.

Der avisierte Funktionsumfang von Human+ ist in einer Reihe von projektinternen Dokumenten festgehalten. Der Projektantrag (Human+ 2016) hat in sehr allgemeiner Weise das übergeordnete Ziel von Human+ skizziert (ungefähr so wie es schon in der Einleitung dargestellt wurde). Diese sehr allgemeinen Leitlinien wurden in einer frühen Projektphase in der sogenannten „Anforderungsanalyse“ genauer spezifiziert, indem MitarbeiterInnen von BOS auf deutscher und österreichischer Seite zu den notwendigen technischen und funktionalen Anforderungen an das Human+ System zur Erreichung der Ziele befragt wurden.² Die Ergebnisse dieser Auswertung mündeten im sogenannten „Kumulativen Anforderungskatalog aller Bedarfsträger an ein Informations- und Darstellungssystem“ (Human+ 2018a und 2018c). Diese Anforderungen wurden im nächsten Schritt hinsichtlich ihrer Machbarkeit im Rahmen des Projekts bewertet und in ein technisches Pflichtenheft überführt (Human+ 2018d).

Die Human+ Technik soll laut Antrag, Anforderungsanalyse und technischem Pflichtenheft in ihrem Kern folgendem Leistungsumfang entsprechen, welcher in Form eines Softwarepakets und der Entwicklung der hierzu notwendigen Algorithmen bereitgestellt werden soll:

- Erstellung eines gemeinsamen Echtzeitlagebildes, auf welches unterschiedliche BOS Zugriff haben und welches sie gemeinsam bespielen können sollen (etwa in Form einer digitalen Karte mit unterschiedlichen Informations-Layern und Zusatzfunktionen, in welchem aggregierte Informationen zu prognostizierten Bewegungen von Gruppen von Geflüchteten, gemeinsamen Ressourcen, allgemeinen Informationen, usw. dargestellt werden können)
- Erstellung von Migrationsprognosen zu Geflüchteten, welches sich im Wesentlichen aus vier Informationsquellen speist, nämlich aus der algorithmenunterstützten Analyse von
 - **Social-Media-Daten**³ (im Rahmen des Forschungsprojekts waren dies ausschließlich Tweets auf Twitter);

² Auf deutscher Seite wurden die Rückmeldungen von insgesamt 115 Personen aus dem Bevölkerungsschutz bzw. der polizeilichen Gefahrenabwehr erhoben, entweder über persönliche, telefonische oder Online-Interviews oder als TeilnehmerInnen von regionalen Fokusgruppen. Auf österreichischer Seite wurden n=22 leitfadengestützte Face-to-Face-Interviews mit BOS-VertreterInnen, die in die Bewältigung der „Flüchtlingslage 2015/2016“ involviert waren.

³ Die Sammlung dieser Daten erfolgt über einen sogenannten „Crawler“, d.h. über ein Programm, das sich mit der Twitterprogrammierschnittstelle (REST API) verbindet und in Echtzeit Posts aus sozialen Medien sammelt (vgl. Human+ 2018d: 26). Dabei ist zu bemerken, dass nicht alle Tweets auf diese Weise gesammelt werden können, weil Twitter nur einen Teil über die API anbietet, sodass kein vollständiges Monitoring aller Tweets über die API möglich ist. Die Datensammlung erfolgt somit über die von Twitter zur Verfügung gestellten offiziellen Wege.

- **In-Situ-Daten**, das heißt von hauptsächlich Videostreams oder Videoaufnahmen, z.B. von Kameras oder Unmanned-Aerial-Vehicles (UAV, vulgo „Drohne“), die sich an „Hot Spots“ befinden. Ein hypothetisches Beispiel während der „Flüchtlingskrise 2015/2016“ wäre etwa der Zugriff auf die Daten von Kameras an Grenzübergängen oder am Wiener Westbahnhof gewesen. Diese Videodaten werden ausgewertet hinsichtlich der Anzahl an Personen, die ein bestimmtes Gebiet betreten und wieder verlassen haben, bzw. dort aufhältig sind;
- **Human-Sensors-Daten**, also Informationen, die Human Sensors in das System einspielen (Human-Sensors bezeichnet im Human+ Projekt die MitarbeiterInnen der oft international operierenden Rettungs- und Hilfsorganisationen wie dem Roten Kreuz, die selbst an unterschiedlichen relevanten Orten des Geschehens sein können, z.B. Grenzübergänge oder Flüchtlingslager);
- **Satellitenfernerkundungsdaten** (wobei hier nicht gemeint ist, dass ein Satellit „aktiv“ permanent Daten liefert oder die Bewegung von Geflüchteten erkennt, sondern dass von Satelliten selektiv erstellte Bilder, z.B. von bekannten Flüchtlingslagern, von Human+ Algorithmen analysiert werden, um Rückschlüsse auf Veränderungen in der Belegung zu erlauben).

Im Rahmen des Human+ Projekts wurde keine anwendungsreife Software entwickelt, sondern obige Funktionen wurden als sogenannter „Demonstrator“ von den technischen Konsortialpartnern in Form einer browserbasierten Applikation verwirklicht, welche eine rudimentäre Echtzeitlagendarstellung sowie Migrationsprognose (mit gewissen Abstrichen, wie Vorprozessierung von Analysen bei entsprechend aufbereitetem Datensatz von im Vorfeld gesammelten Daten) erlaubt.

Um zu verstehen, weshalb von einem „Demonstrator“ die Rede ist, muss zunächst auf die im Projektantrag bzw. die im Forschungsförderungsprogramm festgelegte, anzustrebende **Produktreife** eingegangen werden, welche sich am sogenannten **Technology Readiness Level (TRL)** festmachen lässt. Zu Projektende sollte laut Projektantrag eine Technikwicklung auf dem TRL 4 vorliegen, d.h. „Funktionsnachweis der Technologie im Labor(-maßstab) auf Systemebene“ (vgl. FFG 2019: 28). Das TRL ist ein Maß für den Grad der Ausgereiftheit einer Technik, und verläuft von TRL 1 („Nachweis der Grundprinzipien im Sinn einer orientierten Grundlagenforschung“) bis hin zu TRL 9 („System hat sich in Einsatzumgebung bewährt, wettbewerbsfähige Produktion im Fall von Schlüsseltechnologien“), mit entsprechenden unterschiedlichen Zwischenkategorien. Das bedeutet, dass die während des Projekts zu entwickelnde Technik noch weit von einem Echteinsatz oder einer kommerziellen Vermarktung entfernt ist, aber die Funktionsweise der entwickelten Algorithmen sich exemplarisch in einer Demonst-

rationssoftware nachweisen lässt. Das Pflichtenheft, das auf Basis des Anforderungskatalogs entwickelt wurde, skizziert somit einen Funktionsumfang, der sich nicht mit dem Funktionsumfang eines möglicherweise nach Projektende auf Basis der Projektergebnisse noch zur „Marktreife“ weiterentwickelten Produkts deckt, weil bei letzterem zahlreichen weiteren Anforderungen genügt werden muss (z.B. sicherheitstechnischen Aspekten für die Einflechtung in die IT-Infrastruktur der verschiedenen BOS, Aspekte der Anwendungsergonomie, usw.).

Für eine technikethische Bewertung bringt die Bewertung auf TRL4 allerdings zwei Schwierigkeiten mit sich: Erstens heißt das, dass der technische Funktionsumfang, der bewertet werden soll, nicht an einem ausgereiften Produkt festgemacht werden kann, welches man sich „anschauen“ kann, sondern teilweise nur in Programmcode und Konzepten vorhanden ist. Zweitens bringt dieses frühe Stadium mit sich, dass die späteren ganz konkreten Anwendungsbedingungen nur teilweise antizipiert werden können, welche aber einen Unterschied hinsichtlich der ethischen Legitimität machen können. Daher werden im Rahmen der vorliegenden Masterarbeit da, wo unterschiedliche Anwendungsbedingungen in Zukunft zutreffen können, die bereits antizipierbaren auf ihre unterschiedlichen Implikationen bewertet, sodass auch ein Vergleich von verschiedenen Szenarien durchgeführt wird.

In der tatsächlichen, nach Projektende stattfindenden Weiterentwicklung über TRL 4 hinaus ist zwar möglich, dass einzelne Konsortialpartner bzw. Partnerinstitutionen die von ihnen hauptverantwortlich entwickelten Komponenten unabhängig von anderen Paketen zu weiterer Marktreife weiterentwickeln, aber da das vernetzte Zusammenspiel der skizzierten Komponenten spezifische Missbrauchsmöglichkeiten eröffnet, soll im weiteren Verlauf der Arbeit von der komplexeren Situation einer Bündelung obiger Komponenten ausgegangen werden.

Die Human+ Teilkomponenten haben das übergeordnete Ziel, BOS im Falle von ähnlichen Fluchtereignissen wie 2015/2016 in eine bessere Informationslage zu versetzen, wobei im Projektantrag auch vom „Monitoring“ von Bewegungen von Geflüchteten die Rede ist. Der Projektantrag sieht dabei eine verbesserte Informationslage zu drei skizzierten Szenarien vor (vgl. Human+ 2016: viii sowie 1).

- Monitoring von Migrationsbewegungen zwischen Österreich und Deutschland [„Unterstützung von kritischen Inlandssituationen auf Grund großer Mengen von MigrantInnen“];
- Monitoring von Übertritten an EU Außengrenzen, um Ankunftszeiten für Österreich und Deutschland abschätzen zu können [„Optimierung der Situation an europäischen Grenzen“];
- Monitoring internationaler Krisenregionen [„Bewertung von Informationen aus internationalen Krisenregionen“].

Das Spektrum der BOS, die den AnwenderInnenkreis der Human+ Technik bilden, ist dabei breit gefächert. Im Human+ Projektkonsortium waren während der Entwicklung sowohl Rettungsorganisationen

(Johanniter, Bayerisches Rotes Kreuz) als auch staatliche Akteure (Bundesanstalt Technisches Hilfswerk, österr. Bundesministerium für Landesverteidigung, die Polizei in Form der Deutschen Hochschule der Polizei) vertreten. Weitere AnwenderInnen eines potentiellen Echtprodukts würden somit auch weitere Einsatzorganisationen oder auch staatliche Behörden wie Ministerien umfassen können. Dies wird bei der späteren Beurteilung der Human+ noch zu beachten sein.

In den nächsten Abschnitten soll nun nach dieser grundlegenden Gesamtskizzierung der Human+ Technik die Erarbeitung des ethischen Beurteilungsrahmens bzw. Frameworks erfolgen. Den Beginn macht dabei eine Einführung in die grundsätzlichen Ziele und Ansätze der Technikfolgenabschätzung sowie der Technikethik, um den interdisziplinären Kontext begreifbar zu machen, innerhalb dessen dieses Framework entwickelt wird.

3 TECHNIKFOLGENABSCHÄTZUNG UND TECHNIKETHIK

3.1 TECHNIKFOLGENABSCHÄTZUNG UND IHRE STRÖMUNGEN

Auf Grund des interdisziplinären Ansatzes der vorliegenden Arbeit sollen zunächst das Verhältnis von Technikethik und Technikfolgenabschätzung [technology assessment] (TA), geklärt werden.

Da es um die Betrachtung von Konsequenzen und der ethischen Zulässigkeit einer konkreten Technik geht, bewegt sich die vorliegende Arbeit in einem interdisziplinären Feld zwischen Technikfolgenabschätzung/Technikbewertung und Technikethik. Um die möglichen Untersuchungsdimensionen besser zu verstehen, ist es trotzdem hilfreich, zunächst beide Disziplinen idealtypisch zu skizzieren und ihre Hauptkenntnisinteressen zu schildern.

Beim Sammelbegriff TA können mehrere Strömungen unterschieden werden. Während gemeinsamer Bestandteil unterschiedlicher Ansätze in der TA stets Forschung, Bewertung (Bewertungsverfahren) und Beratung in unterschiedlichen Mischverhältnissen sind, lässt sich doch eine Pluralität von Disziplinen unterscheiden, die sich mit anderen Schwerpunkten in unterschiedlichem Umfang an bestimmte Adressatengruppen wenden, was Grunwald auch als „Buntheit und Einheit“ der TA bezeichnet (vgl. für einen Überblick Simonis 2013a: 16; Grunwald 2010: 87-119). Diese Buntheit ist nicht Beliebigkeit, sondern eine Folge unterschiedlicher Anforderungen. So lassen sich etwa exemplarisch Ansätze wie die „partizipative Technikfolgenabschätzung“ nennen, die Legitimationsdefizite technologischer Entscheidungen seitens der Exekutive durch offenere, breite Beteiligungsformen kompensieren will, die „parlamentarische TA“, die sich an gewählte VolksvertreterInnen richtet, die technologische Entscheidungen treffen sollen, oder auch die „strategische TA“ die in an sozialverträglicher Entwicklung interessierten Unternehmen stattfindet. Deziert stärker an ethischen Gesichtspunkten ausgerichtet ist die „rationale Technikfolgenbeurteilung“, die Fragen der prinzipiellen Zumutbarkeit in den Vordergrund stellt. Die „Technikbewertung“ (TB) wiederum fokussiert stärker auf Technikentwicklung in Unternehmen bzw. Normierungsverfahren in wissenschaftlich-technischen Vereinen, um nur eine Auswahl zu nennen.

Außerdem lässt sich TA in historische Entwicklungsstufen unterteilen. War die Aufgabe der Technikfolgenabschätzung zunächst eng an Politikberatung gekoppelt, wurde die Technikfolgenabschätzung zur Technikfolgenforschung erweitert, die sich zunehmend systematisch um die Wirkung von Technik kümmerte (vgl. Kunzmann 2006: 258, der sich auf Petermann 1999 bezieht). Die Technikfolgenforschung wurde um die Technikgeneseforschung erweitert, mit dem Ziel, rechtzeitig auf die Genese bzw. Entwicklung von Technik Einfluss zu nehmen. In Europa wurde die Disziplin zudem immer stärker von diskursiven Aspekten bestimmt, welche darauf abzielen, dass die von einer Technik Betroffenen auch

an der Gestaltung mitwirken können. Partizipation von BürgerInnen und Laien nahm daher einen immer gewichtigeren Stellenwert ein, wofür Verfahren wie etwa Bürgerforen entwickelt wurden.⁴

Eines der am weitesten verbreiteten Dokumente zur Technikfolgenabschätzung ist die VDI- bzw. Verein-Deutscher-Ingenieure-Richtlinie 3780, die das Konzept der „Technikbewertung“ spezifiziert, und deren Ziele an dieser Stelle exemplarisch dargestellt werden sollen. Der Begriff „Technikbewertung“ wurde vom VDI im Rahmen der Auseinandersetzung mit der TA favorisiert, da Technikfolgenabschätzung eine fixierte Perspektive auf kritische Folgen suggeriere, und der alternative Terminus Technikbewertung negativen Assoziationen einer primär risikofokussierten Folgenbetrachtung vorbeugen solle (vgl. Zweck 2013: 147). TB fokussiert sich besonders, aber nicht ausschließlich, auf Technikentwicklung innerhalb betrieblicher Entwicklungssettings (vgl. Zweck 2013: 150). Das Gegenstück zur betrieblichen ist die öffentliche Technikbewertung, etwa durch Ministerien oder Parlamente. Unter TB wird verstanden, in Anlehnung an die VDI-Richtlinie 3780,

„[...] das planmäßige, systematische, organisierte Vorgehen, das

- den Stand einer Technik und ihre Entwicklungsmöglichkeiten analysiert,
- unmittelbare und mittelbare technische, wirtschaftliche, gesundheitliche, ökologische, humane, soziale und andere Folgen dieser Technik und möglicher Alternativen abschätzt,
- aufgrund definierter Ziele und Werte diese Folgen beurteilt oder auch weitere wünschenswerte Entwicklungen fordert,
- Handlungs- und Gestaltungsmöglichkeiten daraus herleitet und ausarbeitet“ (VDI 1991: 2, zitiert nach Zweck 2013: 151)

Während zwar grundsätzlich jegliche Art von Technik auch einem TA-Prozess unterworfen werden kann, zielt TA besonders auf „politikpflichtige“ Aspekte von Technik für die politische Entscheidungen getroffen werden müssen, also auf allgemeinverbindliche Aspekte im Umgang mit Folgen des wissenschaftlich-technischen Fortschritts, wie etwa Grenzwertsetzungen, Sicherheitsstandards, oder Forschungsförderung durch öffentliche Gelder (vgl. Grunwald 2010: 48).

In der TA gibt es eine Reihe von Ablaufschemata, um Projekte zu strukturieren, die aber auf die Spezifika des jeweiligen Ansatzes zugeschnitten sind. Im VDI-Ansatz beispielsweise werden diese Komponenten in Form folgenden Ablaufschemas bewältigt: Definition und Strukturierung des Themas, Folgenabschätzung, Bewertung, Entscheidung (der sogenannte „weite“ Technikbewertungsbegriff), mit

⁴ An dieser Stelle wird bereits deutlich, dass die vorliegende Arbeit keinesfalls all diese Facetten von TA einlösen kann. So wird etwa nicht die Perspektive von durch die Human+ Technik betroffenen MigrantInnen aktiv eingeholt, lediglich durch Rückgriff auf Literatur, sodass etwa hier nicht beansprucht wird, partizipative Technikfolgenabschätzung zu betreiben.

einer nachgelagerten normativen bzw. prognostischen Phase (Folgenabschätzung, welche auch als engere Begriff der Technikbeurteilung bezeichnet wird, vgl. Zweck 2013: 151). Demgegenüber kritisiert Grunwald am Ablaufschema, dass die entscheidende Definitionsphase nicht differenziert sei und stattdessen diese Phase „dezisionistisch“ dem Auftraggeber zugewiesen würde, und dass zudem die Phase der eigentlichen Technikfolgenabschätzung nur grob und unkritisch positivistisch in Folgenforschung und Bewertung aufgeteilt würde (vgl. Grunwald 2010: 122). Detaillierter und die Definitionsphase stärker differenzierend und vorhergehende Untersuchungen einbeziehend ist das Strukturmodell der US-amerikanischen Beratungsagentur MITRE, welches in der (frühen) TA als Idealmodell betrachtet wurde (vgl. hierfür Grunwald 2010: 122), und aufgeteilt wird in

- **Konzeptionsphase:** Themenfestlegung, Präzisierung der Aufgabenstellung, Bestimmung des Untersuchungsumfanges und der Vorgehensweise sowie Bestimmung des Teams und der zu beteiligenden Personen oder Gruppen
- **Systemdefinition:** Bestandsaufnahme der Ist-Situation, Informationsbeschaffung, Beschreibung der Steuerungsgrößen, der gestaltbaren Parameter und der relevanten Rahmenbedingungen, Systemabgrenzung, Systembeschreibung und Modellbildung
- **Potenzialabschätzung:** Stand von Forschung und Entwicklung, absehbare technologische Entwicklungen und ihre Potenziale, Einschätzung der Anwendungsreife, der Anwendungsfelder und der Marktdurchdringung sowie die Betrachtung alternativer und konkurrierender Technikfelder
- **Szenariobildung:** Festlegung des Szenarientyps, Beschreibung von Entwicklungsmöglichkeiten und ihre Bündelung zu in sich konsistenten ‚möglichen Zukünften‘ und Analyse der Konsequenzen und Implikationen, die sich aus den Szenarien ergeben
- **Folgenabschätzung:** Bestimmung der relevanten Folgendimensionen, Untersuchung der Folgen in diesen Dimensionen, Beschreibung daraus sich ergebender möglicher gesellschaftlicher Konfliktfelder oder Akzeptanzprobleme
- **Bewertung:** Auswahl der Bewertungskriterien, Bewertung der Folgen und Erarbeitung von Handlungsalternativen auf der Basis dieser Bewertung

Zentral ist die Feststellung, dass TA kein einheitliches Programm mit einem festen Methoden- oder auch nur Ablaufkanon ist, sondern eher ein Bündel unterschiedlicher Ansätze mit unterschiedlichen Schwerpunktsetzungen und mitunter „best practice“ Ablaufschemata, die unter der Überschrift TA vereint werden. Die gemeinsame normative Basis der TA-Ansätze kann aber auf folgende Kurzformel gebracht werden: „Es geht darum, unter Mobilisierung möglichst des gesamten, zu einem bestimmten Zeitpunkt vorhandenen und generierbaren Wissens, Entscheidungen über den Einsatz oder auch Nicht-einsatz einer Technologie zu ermöglichen, die das Gemeinwohl bestmöglich fördern“ (Simonis 2013a: 11). Der Verweis auf die normative Basis ist wichtig, denn rein empirisch und ohne normative Bezüge

auf das gemeinschaftliche Gemeinwohl könne TA nicht (mehr) betrieben werden (vgl. Simonis 2013a: 13).

In anderen Worten kann das „Minimalziel“ einer Technikfolgenabschätzung auch so beschrieben werden: „Mit einem Technikfolgenabschätzungsprozess verbunden sind die Ziele einer möglichst realistischen Abwägung der Chancen und Risiken sowie eine Formulierung von Optionen zur Minimierung der identifizierten Risiken“ (Zweck 2013: 245). Dieses Anliegen verfolgt auch die vorliegende Arbeit.

Im thematischen Umfeld der Technikfolgenabschätzung wird mitunter auch der Begriff der „**Begleitforschung**“ verwendet, welche zwar nicht ausschließlich TA im engeren Sinne umfasst, aber beträchtliche Überschneidungen aufweist, und mitunter wird TA auch als Begleitforschung verstanden (v.a. in forschungspolitischen Kontexten, vgl. Grunwald 2010: 78f). Darunter fallen erstens Forschungen zu EHS-Themen (environment, health, safety) wie etwa human- und ökologietoxikologische Studien, zweitens Forschungen zu ELSI-Themen (**e**thical, **l**egal, **s**ocial **i**mplications), die von Akzeptanzuntersuchungen über ethischen Reflexionen und Fragen zukünftiger Regulierung bis hin zu Technologievorausschau und Marktpotentialanalysen reichen und sozial- und kulturwissenschaftliche Aspekte umfassen, sowie drittens Untersuchungen der Risikowahrnehmung und Risikokommunikation. Häufig umfassen diese Ansätze also eine „ethical parallel research“, wobei das Ziel ist, dass ethische Forschung direkt Umsetzung in die technische Entwicklung finden – wenngleich dieser Ansatz einer parallelen Forschung aufgrund der Eingebundenheit in das technische Forschungsprojekt wegen möglicherweise zu großer Nähe zu den EntwicklerInnen der Technik und damit einer fehlenden Unabhängigkeit kritisiert wird: „Eine Technikfolgenabschätzung als enge Begleitforschung wäre Teil eines Entwicklungsprojekts, nicht mehr unabhängiger externer Beobachter und Berater – mit den entsprechenden Vor- und Nachteilen“ (Grunwald 2010: 79). Die Vorarbeiten des Autors dieser Masterarbeit, die in Form eines dediziert als ELSI-Arbeitspaket titulierten Projektabschnitts während der Laufzeit des Human+ Projekts durchgeführt wurden, müssen sich auch dieser Form der Kritik aussetzen, die allerdings dadurch gemildert wird, dass die eigentliche akademische Vertiefung hierzu nach offizieller Beendigung des Human+ Projekts stattgefunden hat, und somit zumindest in den letzten Zügen keine direkten Verbindungen mehr gegeben waren.

Die oben angeführten Ziele und Ansprüche umfassender TA einzulösen kann sehr ressourcenintensiv sein. Ropohl schätzt, dass für eine umfassende Technikbewertung mindestens fünf Personen zwei Jahre lang mit einem Forschungsbudget von eineinhalb Millionen DM⁵ „ungestört an einem Thema arbeiten“ müssen (vgl. Ropohl 2016: 213). Auf welche Art von Technik bezieht sich Ropohl bei dieser pauschalen Kostenprognose? TA bezieht sich nach Ropohl nicht auf eine abgegrenzte Sachtechnik,

⁵ Die Erstauflage von Ropohl (2016) ist aus dem Jahr 1996, daher die Angaben in DM statt in Euro.

sondern auf die Dynamik öko-sozio-technischer Systeme (vgl. Ropohl 2016: 212). Besieht man sich die Beispiele genauer, mit denen Ropohl arbeitet, so beziehen sich diese zwar zunächst auf vermeintlich abgegrenzte Thematiken wie die Analyse von Großraumbüros (vgl. Ropohl 2016: 42ff) oder von neuen Techniken im Gebäudebereich (z.B. die Wärmepumpe als alternative Heizmöglichkeit). Diese architektonischen oder technischen Thematiken müssen dann allerdings als soziotechnische Systeme in den Blick bekommen werden, sodass z.B. nicht die Sachtechnik Wärmepumpe analysiert wird, sondern insgesamt das soziotechnische System „Wohnung“. Bei Mobiltelefonen dürfe nicht nur die physische Auswirkung von elektromagnetischen Feldern, sondern die gesamten Verflechtungen im soziotechnischen Kommunikationssystem und zum sozialen System betrachtet werden. Auf diese Weise wird sichtbar, dass sich die vermeintlich hohen Kosten für umfassende TA nicht erst bei den klassischen Anstoßgebern der TA, wie etwa Atomkraftwerken, ergeben. Lingner gibt für den Ansatz der stärker ethisch orientierten „rationalen Technikfolgenbeurteilung“ ebenfalls einen Zeitraum von in etwa zwei bis drei Jahren mit fünf bis zehn ProjektmitarbeiterInnen an (vgl. Lingner 2013: 85). Lingner bezieht sich in seiner Schätzung auf „aktuelle Probleme der Technikdebatte oder eine neue Entwicklung, deren absehbare Konsequenzen hinreichend gesellschaftlich relevant werden können“ (Lingner 2013: 84). Ein Beispiel hierfür ist etwa „Synthetische Biologie“, welche nach Lingner verspreche, neuartiges Leben zu schaffen. Auch wenn sich Ropohl und Lingner in den konkreten Beispielen wie auch den Aufwandschätzungen unterscheiden: Die von ihnen als Faustregel veranschlagten Ressourcen standen für die Betrachtung von ethischen und soziologischen Aspekten während der Laufzeit von Human+ nicht zur Verfügung, weshalb auch in der analytischen Aufbereitung und Vertiefung in Form dieser Masterarbeit diese oben skizzierte, umfassende „Idealvorstellung“ einer Technikbewertung nicht eingelöst werden kann, sondern bestimmte Schwerpunkte gesetzt werden. Dies verschärft ein grundsätzliches Problem der Technikfolgenabschätzung, denn die Forderung nach umfassender Betrachtung von Technikfolgen (inkl. der Beteiligung aller Betroffenen, dem Gehört-Werden aller Expertenpositionen etc.) lässt sich aus ökonomischen, praktischen und wissenschaftstheoretischen⁶ Gründen nie völlig einlösen, sehr wohl muss aber die Auswahl von zu betrachtenden Dimensionen begründet und transparent gemacht werden (vgl. Grunwald 2010: 164).

Für die vorliegende Arbeit bedeuten diese Ausführungen, in der Erarbeitung des Beurteilungsframeworks besonders transparent die Bewertungsdimensionen (d.h. den Ablauf des Schemas und die zu Grunde gelegten Prinzipien) klarzumachen. Während die vorliegende Arbeit auf Grund obiger Beschränkungen nicht einlösen kann, allen Anforderungen an eine „ideale“ TA genügen zu können (welche wie erwähnt ohnehin nie völlig einlösbar sind), beansprucht sie doch Orientierung an etwas, was

⁶ So erzeugt das Generieren von Wissen über einen Sachverhalt durch Forschung ständig neue Anknüpfungspunkte für Folgeforschung, sodass streng genommen, nimmt man den Begriff „umfassende Betrachtung“ ernst, dieser Prozess nie abgeschlossen werden kann.

lose in Anlehnung an Grunwald als „TA-Denken“ bezeichnet werden könnte. Denn „[...] Technikfolgenabschätzung bzw. das hinter ihr stehende Denken [ist] in viele Anwendungsbereiche und wissenschaftliche Disziplinen hinein diffundiert, ohne dass es immer so bezeichnet wird“ (Grunwald 2010: 309).

3.2 TECHNIKETHIK

Wie bereits sichtbar wurde, enthalten Arbeiten im Bereich der TA oft Anteile an sozialwissenschaftlichen, psychologischen oder ökonomischen Untersuchungen. Worin liegen aber Ähnlichkeit und Unterschied zu Technikethik?

Um erläutern zu können, womit sich Technikethik befasst, muss zunächst der im Alltag oft als selbstverständlich gebrauchte Terminus Technik geklärt werden. Fenner stellt ihrer Definition von Technikethik zunächst eine Typologie möglicher Bedeutungen des Begriffs Technik voran (vgl. Fenner 2010: 213): Technik (1) als **Realtechnik** (Gesamtheit von Werkzeugen und Maschinen zur Erreichung praktischer Ziele wie Umwandlung, Transport, Speicherung von Stoff, Energie oder Information), Technik (2) als **Menge aller Handlungen im Umgang** mit solchen Artefakten (wie Herstellung oder Verwendung), sowie Technik (3) als **Wissen um technische Verfahren und um die Herstellung, Gebrauch und die Reparatur technischer Artefakte** („Technologie“ im engeren Sinne). Dies ähnelt dem „Technikbegriff mittlerer Reichweite“ von Ropohl (2016: 84), der neben nutzenorientierten, gegenständlichen Gebilden (Artefakten, Sachsysteme) auch diejenigen Handlungen einbezieht, durch die diese Sachsysteme hervorgebracht werden oder durch die diese verwendet werden (also das Herstellungs- und Gebrauchshandeln).

Technikethik kann demzufolge definiert werden als „Bereichsethik, die sich mit den ethischen Problemen bezüglich der Bedingungen, Ziele und Folgen der Herstellung, Nutzung und Entsorgung von Technik (1) befasst“ (Fenner 2010: 215). Fragen zu Wissen und Know-How, also Technik (3), ähneln laut Fenner sehr Fragestellungen im Bereich der Wissenschaftsethik, auch wenn Fenner diese nicht aus dem Bereich der Technikethik ausklammert. In einer allgemeineren Fassung wird Technikethik auch definiert als „[...] prinzipielle Reflexion auf die moralischen Fragen, welche die Technik aufwirft“ (Kunzmann 2006: 249). Adressaten der Technikethik sind u.a. Ingenieure und TechnikerInnen selbst, EntscheidungsträgerInnen in der Politik und Wirtschaft, sowie letztendlich alle potentielle TechniknutzerInnen (oder, wie man noch ergänzen könnte, auch die von einer Technik lediglich Betroffenen) in der breiten Öffentlichkeit (vgl. Fenner 2010: 215). Nagenborg unterscheidet die Ebene der (Forschungs-)Politikberatung und die Ebene der konkreten Technikgestaltung, in denen Ethik im Bereich Technik tätig werden kann (vgl. Nagenborg 2014: 243). Terminologisch versucht Nagenborg den Terminus „Technikethik“ im Sinne einer Bindestrichethik zu vermeiden, da er befürchtet, dass dies durch Miss-

verständnisse im Rahmen interdisziplinärer Zusammenarbeit oder durch unangemessene Gewichtungen zum Ausschluss der Betrachtung anderer relevanter Rahmenbedingungen, wie etwa wirtschaftlicher Rahmenbedingungen, führen könnte (vgl. Nagenborg 2014: 246). Für den Kontext der vorliegenden Arbeit sollte durch die obige, über bloße Realtechnik hinausgehende Definition von Technik allerdings der übergreifende Charakter der nachfolgenden Reflexionen hinreichend klargestellt worden sein.

Dass eine Ethik der Technik überhaupt möglich bzw. sinnvoll ist, musste in der Geschichte der Philosophie der letzten Jahrzehnte erst argumentativ gezeigt werden. „Technik ist an sich weder gut noch böse, aber zum Guten und Bösen zu gebrauchen“ (Jaspers 1952: 149). Diese „**Neutralitätsthese**“ besagt demnach, dass technische Geräte und Verfahren (Realtechnik), Handlungen des Entwickelns und Herstellens technischer Artefakte, sowie Wissen um die Herstellung, den Gebrauch und die Reparatur technischer Artefakte ethisch neutral wären, was weniger Raum für technikethische Überlegungen lässt (vgl. Fenner 2010: 224). Gegen die Möglichkeit einer Technikethik würde auch das Zutreffen der These eines „**technologischen Determinismus**“ sprechen, wonach technische Entwicklungen und ihre Folgen einer eigenen Gesetzmäßigkeit folgen, und von menschlichem Wollen und Planen nicht beeinflusst werden können (oder sollten; vgl. Ropohl 2016: 21). Nach Grunwald können beide Thesen mittlerweile sowohl aus handlungstheoretisch-philosophischer Sicht wie auch aus der Perspektive der empirisch-sozialwissenschaftlichen Technikforschung als widerlegt gelten (vgl. Grunwald 1996: 443). So greift etwa Fenner die Neutralitätsthese durch den Hinweis an, dass etwa Artefakte oder Verfahren zum Erreichen von bestimmten „Zwecken“ gemacht werden, also Ergebnisse von subjektiven Tätigkeiten des Planens und Herstellens von TechnikgestalterInnen sind, die Verwendungszwecke vor Augen haben, und keineswegs „neutral“ sind, Kunstwerke die als Selbstzwecke geschaffen werden ausgenommen (vgl. Fenner 2010: 225). In seiner Analyse der Gegenbewegungen zum technologischen Determinismus wiederum verweist Ropohl etwa auf die in den 70er Jahren (in Deutschland) in Fahrt gekommene Ökologische Bewegung, die ausgehend von der Auseinandersetzung mit Atomkraftwerken zu einer umweltpolitisch begründeten Technikkritik gelangte und sich als politische Partei etablierte (vgl. Ropohl 2016: 26ff). Diese Entwicklung machte der Öffentlichkeit den Umstand breit bewusst, dass es Alternativen im Technisierungsprozess gibt, die keinem Sachzwang, sondern einer wertenden Wahl unterliegen.

Technikethische Reflexionen sind nicht (nur) ein Selbstzweck, sondern Ethik kann, etwa bei offenen Konflikten zwischen beteiligten AkteurInnen in heterogenen Kooperationen (wie sie interdisziplinär angelegte Technikentwicklungsprojekte sind), die sich aus abweichenden moralischen Einstellungen der Projektbeteiligten ergeben oder von außen an der Projekt herangetragen werden, dazu beitragen, die Ursachen des Konflikts zu identifizieren. Sie kann im Anschluss dann entweder zu einer Lösung oder

aber wenigstens zu einem begründeten Dissens beizutragen, indem etwa durch Begriffsanalyse eine Art moderierende Funktion eingenommen wird (vgl. Nagenborg 2014: 247-250). Zudem kann Ethik dazu beitragen, Wert- und Normorientierungen von beteiligten AkteurInnen bewusst zu machen, indem „dichte Szenarien“ entworfen werden, die zeigen, wie sich Technik auf das Leben oder die soziale Praxis von Betroffenen auswirken kann, oder die Frage aufgeworfen wird, welche Art von Gesellschaft angestrebt werden soll. Nagenborg verweist hier auf die notwendige Interdisziplinarität eines solchen Unterfangens, wenn etwa auch Ergebnisse aus unterschiedlichsten Disziplinen wie der Techniksoziologie oder der kulturwissenschaftlichen Technikforschung hierfür einbezogen werden müssen, um über die Aussagen der beteiligten ProjektpartnerInnen bzw. der im Projektteam beteiligten Disziplinen hinauszugehen, und so nicht unkritisch einseitige Sachverhaltsdarstellungen zu übernehmen. VertreterInnen der Ethik können hier also bei der Wahl zwischen moralisch unterschiedlich zu bewertenden Alternativen helfen, und noch im Konstruktionsprozess beraten und die Möglichkeit der Auswahl zwischen Alternativen einfordern. Nagenborg (2014: 247) wendet kritisch ein, dass Ethik jedenfalls nicht als Repräsentantin der „Öffentlichkeit“ betrachtet werden sollte, und sie somit als (preisgünstige) Alternative zur demokratischen oder partizipativen Technikgestaltung zu missverstehen.

Methodologisch finden sich im Bereich der Technikethik die traditionellen Schulen der philosophischen Ethik, etwa beispielsweise Werte-Ethik, Verantwortungsethik, Dialektik, Diskursethik und Konstruktivismus (vgl. für einen Überblick über Proponenten Grunwald 1996: 443). Ergänzend führt Fenner noch Menschenrechtstheorien an (vgl. Fenner 2010: 223).

Ebenfalls im Umfeld der Technikethik bewegt sich die Ingenieursethik. Sie befasst sich mit der moralischen Verantwortung, die IngenieurInnen gegenüber ihrer Tätigkeit und den durch sie hervorgebrachten Verfahren und Produkten haben. Im Rahmen der Ingenieursethik wird der Versuch einer moralischen Selbstverpflichtung von IngenieurInnen unternommen, was sich etwa in Form von Ethik-Codizes niederschlägt. Ein Beispiel hierfür sind die „Ethischen Grundsätze des Ingenieurberufs“, herausgegeben vom Verband Deutscher Ingenieure (VDI). Darin wird in der Präambel formuliert: „In der Erkenntnis, daß Naturwissenschaft und Technik wesentliche Gestaltungsfaktoren des modernen Lebens und der Gesellschaft in Gegenwart und Zukunft darstellen, sind sich Ingenieurinnen und Ingenieure ihrer besonderen Verantwortung bewußt“ (VDI 2002: Präambel). In weiterer Folge wird diese Verantwortung näher spezifiziert als Verantwortung gegenüber dem Berufsstand, den gesellschaftlichen Institutionen, den ArbeitgeberInnen, den AuftraggeberInnen und den TechniknutzerInnen selbst (vgl. VDI 2002: Abschnitt 1.2). Zudem sollen technische Entwicklungen möglichst so gestaltet sein, Sachzwänge zu vermeiden und die TechniknutzerInnen in ihrer Selbstverantwortung zu unterstützen (vgl. VDI 2002: Abschnitt 2.2). Ropohl verweist darauf, dass in den USA seit 1947 erstmals die seit dem Ende des 19. Jahrhunderts existierenden ständischen Verhaltensregeln der IngenieurInnen, die sich auf fachliche

und kollegiale Korrektheit beschränkt hatten, um weitere Aspekte wie das Wohl der Menschheit und die Sicherheit des Lebens ergänzt wurden (vgl. Ropohl 2016: 62). Seit den 1970er Jahren haben die dedizierten Ethik-Codizes unterschiedlicher Ingenieursvereinigungen zugenommen, wovon der oben bereits genannte Ethik-Codex des VDI ein prominentes deutschsprachiges Beispiel ist. Nach Ropohl kann Techniksteuerung⁷ sich aber nicht auf Ingenieursethik beschränken, da die Ansprüche und Hoffnungen der Ingenieursethik der soziotechnischen Praxis nicht gerecht werden (vgl. Ropohl 2016: 97). So ist die Tätigkeit von IngenieurInnen hochgradig arbeitsteilig organisiert. Diese Handlungen sind eingebettet in gesellschaftliche Aggregate (also Korporationen wie z.B. Firmen), sodass ein zu individualistisches Handlungs- und Verantwortungsparadigma (wie Ropohl es in einem naiven Glauben an die Wirksamkeit von Ingenieurscodizes am Werk sieht) sich nicht praktisch auswirken kann (vgl. Ropohl 2016: 100). Wenn daher verlangt wird, dass IngenieurInnen ihre Arbeit und die Produkte die daraus hervorgehen, bewusst auf das Allgemeinwohl ausrichten, alle denkbaren Folgen der Projekte erwägen und diese auf die Vereinbarkeit mit maßgeblichen Wertvorstellungen der Gesellschaft prüfen, ggf. das Vorhaben modifizieren oder im Falle einer Verbesserung aufgeben sollen, sowie ihr Handeln auch noch jederzeit gegenüber anderen Personen (moralisch) erklären und verantworten zu können, dann könne das nicht in Form einer individuellen Ethik für diejenigen erfolgen, die im weitesten Sinne an technischen Entwicklungen mitwirken (vgl. Ropohl 2016: 109-120). In der Praxis hätten IngenieurInnen nämlich zwar gewisse Spielräume, um ihr Handeln moralisch auszurichten – ihre Verantwortungsfähigkeit wird aber begrenzt durch die Erfordernis, zusätzlich zur technischen Sachkenntnis auch noch ethische Wertkompetenz mitzubringen, weiters durch die (außer bei spezialisierten Sachsystemen) vielfältigen Anwendungsmöglichkeiten der technischen Entwicklungen und die durch den Markt erzeugte anonymisierte Verbreitung ihrer Produkte, die sie nur schwer begrenzen können, sowie durch die im Rahmen der arbeitsteilig erfolgten Hervorbringung der Produkte reduzierte persönliche Autonomie.⁸

3.3 KONVERGENZ VON TECHNIKFOLGENABSCHÄTZUNG UND TECHNIKETHIK

Während beiden Ansätzen gemein ist, dass sie Orientierungswissen, etwa für politische Entscheidungsträger oder für Fragen der Technikförderung, bereitstellen sollen, basier(t)en sie laut Grunwald (zunächst) auf fundamental unterschiedlichen Annahmen, wie diese Orientierung bereitgestellt werden

⁷ Techniksteuerung im Sinne der TA meint die informierte, politische Entscheidung von staatlichen Instanzen über die Unterbindung oder [modifizierte] Zulassung von anstehenden technischen Entwicklungen (vgl. Ropohl 2016: 36). Ropohl kritisiert diese Auffassung allerdings auch dahingehend, dass es über staatliche Akteure hinausgehend einer „konzertierten Techniksteuerung“ bedarf, bei der neben staatlichen Akteuren auch solche aus Wissenschaft, Wirtschaft und Gesellschaft gefragt sind.

⁸ Ropohl nennt seinen eigenen Lösungsvorschlag „konzertierte Techniksteuerung“, welche auf eine institutionalisierte, interdisziplinäre Zusammenarbeit von Gesellschaft, Politik, Industrie, IngenieurInnen und ethischen Fachleuten hinausläuft (vgl. Ropohl 2016: 259-356).

soll (vgl. Grunwald 1999: 170). Während es in der ethischen Reflexion um die **normativen Implikationen** von technikbezogenen Entscheidungen und die Bedeutung **moralischer Konflikte** ginge, würde Technikfolgenabschätzung stärker auf **(deskriptiver) soziologischer oder wirtschaftlicher Forschung** beruhen. Beide Ansätze, so Grunwald, wurden lange ohne Bezug aufeinander praktiziert und weiterentwickelt. In Deutschland etwa hatte sich die Diskussion zwischen beiden Disziplinen erst Anfang der 1990er Jahre entsponnen und zu einem fruchtbaren Austausch geführt.⁹

Grunwald skizziert schematisch Kernkritikpunkte, die lange gegenüber der jeweils anderen Disziplin vorgebracht wurden (vgl. Grunwald 1999: 171-176). Technikethik wisse demnach etwa zu wenig über die „echte Welt“ und wie tatsächlich Technik entwickelt wird und wie Entscheidungen zur Technikpolitik getroffen werden („operational deficit“), sodass technikethische Reflexionen praktisch wenig Relevanz haben, sondern wie Ratschläge aus dem akademischen Elfenbeinturm wirken. Was nützt es etwa, zu präskribieren, welches Risiko vor dem Hintergrund einer bestimmten Güterabwägung akzeptiert werden „soll“, wenn die Betroffenen einer neuen Technik dieses „Sollen“ faktisch anders sehen, und normative Akzeptabilität und faktische Akzeptanz auseinanderklaffen?¹⁰ VertreterInnen der Technikethik geben sich dieser Kritik zufolge einem normativen Irrglauben hin. Die gesellschaftliche Pluralität hinsichtlich moralischer Wertvorstellungen würde einen ethischen Konsens verunmöglichen und somit verhindern, dass ethische Ratschläge bei betroffenen Gesellschaftsgruppen auch akzeptiert würden. Technikethischen Überlegungen fehle somit insgesamt der Eingangspunkt in das faktische Geschehen der Technikentwicklung bzw. Technikanwendung.

Den Ansätzen der Technikfolgenabschätzung dagegen wiederum wurde lange Zeit aus dem „Lager“ der philosophischen Technikethik ein normatives Defizit unterstellt, wodurch TA gerade eben keine Orientierungsfunktion bieten könne. „Klassische“ Technikfolgenabschätzung etwa stelle idealtypisch lediglich umfassendes deskriptives (soziologisches oder ökonomisches) Wissen zu potentiellen Folgen und Risiken von (alternativen) Techniken bereit, auf welches sich dann politische Entscheidungsträger beziehen (so praktiziert etwa vom 1972 gegründeten „Office of Technology Assessment“ des Kongresses der USA, welches in seinen Konzepten auch Vorbildfunktion für die Einrichtung eines ähnlichen Büros im Deutschen Bundestag hatte). Die Entscheidung und die Evaluation von Alternativen würden somit dem politischen System überlassen, und TA beschränke sich auf bloße Deskription. An anderen, neueren „partizipativen Ansätzen der TA“, welche den klassischen Ansatz als zu expertenorientiert betrachteten, wird kritisiert, dass der diskursive Einbezug von Meinungen betroffener Personengruppen technikbezogene Konflikte zwar zu schlichten versuche, ein Konsens oder Kompromiss aber nicht per

⁹ Bleisch/Huppenbauer (2011: 35) bezeichnen dagegen die Technikfolgenabschätzung als eine Art Vorläufer der Angewandten Ethik, welche sich der Frage widmet, welchen Risiken man andere Menschen aussetzen darf.

¹⁰ Akzeptanz und Akzeptabilität werden als Begriffspaar auch im Rahmen der Behandlung der Hauptfrage 3 noch relevant werden (vgl. S. 151).

se automatisch ethisch gerechtfertigt sei, sodass auch hier letztendlich die normative Grundlage fehle, weil statt normativer Akzeptabilität lediglich faktische Akzeptanz ermittelt würde.

Diese idealtypische Auftrennung in Deskription/Technikfolgenabschätzung und Präskription/Technikethik ist aber mittlerweile nicht mehr aufrechtzuerhalten. Schon Grunwald zeigt in seinem Artikel aus 1999, inwiefern diese holzschnittartigen Kritikpunkte an der jeweils anderen Disziplin letztendlich nicht gerechtfertigt sind (so ist es z.B. kaum möglich, etwas lediglich deskriptiv zu beschreiben, weil bereits die Wahl der Terminologie normative Konnotationen aufweist), und sein eigener Ansatz einer „rationalen Technikfolgenabschätzung“ (vgl. Grunwald 1999: 176-179) bezieht ausdrücklich ethische Reflexionen für die Entwicklung rational gerechtfertigter Kriterien und Prozeduren ein. Auch Ropohl (2016), der den Begriff Technikbeurteilung gegenüber Technikfolgenabschätzung bevorzugt, betont ausdrücklich die Wichtigkeit normativer Reflexionen und Zielvorstellungen (etwa hinsichtlich der gewünschten gesellschaftlichen Zukunft) im Prozess der TA, da sonst gar nicht entschieden werden könne, ob technikpolitikbezogene Entscheidungen überhaupt dieser Zielvorstellung zu- oder abträglich sind. Dass auf der anderen Seite auch auf technikethischer Seite der „akademische Elfenbeinturm“ verlassen wurde, ist etwa ersichtlich an Fenner (2010: 30ff), welche als Fundament für Angewandte Ethik (und so auch für Angewandte Technikethik) Diskursethik als Rahmentheorie präferiert, und so im Kontext der Angewandten Technikethik auch vorsieht, die von einer fraglichen Technik Betroffenen zu Wort kommen zu lassen. Zudem zeigen Reijers et al (2018) auf, wie Technikethik in Form von Forschungsprojekten im Bereich Research & Innovation forschungspraktisch durchgeführt wird – und die dabei angeführten Methoden und Ansätze legen dabei nahe, dass in einem interdisziplinären Zugang empirische Forschungsarbeiten etwa auch aus sozialwissenschaftlichen Disziplinen herangezogen werden müssen, um Implikationen der Forschungsentwicklung in unterschiedlichen Dimensionen (sehr ähnlich etwa zum Werteoktagon des VDI) berücksichtigen zu können.

Technikfolgenabschätzung wie auch Technikbewertung zielen beide darauf ab, die Geltung des „**technologischen Imperativs**“ zu beschränken, wie ihn Mumford (in kritischer Absicht) geprägt hat: „Die westliche Gesellschaft hat einen technologischen Imperativ als unanfechtbar akzeptiert, der ebenso willkürlich ist wie das primitivste Tabu: nicht bloß die Pflicht, Erfindungen zu fördern und fortlaufende technologische Neuerungen herbeizuführen, sondern ebenso die Pflicht, sich diesen Neuerungen bedingungslos zu unterwerfen, nur weil sie angeboten werden, ohne Rücksicht auf ihre Folgen für die Menschen“ (Mumford 1974: 548). Zweitrangig wäre demnach laut Mumfords etwas kulturpessimistischer Diagnose daher die Frage, ob ein Produkt tatsächlich Verbesserungen brächte – denn allein die Tatsache, das Resultat einer neuen wissenschaftlichen Entdeckung oder eines neuen technologischen Verfahrens oder einer neuen Investitionsmöglichkeit zu sein, würde vermeintlich bereits den Wert beweisen. Ropohl bringt diesen Gedanken so auf den Punkt bringt: „Alles, was in der Technik gemacht

werden kann, soll gemacht werden“ (Ropohl 2016: 22). Wenn die Gefahr nun darin besteht, dass der Mensch die angebotenen technologischen Möglichkeiten auch nutzt, dann ist für Mumford die Absurdität dies ohne Schranken zu akzeptieren, offensichtlich, denn: „Wenn der Mensch die Macht hat, alles Leben auf Erden auszurotten, wird er es tun“ (Mumford 1974: 549).

Wie ersichtlich wurde, gibt es zwischen TA und Technikethik mittlerweile sowohl inhaltliche wie methodische Überschneidungen. Ganz gewiss unterscheiden sie sich aber hinsichtlich ihrer Institutionalisierung: Während die TA häufiger in Form von außeruniversitären Forschungsinstituten institutionalisiert wurde, ist Technikethik ein akademisches Fach, welches an Universitäten betrieben wird.

Ein gemeinsames Problem, mit welchem sowohl Arbeiten der Technikfolgenabschätzung wie auch der Technikethik zu kämpfen haben, ist das **Problem der Verallgemeinerbarkeit** (vgl. Grunwald 2010: 151f). Grunwald argumentiert gegen die Möglichkeit, normative Postulate oder gesellschaftliche Werte als allgemeingültig oder allgemein verbindlich zu erklären, die dann als Fundament für Ableitungen hinsichtlich der Akzeptabilität, Wünschbarkeit oder Verpflichtung gegenüber dem Einsatz oder der Entwicklung von Technik fungieren. Er plädiert für einen **konditionalen Bezug** auf diese normativen Ausgangspunkte, im Sinne von: „Wenn man bestimmte normative Ausgangspunkte verwendet, hat dies folgende Konsequenzen oder Implikationen: ...“ (Grunwald 2010: 151). Dies verhindere die Vorwegnahme einer politischen oder gesellschaftlichen Bewertung, und es obliege weiterhin politischen oder gesellschaftlichen Aushandlungsprozessen, über das Vorliegen der „Wenn“-Anteile zu entscheiden. Für die vorliegende Arbeit bedeutet dies umso mehr, transparent zu machen, ob von „Wenn“-Anteilen ausgegangen wird, und die Ableitung der „Dann“-Anteile rational zu erläutern.

Schließlich: Was Simonis für die TA feststellt, gilt nach Ansicht des Verfassers der vorliegenden Arbeit auch für die Technikethik: beide erheben einen universalisierbaren Anspruch ihrer Aussagen und Gestaltungsempfehlungen (oder machen wenigstens ihre Geltungsbedingen transparent, d.V.), und stellen sich somit gegen den Ausschluss bestimmter Gruppen von den Vorteilen der technischen Entwicklung wie auch gegen Strategien individueller Nutzenmaximierungen (vgl. Simonis 2013a: 13).

Nach diesen Ausführungen kann festgestellt werden, dass sich die vorliegende Arbeit methodisch im Überschneidungsbereich von TA und Technikethik verorten lässt. Den Anlass zu ihr haben Vorarbeiten aus der Human+ bezogenen ELSI Begleitforschung gegeben, in welche der Verfasser der vorliegenden Arbeit involviert war, und es geht um eine Bewertung der (ethischen) Akzeptabilität und der möglichen Folgen je nach Art der Ausgestaltung. Es werden Vorschläge für die verträgliche(re) Weiterentwicklung von Human+ oder seiner Komponenten formuliert. Insofern dient auch sie der Begrenzung des technologischen Imperativs, und möchte einen Beitrag zu verantwortungsvoller Forschung leisten.

Nach dieser Selbstverortung der vorliegenden Arbeit werden in den nächsten Kapiteln die Grundlagen erarbeitet, aus denen sich dann das Bewertungsframework für Human+ zusammensetzt. Den Anfang macht im nächsten Kapitel eine Erarbeitung des negativen Utilitarismus, um am Ende des Abschnitts das „Metaprinzip des Negativen Utilitarismus“ zu formulieren. Daran anschließend werden im übernächsten Kapitel stärker anwendungsorientierte Frameworks, wie sie schon jetzt in der Technikethik bewertet werden, in den Blick genommen, und in diesem Zuge v.a. auch eine Erarbeitung der Grundlagen des Principlism gegeben, da dieser im Hintergrund zahlreicher dieser Ansätze steht.

4 NEGATIVER UTILITARISMUS

4.1 UTILITARISMUS – GRUNDLAGEN UND KRITIK

Um den „negativen Utilitarismus“ (NU) zu erläutern, ist es zunächst erforderlich, den „klassischen“ (oder positiven) Utilitarismus im Allgemeinen und die wichtigsten Kritikpunkte an ihm zu skizzieren, um im Anschluss darzulegen, in welchen Punkten sich der NU vom KU unterscheidet. Generell bezeichnet Utilitarismus die bekannteste Spielart des Konsequentialismus. **Konsequentialismus**, auch konsequentialistische oder teleologische Auffassung genannt, beurteilt Handlungen nach dem Ausmaß, in welchem diese Handlungen bestimmte Ziele fördern (vgl. Singer 1994: 17). Der moralische Wert von Handlungen, Institutionen, Gesetzen und Regeln kann im Konsequentialismus nur durch Bezug auf ihre Konsequenzen gerechtfertigt werden (vgl. Williams 1979: 42).

Utilitarismus ist eine konsequentialistische Auffassung, die im Wesentlichen als Beurteilungsmaßstab die Auswirkungen von Handlungen auf das **Glück** [happiness] oder den **Nutzen** bzw. die **Nützlichkeit** [utility] auf die von einer Handlung betroffenen Personen heranzieht, wobei sich je nach VertreterIn dieser Position die genaue Terminologie unterscheidet. Er ist gekennzeichnet durch eine empiristische Grundhaltung, die gewisse anthropologische Grundannahmen zum Streben von Menschen enthält, wie im Weiteren gleich sichtbar werden wird.

Frühe Hauptvertreter des klassischen Utilitarismus sind u.a. Jeremy Bentham und John Stuart Mill. Den Kerngedanken des Utilitarismus bezeichnet Jeremy Bentham in seinem 1789 verfassten und 1823 überarbeiteten Text „An Introduction to the Principles of Morals and Legislation“ (zu deutsch: „Eine Einführung in die Prinzipien der Moral und Gesetzgebung“) als das **Prinzip der Nützlichkeit** [principle of utility], welches

„[...] approves or disapproves of every action whatsoever, according to the tendency which it appears to have to augment or diminish the happiness of the party whose interest is in question: or, what is the same thing in other words, to promote or to oppose that happiness“ (Bentham 1948: 126).

Nützlichkeit wird als Eigenschaft definiert, die dazu neigt, Gewinn, Vorteil, Freude, Gutes oder Glück hervorzubringen, bzw. vor Unheil, Leid, Bösem oder Unglück zu bewahren. Bentham macht deutlich, dass sich Glück und Unglück aufrechnen lassen: „An action then may be said to be conformable to the principle of utility, or for shortness sake, to utility, (meaning with respect to the community at large) when the tendency it has to augment the happiness of the community is greater than any it has to diminish it“ (Bentham 1948: 127).

Der zweite ihrer klassischen Hauptvertreter, John Stuart Mill, benennt das Hauptprinzip des Utilitarismus dagegen als das **Prinzip des größten Glücks** [greatest happiness principle]:

„The creed which accepts as the foundation of morals, Utility, or the Greatest Happiness Principle, holds that actions are right in proportion as they tend to promote happiness, wrong as they tend to produce the reverse of happiness. By happiness is intended pleasure, and the absence of pain; by unhappiness, pain, and the privation of pleasure“ (Mill 2014: 10).

Mill macht deutlich, dass nicht das größte Glück des Handelnden, sondern das größte Glück insgesamt gemeint ist: „[...] the happiness which forms the utilitarian standard of what is right in conduct, is not the agent's own happiness, but that of all concerned“ (vgl. Mill 2014: 24).

Schon jetzt wird klar, dass es *den* Utilitarismus folglich nicht gibt. Bentham und Mill sehen zwar beide das Ziel der Ethik im Verwirklichen des größten Glücks für die größte Zahl, unterscheiden sich aber in der Definition von Glück (vgl. Kaufmann 1994: 17): Während Bentham Glück als Nützlichkeit und als zu verteilende „Substanz“ auffasst, vertritt Mill einen stärker qualitativen Glücksbegriff. Dies wird etwa sichtbar in Mills Ausführungen zu den unterschiedlichen Arten von Freuden bzw. Gütern [pleasures], etwa solchen des Körpers, des Intellekts oder des moralischen Gefühls, in welchen er sich schließlich gegen ein rein quantitatives Glückskalkül wendet:

„It is quite compatible with the principle of utility to recognise the fact, that some kinds of pleasure are more desirable and more valuable than others. It would be absurd that while, in estimating all other things, quality is considered as well as quantity, the estimation of pleasures should be supposed to depend on quantity alone“ (Mill 2014: 11f).

Höffe formuliert das Anliegen des Utilitarismus etwas abstrakter: „Diejenige Handlung bzw. Handlungsregel ist moralisch richtig, deren Folgen für das Wohlergehen aller Betroffenen optimal sind“ (Höffe 2008: 11) und bietet auch eine als Imperativ formulierte Wendung an: „Handle so, daß die Folgen deiner Handlung bzw. Handlungsregel für das Wohlergehen aller Betroffenen optimal sind“ (Höffe 2008: 11).

Aus obigen klassischen Definitionen wird deutlich, dass zwei Leitbegriffe im Utilitarismus eine große Rolle spielen: **Empfindungsfähigkeit**, da sie die Voraussetzung für das Empfinden von Glück darstellt, und **universale Gleichheit** (Akerma 2000: 196). Akerma verweist darauf, dass konsequenterweise der Utilitarismus somit statt eines Anthropologismus eigentlich einen Zoologismus implizieren würde, und nicht nur alle Menschen, sondern alle empfindungsfähigen Wesen als gleichrangig zu betrachten sind,

etwa wenn Mill von der gesamten fühlenden Natur spreche¹¹, wie auch Bentham in einer Fußnote als essentielles Kriterium nicht die Fähigkeit zur Vernunft oder zur Sprache, sondern die Leidensfähigkeit hervorstreicht und damit für die Etablierung von Rechten für Tiere in der Zukunft plädiert.¹²

In der Literatur haben sich gängige Bezeichnungen für unterschiedliche Ausformungen des Utilitarismus etabliert, die sich hinsichtlich der Anwendung des obersten utilitaristischen Prinzips unterscheiden. Im **Handlungsutilitarismus** wird das utilitaristische Prinzip auf Handlungen direkt angewandt, d.h. es wird danach gefragt, welche Folgen die Ausführung einer ganz bestimmten Handlung in einer ganz bestimmten Situation durch eine ganz bestimmte Person hat (vgl. Frankena 1994: 55f). Im **Regelutilitarismus** wird dagegen etwas allgemeiner darauf abgezielt, solche Handlungsregeln festzulegen und zu berücksichtigen, die entsprechend dem Utilitätsprinzip den größten Nutzen für die größte Zahl bedeuten, wenn jeder in ähnlichen Situationen nach der jeweiligen Regel handeln würde. Ein Hauptkritikpunkt gegen den Handlungsutilitarismus besteht darin, dass es aus Sicht der Kritiker unpraktikabel sei, umfangreiche Abwägungen für einzelne Handlungen in der erforderlichen Konkretheit vorzunehmen, bzw. dass es problematisch werden könne, wenn jede Person in jeder Situation selbst entscheiden könne, was das Richtige sei, und dass es stattdessen für das allgemeine Wohl vorteilhafter sei, wenn jeder das eigene Handeln (weitgehend) festen Regeln unterwerfen würde (vgl. Frankena 1994: 56). Williams bezeichnet diese beiden Spielarten auch als direkten und indirekten Konsequentialismus (vgl. Williams 1979: 43).

Eine weitere häufige Unterscheidung ist die zwischen dem **klassischen Utilitarismus** (der den Nutzen von Handlungen in Abhängigkeit von der Maximierung von Lust und der Minimierung von Leid bewertet) und dem **Präferenzutilitarismus**, in welchem Handlungen danach beurteilt werden, in welchem Grade sie mit den Präferenzen der von einer Handlung bzw. deren Konsequenzen Betroffenen übereinstimmen (vgl. Singer 1994: 128).

Neben diesen Unterscheidungen werfen utilitaristische Zugänge zahlreiche Detailfragen auf, von denen zum besseren Verständnis im Folgenden eine Reihe angesprochen werden. Können etwa beliebige Ziele oder Zwecke utilitaristisch gerechtfertigt werden? Nach Höffe müssen die zu rechtfertigenden Ziele und Zwecke bereits „in sich gut sein“ (kontra einem unterstellten Wertnihilismus), was jeweils eine Werttheorie erfordere. Bei den klassischen Vertretern Bentham und Mill besteht die werttheoretische Position in der Annahme der Erfüllung menschlicher Bedürfnisse und Interessen, also des menschlichen Glücks, sodass das Ziel die maximaler Bedürfnis- und Interessenbefriedigung bzw. minimale Frustration sei (vgl. Höffe 2008: 10). Weil grundsätzlich für unterschiedliche „höchste Werte“

¹¹ So definiert Mill als den Maßstab der Moralität [standard of morality] diejenigen Grundsätze für menschliches Verhalten, bei deren Befolgung eine glückliche Existenz sichergestellt werden kann, und zwar „[...] secured to all mankind; and not to them only, but, so far as the nature of things admits, to the whole sentient creation“ (Mill 2014: 17).

¹² „[T]he question is not, Can they *reason*? Nor, Can they *talk*? But, Can they *suffer*?“ (Bentham 1948: 412)

argumentiert werden könne, bilde der Utilitarismus somit keine einzelne homogene Position, sondern eher eine große Theriefamilie.

Eine weitere Frage ist die, ob es etwa um die Maximierung der Gesamtsumme des Glücks aller Individuen aller Zeiten geht (**Nutzen-Summen-Utilitarismus**), oder ob es um die Maximierung des Glücks aller (aktuell existenten) Individuen gehen soll (**Durchschnitts-Nutzen-Utilitarismus**, vgl. Akerma 2000: 200). Akerma verweist darauf, dass sich bei Mill Passagen finden, mit denen sowohl die eine als auch andere Leseart gestützt werden kann (Maximierung des „Glücks der Menschheit insgesamt“ versus Maximierung des „größten Glücks aller von einer Handlung betroffenen“). Besonders in der ersten Lesart erweise sich der Utilitarismus als Morallehre „par excellence“ um Fragen über Rechte zukünftiger Generationen zu erörtern, da sich Betroffene im Utilitarismus nicht erst durch Tugend oder Vernunft des Glücks für würdig erweisen müssen, was zukünftige Generationen mangels aktueller Existenz gar nicht können, „[a]ber sie werden existieren, und die Weichen für ihr (Un)Glück werden jetzt gestellt“ (Akerma 2000: 198).

Würde Human+ Nutzen-Summen-utilitaristisch bewertet werden, stünde man vor der Aufgabe zu bewerten, wie sich die Bereitstellung einer Technik für ein gemeinsames Echtzeitlagebild für BOS und die Erstellung von Migrationsprognosen auf Basis verschiedener Datenquellen (Social-Media-Daten, etc.) auf das Glück aller zukünftig betroffener Stakeholder-Gruppen auswirkt.¹³ Das erfordert offenbar eine ganze Reihe prognostischer Leistungen (z.B. Human+ Anwendungssetting auch in der fernen Zukunft, Migrationsbetroffenheiten der Zukunft, Social-Media-Verhalten von Personen und wie sich deren Einstellungen zur Verwendung ihrer Posts durch Dritte entwickelt, Missbrauchsmöglichkeiten der Technik, welche weiteren Technikpotentiale durch die Human+ Algorithmen eröffnet werden, wie diese mit anderen potentiell erst emergierenden Techniken interagieren, und viele mehr). Wie sich später in Kapitel 6 zeigen wird, ist es bereits für die gegenwärtigen Anwendungsbedingungen schwierig, konkrete ethische Analysen aufgrund vieler Unbestimmtheiten durchzuführen, was schon jetzt gegenwartsnahe konditionale Bezüge erfordert („Was wäre wenn?“) – geschweige denn die unterschiedlichen möglichen, sich immer mehr verzweigenden Entwicklungspfade auch für die ferne Zukunft durchzudeklinieren. Eine ernst gemeinte Nutzen-Summen-utilitaristische Analyse scheint für Human+ daher kaum durchführbar zu sein – zumal ein Ziel der Arbeit auch die Formulierung von konkreten Weiterentwicklungsvorschlägen für Human+ ist. Je allgemeiner oder abstrakter technikethische Analysen durchzuführen sind, so scheint es, desto eher lassen sich Nutzen-Summen-utilitaristische Erwägungen einbringen. Je mehr es um anwendungsnahe Analysen geht, um so eher scheint die Ausrichtung an der Betroffenheit von aktuell existierenden Individuen notwendig zu sein, wie sie etwa im Durchschnitts-

¹³ Zwar spricht Akerma von „allen Individuen aller Zeiten“, aber aus Beispielgründen sei das hier eingeschränkt auf „alle Angehörigen der Stakeholdergruppen aller Zeiten“.

Nutzen-Utilitarismus vorgesehen ist. Damit läuft man allerdings zumindest prinzipiell in Gefahr, Auswirkungen auch für spätere Generationen aus dem Blick zu verlieren.

In jedem Fall stellt sich bei utilitaristischen Erwägungen die Frage, welche Art von Folgen bei der Abwägung relevant sein sollen: die **voraussichtlichen**, die **tatsächlichen**, oder die **feststellbaren**¹⁴ (vgl. Zecha 1999: 421)? Fraglich ist auch, wo die Folgenberücksichtigung abgebrochen werden soll. Schließlich können nicht nur **direkte Handlungsfolgen** beachtet werden, sondern auch **deren Konsequenzen**, und wiederum die Konsequenzen von diesen Konsequenzen, ad infinitum.

Erneut lassen sich hier unter Bezug auf Human+ Beispiele formulieren. So wird weiter unten unter Bezug auf Videokamerasysteme der Begriff der Normalisierung des Verhaltens eingeführt (vgl. Seite 137). Kurz gesagt bedeutet Normalisierung, dass das bloße Vorhandensein einer Kamera dazu führen kann, dass betroffene Personen ihr Verhalten an Normvorstellungen ausrichten. Bewertet man die angedachte Etablierung eines Videokamerasystems, dann ist Normalisierung also zunächst eine aus der Literatur ableitbare, voraussichtliche Folge, die eintreten kann, aber nicht unbedingt eintreten muss. Will man **Risiken** in den Blick bekommen, so müssen folglich auch bereits nur voraussichtliche oder möglicherweise eintretende Folgen inkludiert werden. Würde man dagegen nach der tatsächlichen Etablierung des Videokamerasystems empirisch beforschen (z.B. durch verdeckte Beobachtung), ob Personen auch tatsächlich ihr Verhalten ändern, wenn sie in einen videoüberwachten Bereich treten (bzw. unter welchen Bedingungen oder bei welchen Personengruppen dies der Fall ist), um zu prüfen, ob tatsächlich Normalisierung festgestellt werden kann und dann zu entscheiden, ob die Videokameras dort weiter bestehen sollen, dann wäre dies zugleich die Berücksichtigung von tatsächlichen wie auch feststellbaren Folgen.

Wenngleich in der vorliegenden Arbeit nicht der Utilitarismus selbst im Vordergrund steht, soll im Weiteren kurz auf die wichtigsten Kritikpunkte an utilitaristischen Positionen eingegangen werden, ohne diese in ihrer Breite erschöpfend zu behandeln (vgl. etwa für eine frühe Auseinandersetzung an zeitgenössischen Kritikpunkten bereits Mill 2014 [1864], für eine grundlegende Beschäftigung Williams 1979, für einen aktuelleren Überblick Höffe 2008: 41-49).

Als einer der Hauptkritikpunkte am Utilitarismus wird häufig vorgebracht, dass er durch die Annahme des Prinzips des größten Glücks als einzelnes, höchstes Prinzip, Aspekte der Gerechtigkeit nicht in ausreichendem Maße berücksichtigen könne, jedenfalls nicht ohne das strenge utilitaristische Fundament zu verlassen. Beispielhaft führt dies Frankena aus, der hierzu ein Gedankenexperiment von Butler und

¹⁴ Zecha unterscheidet zwar tatsächliche und feststellbare Folgen – allerdings scheint es fraglich, wie nicht-feststellbare tatsächliche Folgen berücksichtigt werden sollen.

Ross adaptiert, um ein Problem des Regelutilitarismus aufzuzeigen (vgl. Frankena 1994: 58f): Angenommen man müsste zwischen zwei moralischen Regeln R1 und R2 entscheiden, von denen weiters angenommen wird, sie würden bei Befolgung jeweils in exakt gleichem Ausmaß ein Übergewicht von guten gegenüber schlechten Folgen auf lange Sicht und für alle Betroffenen bewirken. Dann wären beide Handlungsoptionen aus regelutilitaristischer Sicht äquivalent, selbst dann wenn bei R1 der durch die Regel erlangte zusätzliche Nutzen nur unter einer kleinen Minderheit verteilt würde, in R2 dagegen der Zusatznutzen „gleichmäßig“ über alle Betroffenen verteilt würde. Dies wäre auch dann der Fall, wenn etwa die dadurch begünstigte Gruppe in R1 noch dazu selbst keinerlei Beitrag hierfür geleistet habe, also keinen „Verdienst“ hierfür gelten machen kann (der Begriff des Verdienstes überschreite zudem bereits das utilitaristische Fundament). Für Frankena ist auf Grund der wesentlich ungleichen Verteilung R1 gegenüber R2 eine „ungerechte“ Regel. Das Nützlichkeitsprinzip selbst müsse somit ergänzt werden durch ein Gerechtigkeitsprinzip, welches die Verteilung des Nutzens gesondert behandelt. Mill, der das größte Glück für die größte Zahl fordere, habe dem Nützlichkeitsprinzip bereits einen Doppelcharakter gegeben, und somit für Frankena den „reinen“ Utilitarismus zugunsten eines gemischten Prinzips bereits verlassen. Man kann nun einwenden, dass der Grad der Gerechtigkeit einer Verteilung in die Liste der „in sich guten“ Dinge aufgenommen werden könnte (wie etwa Vergnügen, Glück oder Erkenntnis), sodass beim Vergleich der Handlungsregeln auch der Wert der Verteilungsweise berücksichtigt werden könnte. Dies weist Frankena aber zurück: während etwa Vergnügen und Erkenntnis als zu maximierende Güter betrachtet werden können, wäre der Grad der Gerechtigkeit einer Verteilung im außermoralischen Sinne nicht „gut“, sondern „richtig“, sodass diese Verteidigungslinie „richtig“ und „gut“ durcheinanderbringe. Unabhängig davon wäre dann auch das Kriterium dafür, ob eine Verteilung gerecht ist, nicht das Nützlichkeitsprinzip selbst, sondern ein anderes, z.B. das der Gleichverteilung, sodass das Nützlichkeitsprinzip nicht auch als Gerechtigkeitsprinzip fungieren könne. Höffe führt hierzu an, dass durch Zusatzannahmen, wie etwa das Gesetz des abnehmenden Grenznutzens, sich der Gegensatz von Utilitarismus und Gerechtigkeit zumindest im Bereich der Verteilungsgerechtigkeit abmildere, ja sogar eine strenge Gleichbehandlung folgen würde (vgl. Höffe 2008: 46).

Mitunter wird diese Kritik mangelnder Gerechtigkeit auch ergänzt mit der Behauptung, der Utilitarismus betrachte die von einer Handlung Betroffenen lediglich als Träger von Empfindungen oder von Interessen, und es komme nur darauf an, ein Maximum an subjektivem Wohlbefinden bzw. an Interessensbefriedigung zu erzielen, ungeachtet einer gerechten Verteilung von Nutzen oder Lasten, sodass der Schutz einzelner Betroffener Wesen vor Schaden vor dem Hintergrund des zu maximierenden Gesamtnutzens in den Hintergrund trete (vgl. Fenner 2010: 143). Kritisch sehen sich VertreterInnen utilitaristischer Strömungen daher häufig mit dem Vorwurf konfrontiert, dass der Utilitarismus (ohne Rekurs auf ergänzende Prinzipien) nicht ausreichend für den Schutz von Minderheiten sorgen könne,

wenn deren Schaden durch einen umso größeren Nutzen der Mehrheit per Nutzenkalkül überwogen würde. Ein Einwand hierzu wäre, dass die Mehrheit nicht glücklich sein könne, wenn es einer Minderheit schlecht gehe bzw. der Glücksgewinn auf deren Kosten erkaufte würde – ein idealistischer Einwand, der sich aber in der Wirklichkeit nicht bewähre (vgl. Kaufmann 1994: 19). Der (positive) Utilitarismus lasse nach Kaufmann aus dieser Sicht keinen Minderheitenschutz zu.

Ein weiterer Einwand gegen den Utilitarismus zielt darauf ab, dass Glück nicht universalisierbar sei, sofern es nicht so abstrakt gefasst würde, dass es bereits inhaltsleer würde (vgl. Kaufmann 1994: 20). Inhaltliche Argumentationen erfordern dagegen einen gehaltvolleren, konkreteren Inhalt des Begriffs Glück, und das könne dann je nach Person ein anderer sein. Während zwar fast alle Menschen etwa gesund, frei, satt etc. sein wollten, gäbe es dennoch zahlreiche Menschen, auf die diese Kriterien zutreffen, und die dennoch nicht glücklich sind.

Das utilitaristische Nutzenkalkül, welches positive gegenüber negativen Folgen für alle Betroffenen abwägt, könne Kritikern zufolge zudem nicht im Sinne einer exakten Anwendung oder Abwägung angewandt werden, wie es etwa Bentham (1948) im mathematischen Sinne für möglich hält, sondern sollte eher lediglich als eine Art „regulative Idee“ betrachtet werden (vgl. Höffe 2008: 42). Schließlich lasse bereits der Begriff „Betroffene/r“ unterschiedliche Definitionen zu, etwa nach direkter oder indirekter Betroffenheit, oder im Sinne von aktuell Betroffenen oder der Ausweitung auch auf zukünftige Generationen. Weiters sei eine für ein Nutzenkalkül notwendige quantitative Bestimmung von Leid und Nutzen auf Grund eines fehlenden einheitlichen, interpersonellen Maßstabs von Glück und Leid nicht möglich, sodass interpersonelle Vergleiche scheitern müssen. Deshalb argumentiert Höffe, dass die Idee eines utilitaristischen Nutzenkalküls eher als Aufforderung verstanden werden sollte, sorgfältig bei alternativen Optionen Vor- und Nachteile zu überlegen und gegeneinander abzuwägen.

Ein weiterer Kritikpunkt ist der, dass bei Befolgung utilitaristischen Kalküls sich Schwierigkeiten für die Wertsphäre der „Integrität“ ergeben können. Aus utilitaristischer Sicht ist es nämlich unerheblich, ob eine Handlung von Person A oder Person B vollbracht (bzw. nicht vollbracht) wird, solange beide Handlungen den selben bzw. größeren Überhang von guten gegenüber schlechten Folgen mit sich bringen. Dies hat besondere Folgen für die eigene Verantwortung im Falle einer Unterlassung. Williams spricht hier von der **starken These negativer Verantwortlichkeit**: „wenn ich weiß, daß, falls ich x tue, E_1 eintreten wird und daß, wenn ich es unterlasse, x zu tun, E_2 eintreten wird und daß E_2 schlechter ist als E_1 , bin ich verantwortlich für E_2 , wenn ich es freiwillig unterlasse, x zu tun“ (Williams 1979: 72). Es wirke daher so, als beinhalte der Utilitarismus eine grenzenlose Verpflichtung, die Welt zu verbessern, und

dies bringe in letzter Konsequenz Schwierigkeiten für die Wertsphäre der „Integrität“¹⁵. Williams illustriert dies an einem Beispiel, das auf J.J.C. Smart zurückgeht (vgl. Williams 1979: 61). In abstrakter Form hat es folgende Gestalt: Wenn Person A einwilligt, 1 unschuldige Person zu töten, lässt Person B dafür 9 andere unschuldige Personen am Leben, die B ansonsten zusätzlich zu der von A zu tötenden Person getötet hätte. Wenn sich Person A weigert, wird Person B dagegen alle 10 unschuldigen Personen töten. Person A hat im Gedankenexperiment keine Möglichkeit, Person B zu hindern – im Originalbeispiel ist Person A ein zufälliger Reisender, und Person B ein Militärkommandant. Einen unschuldigen Menschen zu töten ist völlig gegen die eigenen Wertvorstellungen von Person A. Aus utilitaristischer Sicht könne eine Weigerung von Person A (die im Endergebnis 9 Personen mehr das Leben kostet) als „hemmungslos übertriebene Gewissenhaftigkeit“ betrachtet werden (vgl. Williams 1979: 66). Wenn nun der Utilitarismus dieser Argumentation folgend von Person A fordert, zu schießen, dann entfremdet er Person A von ihren eigenen moralischen Gefühlen und führe so zum Verlust der moralischen Integrität von Person A.

4.2 NEGATIVER UTILITARISMUS UND DIE ASYMMETRIE VON LEID UND GLÜCK

Gegenüber dem positiven Utilitarismus (PU)¹⁶ wird im negativen Utilitarismus eine Akzentverschiebung vorgenommen: im Vordergrund steht hier nicht das größte Glück der größten Zahl, wie es bei Bentham gefordert wird, sondern das „kleinste Leid der kleinsten Zahl“ (vgl. Ropohl 2016: 314). Der Begriff des negativen Utilitarismus (NU) wurde ursprünglich in einer Replik an Popper von R.N. Smart geprägt (vgl. Smart 1958, Akerma 2000: 224). Wenngleich der Begriff neu ist, weist Akerma nach, dass bereits bei Mill neben einer Positivbestimmung auch eine Negativbestimmung des Glücks erkennbar ist, wenn er Glück als Freisein von Unlust bezeichnet, und dass Mill wesentlich optimistischer (und drängender) gegenüber einer Reduzierung von Unglück als gegenüber einer Beförderung von Glück gewesen wäre (vgl. Akerma 2000: 217-220). Das grundlegende Konzept des NU lässt sich somit bis in die ersten modernen Formulierungen des Utilitarismus bei Mill zurückverfolgen, und ist weniger neu als der es bezeichnende Terminus.

Popper selbst hat ebenfalls zum NU geschrieben. Seine Ausführungen finden sich in einer Endnote zum Abschnitt der Sozialtechnik der kleinen Schritte (vgl. Popper 2003: 188), die er als Gegenkonzept zum Entwurf von politischen und gesellschaftlichen Utopien entwickelt. Während Utopien versuchen,

¹⁵ Williams gibt zwar keine Definition von „Integrität“, spricht aber an einer Stelle davon, dass dann das „Bild“ von der eigenen moralischen Identität intakt ist (vgl. Williams 1979: 67). Damit spielt er offenbar darauf an, dass zwischen dem moralischen Handeln und dem moralischen Denken bzw. Empfinden Kongruenz besteht.

¹⁶ Im Rahmen der vorliegenden Arbeit bezeichnet PU jegliche Art von nicht-negativem Utilitarismus. Der alternativ zu PU mögliche Begriff „klassischer“ Utilitarismus wird bei verschiedenen Autoren allerdings unterschiedlich genutzt. So verwendet beispielsweise Singer (1994) ihn für nicht-präferenzutilitaristische Formen. Um Verwirrungen zu vermeiden, wird hier daher die Bezeichnung PU verwendet.

durch umfangreiche Neuaufstellungen den idealen Staat zu verwirklichen, „[...] wird sich der Anwalt der Sozialtechnik der kleinen Schritte nach den größten und dringlichsten Übeln in der Gesellschaft umsehen, und er wird versuchen, die zu beseitigen; er wird nicht dem höchsten Gut nachspüren und sich für seine Verwirklichung einsetzen“ (Popper 2003: 188). In der Endnote zu dieser Bemerkung argumentiert Popper gegen die Annahme einer **Symmetrie von Freude und Leiden** bzw. von Lust und Schmerz (vgl. Popper 2003: 391). Leiden beinhaltet den direkten moralischen Appell zu helfen, während es keine gleichgelagerte Notwendigkeit dafür gebe, Glück oder Freude derjenigen zu vermehren, denen es ohnehin gut gehe. Die Annahme einer „Glückseligkeitsskala“, bei welcher Schmerz als negative Glückseligkeit aufgewogen werden könne durch positive Glückseligkeit (wie etwa von Bentham skizziert), ist somit falsch, v.a. wenn der Schmerz des einen durch Freude des anderen aufgewogen werden soll. „Statt der größten Glückseligkeit für die größte Zahl sollte man – etwas bescheidener – das kleinste Maß an vermeidbarem Leid für alle fordern; und man sollte weiterhin verlangen, daß unvermeidbares Leid – wie Hunger in Zeiten eines unvermeidlichen Mangels an Nahrungsmitteln – möglichst gleichmäßig verteilt werde“ (Popper 2003: 391). Nach Akerma ist es eine grundsätzliche Diskussion im Utilitarismus, ob es eine solche, von Popper angenommene **Asymmetrie zwischen Leid und Glück** gebe (und das Leid schwerer zu gewichten sei) oder ob, quantifizierend gesprochen, ein bestimmter Beitrag an Leid durch einen äquivalenten Beitrag an Glück ausgeglichen werden kann (vgl. Akerma 2000: 223). Akerma selbst argumentiert für das tatsächliche Vorliegen dieser Asymmetrie, und führt hierfür drei Kategorien von Grundbefindlichkeiten ein (vgl. Akerma 2000: 235f): **Leiden** (wenn grundlegende oder höhere Bedürfnisse nicht befriedigt sind), **Zufriedenheit** (bei Erfülltsein grundlegender Bedürfnisse) und **Glück** (bei Erfülltsein von höheren, extravaganten bzw. „ausnahmehaften“ Bedürfnissen). Wer nun „leidet“, deren bzw. dessen Weiterleben ist akut bedroht. Wessen „Glück“ bedroht ist, dem fehlen zwar (womöglich wichtige) Bestandteile des Lebens, wie etwa der Lebenssinn – dieser kann aber wiedergewonnen werden, während im Falle des Leidens im Sinne einer akut die eigene Existenz gefährdenden Bedrohung eine einmal verlorene Existenz nicht mehr wiederzuerlangen ist. Daraus ergebe sich, dass die Aufhebung grundsätzlichen Leids ethisch dringlicher sei als das Hervorrufen höchsten Glücks. Zu Gute kommt dieser Argumentation, wie schon angeführt, auch der Umstand, dass individuelle Abweichungen dessen, was als Übel wahrgenommen wird, geringer sind als im Bereich der Güter (was nach Akerma gegen die **Privations-Theorie des Übels** spreche, wonach das Übel die Abwesenheit des Guten sei). Kaufmann spricht hierbei davon, dass Leid universalisierbar sei, Glück dagegen nicht (vgl. Kaufmann 1994: 23).

Das am häufigsten vorgebrachte Argument zugunsten eines NU gegenüber eines PU liegt in der weniger kontroversiellen Bestimmbarkeit der zu vermeidenden Übel gegenüber den zu erstrebenden Gütern:

„Über die Güter, die sie zu ihrem Glück zu benötigen glauben, sind die Menschen bekanntlich höchst uneins [...]. Über die Übel hingegen, die sie zur Vermeidung von Leiden auf keinen Fall erleben wollen, haben fast alle Menschen die gleiche Meinung: Niemand will getötet oder verletzt werden, niemand will in Not und Elend, niemand unter Zwang und Unterdrückung leben, und niemand will betrogen und getäuscht werden“ (Ropohl 2016: 316).

Kaufmann bringt den NU in eine Imperativ-ähnliche Formel: „Handle so, daß die Folgen deiner Handlung verträglich sind mit der größtmöglichen Vermeidung oder Verminderung menschlichen Elends“ (Kaufmann 1994: 24).

Auch der im PU oft angezweifelte Minderheitenschutz lässt sich nach Ansicht der ProponentInnen eines NU in ihm besser verwirklichen, da gerade diejenigen, die von Handlungen negativ betroffen werden, im Zweifel höher gewichtet würden. Kaufmann zufolge ist im NU somit eine dem Rawls'schen Differenzprinzip ähnliche Prioritätsregel gültig, welche im Falle einer Schlechterstellung von Betroffenen diesen ein besonderes ethisches Gewicht einräumt (vgl. Kaufmann 1994: 25). Weiterhin kann aber kritisch eingewandt werden, dass auch beim Versuch, so wenig Elend für so viele Menschen wie möglich entstehen zu lassen, eine Minderheit übrig bleibt: nämlich diejenigen, die nicht Bestandteil der Menge an Personen sind, für die so wenig Elend wie möglich verwirklicht werden konnte (vgl. Zecha 1999: 424).

Ein weiterer Einwand gegen den NU besteht im sogenannten „wohlwollenden-Weltzerstörer-Argument“ (vgl. Smart 1958; darauf aufbauend und ergänzend in kritischer Hinsicht Akerma 2000: 226-229). Wenn das Ziel des NU in der Minimierung von Leid besteht, dann ist sozusagen der direkteste, nachhaltigste und extremste Weg zu diesem Ziel die schmerzfreie, sofortige und restlose Auslöschung der Menschheit: Angenommen, es gäbe die Möglichkeit, durch Auslösen einer Superwaffe sämtliches menschliche Leben von einer Mikrosekunde auf die andere auszulöschen, sodass in keinem Moment jemand bei diesem Prozess leiden würde und es auch keine Hinterbliebenen gäbe, die den Verlust betrauern würden, so würde dies auf einen Schlag sämtliches (menschliche) Leid auf Null minimieren. Würde die Waffe dagegen nicht eingesetzt werden und das Leben seinen gewöhnlichen Verlauf nehmen, so wäre dies mit zumindest etwas Leid für die Menschheit (bzw. für manche Menschen) verbunden. Wenngleich der Einsatz dieser Waffe aus NU-Sicht geboten erscheint, wäre dies nach Smart eine Handlung, die für die meisten Menschen intuitiv moralisch falsch wirkt. Akerma wendet gegen Smarts Beispiel ein, dass die leidlose Auslöschung der Menschheit nur dann geboten scheint, wenn nicht eine Reduzierung des menschlichen Leides, sondern eine Eliminierung des menschlichen Leides angestrebt wird.

Im Rahmen der vorliegenden Arbeit wird kein reiner NU angestrebt, der ein einzelnes Prinzip zum höchsten Prinzip erklärt. Das bedeutet, dass Konsequenzen wie die vorhin genannte durch Einbezug anderer Prinzipien verhindert werden können. So würde etwa das Prinzip des Respekts vor der Autonomie (auf welches noch eingegangen wird) mit Sicherheit dagegensprechen, dass eine einzelne Person wie in Smarts Gedankenexperiment die Auslöschung der Menschheit über deren Kopf hinweg entscheidet. Dagegen werden im Weiteren vor allem auf Grund der Asymmetrie von Glück und Leid die Reduzierung von negativen Konsequenzen gegenüber der Förderung von positiven Konsequenzen der Human+ Technik stärker gewichtet werden. Wenn unter Rekurs auf Principlism-Ansätze daher unterschiedliche zu betrachtende Wertdimensionen und Prinzipien herangezogen werden, kann das NU-Prinzip als ein bei der Gewichtung hilfreiches Metaprinzip die Anwendung und Interpretation dieser anderen Prinzipien unterstützen, welches etwa folgendermaßen als **Metaprinzip des Negativen Utilitarismus** formuliert werden kann:

„Die Verminderung von Leid (Unglück, Übeln, ...) hat Vorrang gegenüber der Förderung von Glück. Wenn sich in technikethischen Betrachtungen Konflikte zwischen unterschiedlichen Handlungsmöglichkeiten oder Bewertungsmöglichkeiten ergeben, sollte im Abwägungsprozess einerseits eine Verminderung von Leid und negativen Konsequenzen Vorrang gegenüber der Förderung von Glück bzw. positiven Konsequenzen haben, und andererseits diejenige Alternative gewählt werden, die das geringste Leid verursacht“.

Ein Beispiel für den Anwendungsbereich des Metaprinzips vor dem Hintergrund der Verminderung von Leid bzw. der Förderung von Glück wäre, wenn man vor der Entscheidung steht, begrenzte Forschungsgelder für Technikprojekte zu verteilen: Projekt A betrifft die Entwicklung eines Verfahrens, welches aus günstigen Ausgangsstoffen preiswert die Herstellung von synthetischen, nahrhaften Lebensmitteln erlaubt. Projekt B avisiert die Entwicklung eines Verfahrens, welches neuartige, synthetische Geschmacksstoffe, die bisher noch niemand gekostet hat, zum Ziel hat. Aus Sicht des Metaprinzips des Negativen Utilitarismus liegt auf Grund von begrenzten Forschungsgeldern hier ein Konflikt zwischen zwei Handlungsmöglichkeiten vor: Förderung von Projekt A und Förderung von Projekt B. Da sich Projekt A auf die Bereitstellung von günstigen Lebensmitteln bezieht, und diese dazu dienen Hunger zu stillen (ein Paradebeispiel für eine personeninvariante Quelle von Leid), ist die Förderung von Projekt

A gegenüber Projekt B zu bevorzugen. Projekt B mag zwar auf Grund der Aussichten auf noch nie erlebte Geschmackswelten bei den KonsumentInnen besonderes zusätzliches Glück hervorrufen, aus NU-Sicht ist dies aber weniger gewichtig als die Verminderung von Leid bei Projekt A.¹⁷

Auf die grundsätzlichen Schwierigkeiten, die mit der parallelen Gültigkeit mehrerer Prinzipien (ohne ein alle anderen überragendes oberstes Prinzip) einhergehen, wird im Rahmen des Abschnitts zur Kritik am Principlism noch eingegangen. Im nächsten Kapitel dieser Arbeit werden verschiedene in der Technikethik übliche Frameworks beleuchtet – der gerade erwähnte Principlism wird dabei eine wichtige Rolle spielen, steht dieser doch im Hintergrund mehrerer dieser spezialisierteren Ansätze.

¹⁷ Dieses Beispiel wurde aus Anschaulichkeitsgründen erdacht. Hiermit soll nicht angedeutet werden, dass etwa für die Beseitigung des Welthungers tatsächlich neue technische Verfahren nötig sind, und dass der Welthunger nicht etwa auch seine Ursachen in ungleicher Verteilung haben kann.

5 ANSÄTZE UND FRAMEWORKS IN DER TECHNIKETHIK

5.1 PLURALITÄT VON ANSÄTZEN UND METHODEN

In der Einleitung wurden drei Hauptfragen dargelegt, deren Beantwortung Aufschluss über die ethische Zulässigkeit der Human+ Technik geben soll. Während damit das Erkenntnisinteresse allgemein festgestellt wurde, ist der Weg zur Antwortfindung bezüglich dieser drei Hauptfragen bisher nur teilweise aufgezeigt worden. Vorausgeschickt wurde bereits, dass zur Beantwortung erst ein reflexionstheoretisches Bewertungsframework erarbeitet werden muss. Dies erfolgt im Abschnitt „6.2 Harmonisierung des Bewertungsframework“. Um ein solches für die fragliche Human+ Technik aber fundiert entwerfen zu können, liegt es nahe, zuerst zu prüfen, mittels welcher Methodologien im Bereich der Angewandten (Technik-)Ethik gearbeitet wird, und welche Fragestellungen dabei im Vordergrund stehen. Weiter oben wurde bereits unter Rückgriff auf utilitaristische Gedanken das Metaprinzip des Negativen Utilitarismus formuliert. Außerdem wurde die häufige Kritik erläutert, dass (zumindest nach Ansicht verschiedener KritikerInnen) das utilitaristische Prinzip weitere Prinzipien als Korrektiv erfordere, z.B. um Gerechtigkeitsüberlegungen ausreichend berücksichtigen zu können. Daher soll in diesem Kapitel, als weitere Vorbereitung auf den Entwurf des reflexionstheoretischen Bewertungsframeworks und in Ergänzung zum schon formulierten Metaprinzip des Negativen Utilitarismus, ein Überblick über verschiedene technikethische Zugänge gegeben werden, um Anhand des Rückgriffs auf die Literatur weitere Prinzipien und Verfahrensweisen für die Beantwortung von technikethischen Fragen zu erarbeiten.

Human+ ist ein Forschungsprojekt im Gebiet Informations- und Kommunikationstechniken. Im Bereich der Technikethik im Allgemeinen und von Informationstechnologie-bezogener Technikethik im Besonderen gibt es eine Reihe von konkreteren Ansätzen, in deren Hintergrund eine Denkweise steht, die sich nicht auf ein oberstes ethisches Prinzip bezieht, sondern mehrere relevante Bewertungsdimensionen (prima facie) gleichrangig nebeneinander sieht: Principlism bzw. Ansätze der Prinzipien mittlerer Reichweite. Deren idealtypische Ausgestaltung findet sich in der Form der ursprünglich auf medizinische und bioethische Kontexte zugeschnittenen Vier-Prinzipien-Variante nach Beauchamp und Childress (2009), auf die zu Beginn dieses Kapitels eingegangen wird. Insgesamt ist im Bereich der Technikethik aber ein Methodenpluralismus vorherrschend.

Dies belegen Reijers et al (2018), die eine Meta-Analyse von 136 Forschungsarbeiten¹⁸ zu im Bereich der Technikerforschung bzw. Technikentwicklung verwendeten ethischen Ansätzen durchgeführt haben. Insgesamt haben Reijers et al 35 unterschiedliche Methodologien identifiziert, die entweder in spezifischen Bereichen (wie Gesundheitstechnologie oder Informationstechnologien) verstärkt Anwendung finden, oder ohne besonderen Fokus sich allgemein auf Ethik im Bereich Forschung und Entwicklung beziehen. Die Analyse zeigt, dass die meisten der durch ihren Suchalgorithmus entdeckten, Technikethik-bezogenen Forschungsarbeiten zu Forschung und Entwicklung im Bereich Health Technology Assessment durchgeführt wurden, dann aber bereits gefolgt von Arbeiten im Gebiet Informationssysteme und Computerwissenschaften (vgl. Reijers et al 2018: 1443ff).

Für den Bereich Informationstechnologien (dem auch Human+ zuzurechnen ist) haben sich demzufolge eine Reihe von Methoden etabliert, die häufiger diskutiert und von unterschiedlichen AutorInnen in mehreren Forschungsarbeiten weiterentwickelt wurden. Reijers et al verweisen hier exemplarisch auf verschiedene Ansätze, z.B. Value Sensitive Design (vgl. u.a. Friedmann et al 2006), ETICA (vgl. u.a. Stahl 2011), Ethical Impact Assessment (vgl. u.a. Wright 2011). Hier sei nur vorausgeschickt, dass die unter Bezug auf Reijers et al in diesem Kapitel wiedergegebenen, exemplarischen Ansätze weiter unten noch genauer erläutert werden.

Zahlreiche der von Reijers et al analysierten Texte befassten sich nicht mit bestimmten Technikbereichen, sondern beleuchteten generell die Thematik von Ethik für Forschung und Entwicklung. Beispielsweise verweisen diese besonders auf die Notwendigkeit von Technikethik im Entwicklungsbereich und welche Schritte und Kriterien dabei wichtig sind (vgl. Palm und Hansson 2006), oder entwickeln bereichsübergreifende Methodologien wie z.B. die Ethical Matrix (vgl. Mephram et al 2006), szenariobasierte Ansätze (vgl. Boenink et al 2010) oder Checklisten-basierte Ansätze (z.B. Brey 2012). Ein bedeutender Teil befasst sich zudem mit der Adaption und Kritik des Principlism bzw. von prinzipienbasierten Ansätzen. Die von Reijers et al 2018 angeführten Quellen weisen dabei mitunter Überschneidungen auf, etwa wenn dieselben AutorInnen sowohl spezifische Arbeiten im Bereich Informationstechnologien wie auch generellere zur Ethik im Forschungsbereich verfassten. Die genaue Klassifikation von Reijers et al 2018 wird daher im weiteren Verlauf der Arbeit nicht im Vordergrund stehen.

¹⁸ Die 136 Forschungsarbeiten sind allesamt auf Englisch publiziert worden. Ein Teil wurde aus Wissenschaftsdatenbanken bzw. Datenbankverbunden wie „Web of Knowledge“, „Scopus“ und „Springerlink database“ entnommen, auf Basis einer Suche über verschiedene Suchbegriffe, die das Themenfeld eingrenzen (z.B. „ethics“, „ethical“, „technology“, „impacts“, „assessment“, verknüpft mit Begriffen die das Themenfeld auf den Entwicklungsbereich eingrenzen, z.B. „research“ oder „innovation“) und einer daran anschließenden manuelle Sichtung des Materials. Ein anderer Teil der Ergebnisse wurde durch Prüfen der Literaturverzeichnisse dieser Quellen eruiert. Thematisch sind sie eingegrenzt auf das engere Feld der Technikethik, und bilden somit nur einen Ausschnitt aus dem möglichen ELSI-Spektrum ab, indem sie Forschungsarbeiten zu legislativen und sozialen Aspekten von Techniken bewusst ausklammern. Während ein kleiner Teil der Arbeiten bis auf 1990 zurückdatiert, ist der überwiegende Anteil erst nach 2005 publiziert worden (vgl. für methodologische Aspekte hierzu Reijers et al 2018: 1439-1442).

Interessanterweise finden sich im sehr ambitionierten Überblick von Reijers et al 2018 kein einziges Mal die Begriffe Utilitarismus [utilitarianism], Kantianisms [kantianism] oder Deontologie [deontology]. Reijers et al gruppieren die von ihnen identifizierten Ansätze stattdessen zu drei verschiedenen Kategorien von Methoden, die sich auf unterschiedliche Stadien der Technikreife bzw. Phasen des Technikentwicklungsprozesses beziehen, und arbeiten für diese unterschiedliche Erkenntnisziele heraus, die somit auch als idealtypische technikethische Erkenntnis- bzw. Betätigungsziele gelten können (vgl. Reijers et al 2018: 1447-1454):

1. **Ex Ante Methoden** im Bereich von gerade erst sich in einer sehr frühen Entwicklungsphase befindlichen Techniken, mit den Zielen
 - a. Identifizierung von potentiellen, emergierenden Techniken
 - b. Konstruktion von Szenarien über zukünftige ethische Auswirkungen
 - c. Evaluation potentieller ethischer Auswirkungen
 - d. Bewertung des Status von unsicheren normativen Behauptungen
2. **Intra Methoden** (bezogen auf Techniken im Prototypstatus, d.h. während der Überführung von Konzepten in konkrete Entwicklungen), mit den Zielen
 - a. Integration von EthikerInnen in den Forschungs- und Entwicklungsprozess
 - b. Aufdeckung ethischer Probleme im Technikdesign
 - c. Einbettung von Werten im Technikdesign
 - d. Organisation der „Anwendung von Ethik“ im Designprozess; sowie
3. **Ex Post Methoden** (die sich auf bereits existierende Techniken beziehen), mit den Zielen
 - a. Identifizierung ethischer Probleme die durch bereits existierende Techniken aufgeworfen werden
 - b. Analyse ethischer Probleme die durch bereits existierende Techniken aufgeworfen werden
 - c. Organisation der Governance hinsichtlich ethischer Analyse
 - d. Unterstützung im Bereich ethischer Entscheidungsfindung hinsichtlich bereits existierender Techniken

Das mit 31.3.2020 abgeschlossene Human+ Projekt durchlief die Ex Ante bzw. Intra Phase, d.h. beginnend mit einem Projektantrag, bei welchem nur erste Konzepte zu den Entwicklungen skizziert wurden, hin zu einer Entwicklung in Richtung eines TRL, welches die grundsätzliche Funktionsfähigkeit der Konzepte in einem labornahen Demonstrator-Setting zeigt (TRL4). Wenngleich das Projekt also bereits abgeschlossen ist, sind die entwickelten Komponenten noch lange nicht auf einem markt- oder anwendungsreifen Niveau.

Für den Kontext der vorliegenden Arbeit lässt sich insgesamt feststellen, dass somit neben „Grundlagentheorien“ wie dem schon vorgestellten Utilitarismus in der Technikethik auch eine Vielzahl an schon an den Technikethikbereich angepassten, spezialisierteren Methodologien Anwendung findet. Wie schon angeführt, steht im Hintergrund einiger dieser spezialisierteren Methodologien der Principlism oder zumindest ein Prinzipien-orientiertes Vorgehen.

Daher wird im nächsten Abschnitt der Principlism in seiner Gestalt von Beauchamp und Childress (2009) beleuchtet, bevor anschließend eine Auswahl dieser spezialisierteren Ansätze (von denen nicht alle, aber einige, ein Principlism-Fundament aufweisen) erläutert wird, um die Grundlagen für deren Harmonisierung zu einem ethischen Referenzrahmen für Human+ in Abschnitt „6.2 Harmonisierung des Bewertungsframework“ ab Seite 93 zu legen.

5.2 PRINCIPLISM

5.2.1 Begriffsbestimmung – Principlism und Prinzipien mittlerer Reichweite

In diesem Abschnitt wird auf den in der Angewandten Ethik zentralen Ansatz des Principlism eingegangen, und es werden ethische Prinzipien rekapituliert, die besonders in der Beantwortung der in der Einleitung angesprochenen Kernfrage 3 zur Gefährdung der Lebensbedingungen bzw. der Lebensqualität von gegenwärtigen oder zukünftigen Menschen durch die Folgen und Nebenwirkungen der planmäßigen Human+ Techniknutzung zur Anwendung kommen werden.

Ideengeschichtlich ist der Begriff „Principlism“ stark mit dessen als paradigmatisch geltenden und als „Principles of Biomedical Ethics“ erstmals 1979 publizierten Ausarbeitung bei Tim L. Beauchamp und James F. Childress verbunden. Beauchamp und Childress hatten das Konzept also zunächst vor allem im Bereich der Medizinethik prominent gemacht, obwohl wie noch gezeigt wird, Principlism mittlerweile in zahlreiche Bereiche der Angewandten Ethik Verwendung findet. „Principlism“ als Begriff wurde erstmals von K. Danner Clouser und Bernard Gert im Jahr 1990 verwendet, die sich auf die Ausarbeitung des Vier-Prinzipien-Ansatzes bei Beauchamp und Childress allerdings in negativ-abschätziger Weise bezogen (vgl. Rauprich 2012: 590). Der Begriff hat mittlerweile aber eine Entfärbung durchlaufen und wird in der gegenwärtigen Diskussion als neutral-deskriptiver Terminus für Prinzipien-mittlerer-Reichweite-Ansätze genutzt. Im deutschen Sprachraum wird mitunter alternativ auch von „Prinzipienethik“ gesprochen, wenngleich einer Anregung von Fenner (2010: 24) folgend dies im Rahmen der vorliegenden Arbeit nicht getan wird, um Verwechslungen mit Ansätzen, die ein einzelnes oberstes und nicht bloß prima facie gültiges Prinzip vertreten, zu vermeiden.

Principlism bezeichnet zumindest in der Form von Beauchamp und Childress im Kern einen ethischen Ansatz, der vollumfänglich aus fünf Komponenten besteht (vgl. Rauprich 2012: 591):

- a. ein Rahmenwerk bestehend aus prima facie gültigen Prinzipien
- b. Spezifizierung [specification] als Methode um moralische Normen auf konkrete Fälle oder Problemlagen zu übertragen
- c. die Analyse von Problemen im Bereich der angewandten Ethik hinsichtlich des ihnen zugrundeliegenden Normenkonflikts
- d. Spezifizierung und Abwägung [balancing] als Methoden um moralische Konflikte aufzulösen
- e. eine Theorie der Rechtfertigung [justification]

Wie später noch sichtbar wird, besteht aber nicht jede ethische Forschungsarbeit, die sich auf ein Principlism-ähnliches Fundament beruft, aus all diesen fünf Komponenten. In jedem Fall zentral ist der Rekurs auf die erwähnten prima facie gültigen Prinzipien, welche auch als „Prinzipien mittlerer Reichweite“ bezeichnet werden. Der Zusatz „mittlere Reichweite“ ergibt sich daraus, dass sie weniger abstrakt sind verglichen mit den Hauptprinzipien „großer“ ethischer Theorien wie des Kantianismus oder des schon vorgestellten Utilitarismus, dem Greatest Happiness Principle (vgl. Bleisch und Huppenbauer 2014: 72). Zudem denotiert die Pluralform „Prinzipien“, dass es kein einzelnes oberstes Prinzip gibt, sondern mehrere. Diese sind **prima facie gültig**, d.h. sie sind zwar zunächst alle gültig, können aber nicht absolute Geltung beanspruchen, sondern müssen letztendlich vor dem Hintergrund eines bestimmten Gegenstandsbereichs jeweils mit- und gegeneinander abgewogen werden (vgl. Rauprich 2012: 591).

In früheren Kapiteln wurde gezeigt, dass es bei der Technikfolgenabschätzung und dem Utilitarismus eigentlich irreführend ist, nur von „dem Utilitarismus“ oder „der Technikfolgenabschätzung“ zu sprechen, weil es eine Vielfalt an möglichen Ausformungen und Schwerpunktsetzungen gibt, die sich um einen zentralen theoretischen Kern gruppieren. Ähnlich ist es nach Ansicht des Verfassers der vorliegenden Arbeit auch beim Principlism, sodass eigentlich von Principlism-Ansätzen gesprochen werden müsste. Zwei davon sollen im Folgenden vorgestellt werden, da sie die Eckpunkte eines Spektrums möglicher Anwendungs- oder Ausarbeitungsgrade markieren, zwischen denen auch das noch zu entwickelnde Human+ Bewertungsframework sich bewegen wird. Der erste Ansatz bezieht sich auf den Principlism, wie er direkt bei Beauchamp und Childress angelegt ist. Der zweite Ansatz bezieht sich auf eine Ausformung, die als Ethical Matrix bezeichnet wird, und die auf eine Arbeitsgruppe um Ben Meppham zurückgeht, welche sich mit der Anwendung des Principlism besonders auf lebensmittel- und ökologiebezogene Fragestellungen beschäftigt hat, mittlerweile ebenfalls in verschiedene Anwendungsbereiche diffundiert ist und dabei methodologisch den Ansatz von Beauchamp und Childress adaptiert hat.

5.2.2 Der Vier-Prinzipien-Ansatz von Beauchamp und Childress

In den „Principles of Biomedical Ethics“ (Beauchamp und Childress 2009) entwerfen die Autoren die Grundlegung zu dem, was mittlerweile als Principlism bezeichnet wird. Im Folgenden werden die vier Prinzipien und die dazugehörige Theorie ihrer Rechtfertigung erläutert, und dann die Methoden der Spezifikation bzw. Abwägung vorgestellt.

Im Zentrum des Principlism stehen **vier Prinzipien mittlerer Reichweite**, die die Autoren auf einem relativ hohen Abstraktionsgrad beschreiben, und sodann hinsichtlich des biomedizinischen Kontexts weiter erläutern bzw. einem Prozess namens „Spezifizierung“ unterziehen. Während daher ein Teil der Ausführungen für den Kontext von IKT-bezogener Forschung wenig relevant erscheint, sind vor allem die grundlegenden, abstrakteren Überlegungen zu den Prinzipien von allgemeinem Interesse (und wurden auch, wie noch dargestellt wird, in technikethischen Analyseframeworks breit rezipiert).

Die vier Prinzipien verstehen sich als analytisches Framework, als abstrakter Ausgangspunkt und als Richtlinien für die Formulierung von spezifischeren Regeln (vgl. ebd.: 12). Streng genommen handelt es sich nicht um vier Prinzipien, sondern um vier Prinzipienbündel [four clusters of moral principles]. **Prinzipien** sind die „[...] most general and comprehensive norms [...]“, während die damit verwandten und nicht ganz klar davon abgegrenzten **moralischen Regeln** „[...] more specific in content and more restricted in scope [...]“ sind (ebd.: 13). Prinzipien sind daher keine konkreten Handlungsanleitungen (wie es die Regeln sind), aber beide begründen Rechte und Verpflichtungen. Für den Kontext der vorliegenden Arbeit bedeutet das, dass in erster Linie die Prinzipien im Vordergrund stehen, während die moralischen Regeln einen bereits stärkeren biomedizinischen Zuschnitt haben und hier weniger von Interesse sind.

Folgende vier Prinzipien(bündel) finden sich bei Beauchamp und Childress (deutsche Begriffe sind entlehnt von Bleisch und Huppenbauer 2014: 72):

- Prinzip der Selbstbestimmung [respect for autonomy]: Nach Beauchamp und Childress akzeptieren alle Theorien der Selbstbestimmung zwei Bedingungen: **Freiheit** (als Unabhängigkeit von kontrollierenden Einflüssen) und **Handlungsfähigkeit** (als Vermögen, intentional zu handeln; vgl. Beauchamp und Childress 2009: 100). Handlungen können dabei in unterschiedlichen Graden autonom sein, je nachdem wie sehr die Voraussetzungen vorliegen (und somit nehmen die Autoren Abstand von einem Konzept von Autonomie, welche umfassende Kontrolle über alle Facetten der eigenen Wünsche, inkl. einer Änderung der eigenen Präferenzstruktur, erfordern würde). Für die Autoren hat das Prinzip der Selbstbestimmung damit sowohl eine positive Verpflichtung wie eine negative Verpflichtung: „As a *negative* obligation: Autonomous actions should not be subjected to controlling constraints by others“ (ebd.: 104). Im notwendigen Prozess der Spezifizierung können zwar

Ausnahmen hierfür geschaffen werden, eine Generalausnahme im Sinne von „außer wenn ihre Handlungen oder Gedanken anderen Menschen schaden“ wäre aber unzulässig. Und weiter: „As a *positive* obligation, this principle requires both respectful treatment in disclosing information and actions that foster autonomous decision making“ (ebd.: 104). Das heißt, es ist Bestandteil dieses Prinzips, betroffene Personen dabei zu unterstützen, autonome Entscheidungen treffen zu können.

- Prinzip der Nichtschädigung [nonmaleficence]: Beauchamp und Childress unterscheiden zwischen Normen, die sich aus Geboten der Nichtschädigung ergeben, und solchen, die sich aus Normen des Wohltuns ergeben. Nichtschädigung meint: „One ought not to inflict evil or harm“ (ebd.: 151). “[...] nonmaleficence requires only intentionally refraining from actions that cause harm. Rules of non-maleficence therefore take the form ‘Do not do X.’“ (ebd.: 151). Schädigung meint dabei nicht zwangsläufig ein moralisches Versagen (im Sinne von „wronging“), sondern ist gemeint im nicht-normativen Sinne von „[...] thwarting, defeating, or setting back some party’s interests. Therefore, a *harmful* action by one party may not be wrong or unjustified, although acts of harming in general are *prima facie* wrong“ (ebd.: 152). Während der Begriff Schädigung somit eine große inhaltliche Breite aufweist (und manche AutorInnen auch Unwohlsein, Erniedrigung, Genervtheit etc. inkludieren), verweisen die Autoren darauf, dass Konsens darüber bestehe, dass bedeutende körperliche Schäden Paradefälle von Schädigungen bedeuten, und sie sich auf physische Schädigungen, v.a. Schmerz, Behinderung, Leiden und Tod konzentrieren (ohne die Bedeutung geistiger Schäden oder anderer Interessensverletzungen zu bestreiten; vgl. Beauchamp und Childress 2009: 153). Verpflichtungen aus diesem Prinzip schließen nicht nur die Vermeidung von Schädigungen ein, sondern auch die Vermeidung, andere Menschen dem **Risiko** einer Schädigung auszusetzen.¹⁹ Diese Regeln verweisen somit auf Unterlassungen, die unparteiisch und universell zu befolgen sind, und der Verstoß gegen sie ist üblicherweise mit juristischen Strafen verknüpft (vgl. ebd.: 199).
- Prinzip der Fürsorge [beneficence]: In Ergänzung und in Unterschied zum Prinzip der Nichtschädigung (welches ein Unterlassen impliziert) steht das Prinzip der Fürsorge für die aktive Handlungskomponente, die darauf abzielt, aktiv das Wohlbefinden anderer zu fördern. Es umfasst zwei (Unter-)Prinzipien: „*Positive beneficence* requires agents to provide benefits to others. *Utility*²⁰ requires that agents balance benefits, risks and costs to produce the best overall results“ (ebd.: 197).

¹⁹ Gerade bei mit größeren Unsicherheiten behafteten Thematiken, wie etwa bei der Human+ Technik, deren genauer späterer Funktionsumfang wie auch die konkreten Anwendungsbedingungen während der Projektlaufzeit nur eckpunkthaft umrissen werden konnten, scheint es ratsam, daher auch die Risiken besonders in den Blick zu nehmen.

²⁰ Der Unterschied zum utilitaristischen Nutzenprinzip besteht darin, dass das Nutzenprinzip im Falle des Principlism nicht ein höchstes Prinzip ist, sondern lediglich als *prima facie* Prinzip besteht, in welchem sich nicht die Fülle moralischer Verpflichtungen erschöpft (vgl. Beauchamp und Childress 2009: 198). Da in der Behandlung des Nutzenprinzips auf Verrechnungsweisen rekuriert wird, die für die vorliegende Masterarbeit nicht relevant sind (z.B. Cost-Effectiveness und Cost-Benefit Analysis, vgl. ebd. 222) wird im Weiteren auf jene Aspekte nicht eingegangen.

Wohltun wird von den Autoren sehr breit gefasst: “We use beneficence to cover beneficent action more broadly, so that it includes all forms of action intended to benefit other persons” (ebd.: 197). Das Prinzip des Wohltuns verweist darauf, dass es zumindest einige Arten des Wohltuns gibt, die moralisch verpflichtend sein können – auch wenn viele es nicht sind (v.a. solche, die Selbstaufopferung oder extremen Altruismus implizieren, sind selten). Unter die moralisch verpflichtenden Formen rechnen die Autoren den Schutz und die Verteidigung der Rechte anderer, die Verhinderung der Schädigung anderer, die Beseitigung von Bedingungen die zur Schädigung anderer führen, die Hilfe von Personen mit Behinderungen, und die Rettung von Personen in Gefahr. Regeln aus diesem Prinzip erfordern somit aktives Tun, müssen nicht zwangsläufig unparteiisch angelegt sein (so können sich z.B. Naheverhältnisse auf die moralische Verpflichtung auswirken) und sind normalerweise nicht mit juristischen Sanktionen behaftet, wenn sie nicht befolgt werden (vgl. ebd.: 199). Ein besonderer Fall, in welchem die Umstände zur unparteiischen Hilfe verpflichten (und somit sich nicht aus vertraglichen Verpflichtungen oder Naheverhältnissen ergeben), sind Fälle der Rettung aus Gefahrensituationen, die noch dazu besonderen Bedingungen genügen müssen. Hierbei handelt es sich um einen Parafall des allgemeinen Wohltuns [general beneficence].²¹ Zusätzlich können sich aus speziellen moralischen Beziehungen (z.B. Freundschaft, oder therapeutische ÄrztInnen-PatientInnen-Beziehungen) oder aus sogenannte Rollenverpflichtungen Erfordernisse des spezifischen Wohltuns [specific beneficence] ergeben (vgl. ebd.: 205).

- Prinzip der Gerechtigkeit [justice]: Grundsätzlich betrifft dieses Prinzip die faire Verteilung von Kosten und Risiken. Die Autoren verweisen darauf, dass in der Beurteilung von (v.a. nicht-therapeutischer) Forschung lange der Schutz der Forschungssubjekte vor Ausbeutung und Schaden (d.h. vor unfairer Verteilung von Lasten) gegolten habe, wobei seit 1990 ein Paradigmenwechsel durch die Forschung an HIV-Arzneien eingetreten sei, und der Fokus dadurch auf die möglichen Vorteile für Forschungssubjekte und den fairen Zugang zu dieser Forschung gelegt wurde. Sowohl **Schutz vor Ausbeutung** wie **fairer Zugang zur Forschung** sind somit zwei Facetten (vgl. ebd.: 241). In ihrer Beschreibung des Prinzips gehen die Autoren auf unterschiedliche Begriffe von Gerechtigkeit ein, z.B. Gerechtigkeit als Fairness, Verdienst [desert] und Anspruch [entitlement], oder als Verteilungsgerechtigkeit [distributive justice] im Sinne einer „[...] fair, equitable and appropriate distribution determined by justified norms that structure the terms of social cooperation“ (ebd.: 241), z.B. Verteilungsstrategien. In einem breiten Sinn sei damit auch die Verteilung aller Rechten und Pflichten

²¹ Beauchamp und Childress führen hierzu fünf Bedingungen an, die allesamt erfüllt sein müssen, damit eine Person X eine moralische Verpflichtung des Wohltuns gegenüber Person Y hat: „1. Y is at risk of significant loss of or damage to life or health or some other major interest. 2. X’s action is necessary (singly or in concert with others) to prevent this loss or damage. 3. X’s action (singly or in concert with others) has a very high probability of preventing it. 4. X’s action would not present very significant risks, costs, or burdens to X. 5. The benefit that Y can be expected to gain outweighs any harms, costs or burdens that X is likely to incur“ (Beauchamp und Childress 2009: 202).

in einer Gesellschaft gemeint, was es von anderen Arten der Gerechtigkeit wie gerichtlicher Gerechtigkeit (als Zuteilung von gerechten Strafen) unterscheidet. Da nicht ein einzelnes Prinzip alle Probleme der Gerechtigkeit abdecken könne, referieren die Autoren auf mehrere Prinzipien der Gerechtigkeit, die untereinander abgewogen und weiter spezifiziert werden müssen (für den Gesundheitskontext). Während das **formelle Prinzip** der Gerechtigkeit auf Aristoteles zurückgehend besagt, dass Gleiches gleich und Ungleiches ungleich behandelt werden soll, ist es völlig inhaltsleer und müsse daher mit **materiellen Prinzipien** „gefüllt“ werden, die spezifizieren, in welchen Hinsichten von (Un)Gleichheit gesprochen wird, und auf welche Weise die Vergleiche angestellt werden sollen (vgl. ebd.: 242). Thematisch schränken die Autoren die Reichweite des Prinzips auf Grundbedürfnisse ein, zu denen sie auch Gesundheit rechnen. Mögliche materielle Verteilungsprinzipien umfassen die Zuteilung zu jeder Person nach gleichem Anteil, nach Bedarf, nach Bemühen [effort], nach Beitrag [contribution], nach Verdienst [merit], oder nach marktwirtschaftlichem Austausch, wobei angenommen wird, dass mit diesen Prinzipien erneut prima facie Verpflichtungen einhergehen, deren Gewicht nicht unabhängig von bestimmten fraglichen Gütern und Bereichen verhandelt werden kann. Sie führen im weiteren unterschiedliche Gerechtigkeitstheorien (z.B. utilitaristischer, libertarianistischer, kommunitaristischer oder egalitärer Natur) an, wobei sie letztendlich eine „kosmopolitische“ Theorie auf egalitärem (Raws'schen) Fundament entwickeln, worauf auf Grund ihres Fokus auf den Gesundheitsbereich hier ebenfalls nicht weiter eingegangen werden muss. Das Prinzip der Gerechtigkeit nimmt insofern einen Sonderstatus unter den Prinzipien der biomedizinischen Ethik ein, weil es über die ÄrztInnen-PatientInnen-Beziehung hinausgeht und somit auch institutionelle Aspekte des medizinischen Bereichs betrifft (vgl. Rauprich 2012: 593).

Um dem Vorwurf eines eklektischen Zusammenstellens der obigen Prinzipien zu begegnen, haben ab der vierten Edition ihres Buches Beauchamp und Childress die bis dahin eher unterentwickelte Begründung der obigen vier Prinzipien mit einer ausdrücklichen Theorie der Rechtfertigung untermauert (vgl. Rauprich 2012: 596). Diese Theorie der Rechtfertigung besteht aus einer **Theorie einer gemeinsamen Moral** [common morality], ergänzt um eine Methode zur Erreichung eines Überlegungsgleichgewichts [reflective-equilibrium].

Gemeinsame Moral meint “[...] the set of norms shared by all persons committed to morality. [...] The common morality is applicable to all persons in all places, and we rightly judge all human conduct by its standard” (Beauchamp und Childress 2009: 3). Damit ist folglich nicht gemeint, dass alle Menschen sich an die gemeinsame Moral halten – sondern diejenigen Personen, die sich moralisch verhalten, halten sich auch an die Normen der gemeinsamen Moral. Beispiele dafür sind nach Ansicht der Autoren

verschiedene Imperative, wie „Töte nicht“, „Füge anderen kein Leid zu“, „Verhindere Böses oder Schaden“, „Befreie Menschen aus Not“, „Sage die Wahrheit“, „Bestrafe nicht die Unschuldigen“, usw. (vgl. ebd.: 3). Daneben beziehen sie auch Tugenden ein, z.B. Ehrlichkeit, Integrität, usw. Die gemeinsame Moral ist dabei nicht überzeitlich, sondern kann sich (im Prinzip) historisch ändern, wird aber dennoch als universell über Kulturgrenzen hinweg als gemeinsam geteilt und wirksam erachtet. Moralischer Pluralismus besteht in spezifischen Moralitäten (die z.B. nur für bestimmte Religionen, Milieus, Berufsstände, etc. gelten), aber nicht in der gemeinsamen Moral, so die Konzeption der Autoren. Die von Beauchamp und Childress aufgestellten, sehr abstrakten vier Prinzipien, stellen einen Grundstock der gemeinsamen Moral dar (vgl. ebd.: 387). Die Autoren halten es zwar nicht für völlig ausgeschlossen, dass sich die allgemeinsten Normen der gemeinsamen Moral verändern, aber historisch sei ein Fall, in dem ein moralisches Prinzip der gemeinsamen Moral nur für eine bestimmte Zeit gültig gewesen wäre, nicht vorgekommen (vgl. ebd.: 390). Während sich die abstrakten Prinzipien nicht ändern, kann sich aber der Raum ihrer Anwendung verändern, z.B. durch Änderungen der Gruppe derjenigen, denen moralischer Status zugeschrieben wird.

Durch die Berufung auf die Idee der gemeinsamen Moral handele es sich nach Ansicht der ProponentInnen des Principlism bei diesem (entgegen Vorwürfen, dass dieser eine Befangenheit in Richtung westlicher Kultur habe) um ein „[...] transcultural, transnational, transreligious, transphilosophical framework for ethical analysis“ (Gillon 1994b: 332; zitiert nach Mepham 2000: 167).

Die vier Prinzipien sowie die übrigen Normen der gemeinsamen Moral sind der Startpunkt für die Ausarbeitung von detaillierteren Normen (z.B. im Bereich der Bioethik), d.h. sie müssen situations- und kontextspezifisch angepasst werden. Der Prozess hierzu wird in Anlehnung an John Rawls' Theorie der Gerechtigkeit als Methode beschrieben, durch welche **wohlüberlegte Urteile** [considered judgments] (also Annahmen, die für uns einen besonders gesicherten Erkenntnisstatus haben und besonders wenig durch Vorurteile oder Eigeninteressen beeinflusst sind) in ein **kohärentes reflexives Gleichgewicht** [reflective equilibrium] mit anderen, weniger gesicherten Urteilen gebracht werden (meistens, aber nicht immer, indem die Urteile aus letzterer Kategorie modifiziert werden). Für Beauchamp und Childress nun sind diese wohlüberlegten Urteile der gemeinsamen Moral zu entnehmen (vgl. Rauprich 2012: 596), welche auch den Rahmen für das reflektive Gleichgewicht bilden, der nicht überschritten werden darf. Als Ausgangspunkt und Korrektiv bilden sie sozusagen den Rahmen, innerhalb dessen man für spezifischere Kontexte umfassendere und inhaltsreichere Prinzipien ableitet, die zueinander in einem

kohärenten, also widerspruchsfreien, Verhältnis stehen müssen (als Ergebnis des Abwägungsprozesses zumindest).²²

Der Prozess, in welchem diese inhaltlich gehaltvolleren, sekundären Normen (Prinzipien, Regeln, etc.) aus den übergeordneten, aber abstrakteren primären Normen der gemeinsamen Moral gebildet werden, nennt sich Methode der **Spezifizierung**:

„Specification is a process of reducing the indeterminate character of abstract norms and generating more specific, action-guiding content [...]. Specification is not a process of producing or defining general norms such as those in the common morality; it assumes that they are available. [...] [It] is accomplished by narrowing the scope of the norms, not by explaining what the general norms mean. [...] Specification, then, does not merely analyze meaning; it adds content“ (Beauchamp und Childress 2009: 17).

Umgelegt auf die Problematik der vorliegenden Arbeit, soviel kann schon vorausgeschickt werden, wird dies bedeuten, bei den verwendeten Prinzipien mittlerer Reichweite eine „informations- und kommunikationstechnologische“ Spezifizierung hinzuzufügen (vgl. als Vorgriff auch „Tabelle 9: Human+ Bewertungsframework: verwendete Prinzipien mittlerer Reichweite“ auf Seite 98). So werden etwa weiter unten Aspekte des Datenschutzes und der Privatheit ausdrücklich unter das Prinzip des Respekts der Autonomie subsumiert.

Ein anderes Beispiel hierfür wäre die Spezifizierung des Prinzips der Selbstbestimmung (bzw. des Respekts vor der Autonomie) zum Prinzip der Informierten Zustimmung (vgl. Rauprich 2012: 593f). Der Prozess der Spezifizierung beinhaltet **detaillierte Explikation**. Es ist eine Form der Interpretation von moralischen Normen, ausgehend von inhaltsthünnen Prinzipien hin zu konkreteren, handlungsleitenden Normen, indem sie mit Inhalt angereichert und dadurch weniger abstrakt werden, ihre Geltungsreichweite bestimmt wird, und so auf tatsächliche, konkrete Situationen anwendbar werden. Wenngleich die Grundidee schon bei Beauchamp und Childress angelegt war, verweist Rauprich auf die erstmalige explizite Ausarbeitung durch Henry Richardson in 1990, wonach Spezifizierung die Herstellung

²² Im Rahmen der vorliegenden Arbeit wird nicht ausdrücklich mit dem Konzept des reflexiven Gleichgewichts gearbeitet. Umgelegt auf Human+ würde dies aber wohl bedeuten, alle Bestandteile eines Systems von Überzeugungen [beliefs] in Einklang zu bringen. Den Anforderungen des Vier-Prinzipien-Ansatzes folgend wären die Prinzipien mittlerer Reichweite (bzw. deren Spezifikationen) Teil eines solchen Systems, von denen nicht ohne sehr guten Grund abgerückt werden soll, da sie den „Rahmen“ für das Reflexionsgleichgewicht bilden. Weitere Bestandteile wären Überzeugungen, die sich auf das Ausmaß der technischen Möglichkeiten des Human+ Systems (z.B. Prognose von Migrationsbewegungen), Missbrauchsmöglichkeiten (z.B. Manipulationsmöglichkeiten), der Anwendungsbedingungen (z.B. welche BOS verwendet das System in welchem Setting) beziehen. Herstellen eines Reflexionsgleichgewichts würde bedeuten, den Gehalt dieser Überzeugungen auf eine Art iterativ zu verändern, dass Kohärenz zwischen diesen Überzeugungen herrschen kann. Beispielsweise die Anwendungsbedingungen auf eine Art und Weise zu skizzieren, dass sie kohärent zu den Erfordernissen der Prinzipien mittlerer Reichweite passen.

einer Relation zwischen einer weniger spezifischen und einer mehr spezifischen Norm ist. Diese Relation muss einerseits der syntaktischen Bedingung [syntactic condition] genügen durch Hinzufügen oder Festlegen von Bedingungen (z.B. wo, wann, wodurch) gegenüber wen oder durch wen eine weniger spezifische Norm in einem bestimmten Kontext zu befolgen ist. Andererseits muss sie der semantischen Bedingung [semantic condition] genügen, durch Eingrenzen ihres Bezuges [extensional narrowing], wonach alles das die spezifische Norm erfüllt auch die abstraktere Norm erfüllen muss, aber nicht umgekehrt. Werden diese zwei Bedingungen befolgt, so ist sichergestellt, dass trotz unterschiedlichem Inhalt die normative Verpflichtung der abstrakteren Norm auch bei der spezifischeren Norm erhalten bleibt. Die solcherart spezifizierten Normen können ihrerseits erneut immer detaillierter auf speziellere Anwendungsgebiete zugeschnitten werden, sodass durch iterative Spezifizierung ein ganzes Netzwerk an Normen erzeugt wird, um mit Problemen der Angewandten Ethik in einem bestimmten Anwendungsbereich umzugehen. Eine bestimmte, weniger abstrakte Norm muss zudem nicht aus einer einzigen übergeordneten Norm angeleitet werden, sondern kann ihre normative Kraft auch aus mehreren abstrakteren Normen erhalten.

Sowohl die abstrakten vier Prinzipien als auch die aus ihnen abgeleiteten Spezifizierungen sind prima facie Pflichten, das heißt, sie können in Konflikt miteinander geraten. In einem solchen Fall muss festgestellt werden, welche in einer bestimmten Situation wichtiger sind, also die „tatsächlichen“ Pflichten sind. Dass Prinzipien (sowie Regeln, Verpflichtungen und Rechte) in Konflikt miteinander geraten, ist für die Autoren ein natürlicher Bestandteil der Moral, und verweist nicht auf ein Defizit: „It is no objection to moral norms that, in some circumstances, they can be justifiably overridden by other moral norms with which they conflict. All general moral norms are justifiably overridden in some circumstances“ (Beauchamp und Childress 2009: 15). Die Autoren übernehmen von W.D. Ross die Unterscheidung zwischen prima facie und wirklichen [actual] Pflichten: „A *prima facie* obligation is one that must be fulfilled unless it conflicts, on a particular occasion, with an equal or stronger obligation. This type of obligation is always binding *unless* a competing moral obligation outweighs it in a particular circumstance“ (ebd.: 15). Ob in einer konkreten Situation auch eine tatsächliche Pflicht vorliegt, müssen die moralisch Handelnden entscheiden: „Agents can determine their *actual* obligations in such situations by examining the respective weights of the competing prima facie obligations (the relative weights of all competing prima facie norms)“ (ebd.: 15).

Rauprich verweist darauf, dass „**moralische Probleme**“ im Principlism (Beauchamp und Childress'scher Art) systematisch als Konflikt von prima facie Pflichten im Netzwerk der vier Prinzipien und ihrer Spezifizierungen betrachtet werden (vgl. Rauprich 2012: 594f). Eine Analyse ethischer Probleme besteht somit in der Suche nach den konfligierenden Pflichten. Eine Möglichkeit hierzu besteht in weiterer Spezifizierung von Normen, um ihren Geltungsbereich besser zu unterscheiden. Im biomedizinischen

Kontext könnte etwa das Abschalten eines Beatmungsgeräts nicht als töten, sondern als sterben lassen interpretiert werden, wodurch das Verbot des Mordes spezifiziert wird, indem Mord vom Abschalten von Maschinen unterschieden wird. Dies erfordert eine sehr genaue konzeptionelle Analyse der involvierten Begriffe und der Auswirkungen auf die Betroffenen, das soziale Umfeld, und dergleichen. In Fällen, in denen eine weitere Spezifizierung den Konflikt nicht auflösen kann, muss eine **Abwägung** [balancing] erfolgen.

Prinzipien, Regeln, (professionelle) Verpflichtungen und Rechte müssen gewichtet und abgewogen werden [weighting and balancing] (vgl. Beauchamp und Childress 2009: 19). „Balancing is the process of finding reasons to support beliefs about which moral norms should prevail. Balancing is concerned with the relative weights and strengths of different moral norms, whereas specification is concerned primarily with their scope (i.e., range)” (ebd.: 20). Abwägung ist, anders als Spezifizierung, stärker auf einen konkreten Einzelfall bezogen und „[...] allows for a due consideration of all norms bearing on a complex, very particular circumstance” (ebd.: 21), während Spezifikation immer noch die Formulierung einer allgemeinen Regel betrifft.

Abwägungen können nicht nach Willkür oder bloßer Intuition erfolgen, sondern müssen **sechs Bedingungen** genügen, um die Übertretung bestimmter prima facie verpflichten zugunsten anderer prima facie Verpflichtungen zu rechtfertigen (vgl. ebd.: 23):

1. Gute Gründe müssen angegeben werden, weshalb auf Basis der übertretenden Norm statt der übertretenen Norm gehandelt wird.
2. Das moralische Ziel, das diese Normenübertretung rechtfertigt, muss eine realistische Aussicht auf Erreichung haben.
3. Es sind keine moralisch zu präferierenden alternativen Handlungsweisen möglich.
4. Der niedrigste Grad an Übertretung wurde gewählt, der für das zu erreichende Ziel nötig ist.
5. Jegliche negativen Effekte, die sich aus der Überschreitung ergeben, wurden minimiert.
6. Alle betroffenen Gruppen wurden unparteilich einbezogen.

Nach Rauprich (2012: 595) muss im Principlism die Abwägung auf Grund einer fehlenden Hierarchie der Prinzipien und Regeln jeweils auf Fall-zu-Fall-Basis mit jeweils situationsbezogener Begründung erfolgen (z.B. hinsichtlich der Frage, warum in einem bestimmten Fall Aspekte des Wohltuns und der Autonomie wichtiger sind als der Nichtschädigung, was in einem strittigen biomedizinischen Fall etwa eine riskante Organtransplantation erlauben könne). Während Spezifizierung vor allem für die Entwicklung von Handlungsvorschriften [policies] wichtig ist, ist Abwägung für die Behandlung von individuellen (Praxis-)Fällen relevant. Damit wird nicht der Anspruch erhoben, sämtliche Normenkonflikte auflösen zu können – manche Konflikte liegen sozusagen in der Natur der Sache, und sind kein

Zeichen von fehlendem Wissen oder einem ethischen Defizit – sehr wohl aber diese Konflikte ernst zu nehmen und auch nicht zu früh als unlösbar zu bezeichnen.

Der Vier-Prinzipien-Ansatz von Beauchamp und Childress wurde vorgestellt, um einen besonders umfassenden Principlism-Zugang vorzustellen, welcher ein regelgeleitetes Vorgehen (Spezifikation) beschreibt, um von sehr abstrakten, in der gemeinsamen Moral, fundierten Prinzipien mittlerer Reichweite zu stärker situativ bezogenen, weniger abstrakten Prinzipien bzw. Regeln zu gelangen. Wie sich zeigen wird, sind nicht alle Principlism-Ansätze methodologisch derart ausgearbeitet mit dem Ziel, Handlungsanleitung für konkrete Situationen zu geben, sondern beziehen sich eher allgemein auf die Hauptidee, ein Problem vor dem Hintergrund mehrerer Prinzipien mittlerer Reichweite zu beleuchten. Darunter fällt nach Ansicht des Verfassers der vorliegenden Arbeit auch die Ethical Matrix (EM), die demnach als eine Art methodologisch reduzierte Fassung des Principlism betrachtet werden kann, und die von Reijers et al (2018) als paradigmatisches Beispiel für einen Principlism-Ansatz im Bereich der Analyse von bereits bestehenden Techniken bezeichnet wird.

5.2.3 Ethical Matrix

Der Ethical Matrix (im Weiteren abgekürzt als EM) Ansatz wurde von einer Arbeitsgruppe rund um Ben Mepham erarbeitet (vgl. Mepham et al 2006). Aufgrund ihrer vielseitigen Einsetzbarkeit und der engen Verwobenheit mit dem vorhin ausführlich dargestellten Ansatz von Beauchamp und Childress soll das Grundprinzip der EM nun skizziert werden, da in Abschnitt 6.7 zu „Hauptfrage 3 (Folgen und Nebenfolgen der planmäßigen Anwendung) – Schritt 2 bis 4“ auch eine EM zum Einsatz kommt (vgl. S. 132).

Erstmals vorgestellt wurde das Konzept der EM von Ben Mepham, als ein „[...] framework for rational ethical analysis“ (Mepham 2000: 165)²³, welches die von Beauchamp und Childress für den medizinischen Kontext ausgearbeiteten Prinzipien auf den landwirtschaftlichen Kontext umlegt. Die Anwendung von Prinzipien mittlerer Reichweite auf einen bestimmten Kontext erfordert “[...] that they be translated into terms appropriate to a wide range of different ‚interest groups‘ [in that area]” (Mepham 2000: 168). Dies bedeutet, dass die Implikationen für jedes angenommene Prinzip jeweils vor dem Hintergrund der betroffenen Stakeholder-Gruppe zu beleuchten sind.²⁴ Anwendungspraktische Hinweise bietet das „Ethical Matrix Manual“ (Mepham et al 2006: 5). Es charakterisiert die EM als ein Werkzeug für die Unterstützung zur Beurteilung oder Entscheidungsfindung hinsichtlich Fragen ethischer Akzeptabilität oder der bestmöglichen Regulierung von bestehenden oder zukünftigen Techniken im Bereich der Nahrungserzeugung und der Landwirtschaft. Mittlerweile wird das Konzept der EM

²³ Dabei handelt es sich nicht um den ersten Artikel von Mepham, in welchem die EM zur Anwendung kommt – allerdings um den ersten, der sich thematisch ausdrücklich auf die Methodologie fokussiert.

²⁴ Nach Ansicht des Verfassers der vorliegenden Arbeit handelt es sich dabei somit um einen Prozess der Spezifizierung, ähnlich wie er schon bei Beauchamp und Childress dargelegt ist, allerdings in weniger formalisierter Gestalt.

aber auch in anderen Anwendungsbereichen eingesetzt, z.B. als Werkzeug um ethische Einstellungen hinsichtlich einer unterirdischen CO₂-Speicheranlage zu verorten (vgl. Gough und Boucher 2013), für die Entscheidungsfindung im Bereich Abfallbehandlung von radioaktiven Abfällen (vgl. Cotton 2009), sowie zur Adressierung ethischer Probleme im Bereich von Bildungsforschung (Tangen 2014).

In der EM-bezogenen Arbeit lassen sich grob zwei Phasen unterscheiden: die ethische **Analyse** [ethical analysis] und die ethische **Bewertung** [evaluation]. Während die Analyse als Erstellung und Aufbereitung der EM betrachtet werden kann, geht es in der Bewertung darum, die in der EM festgehaltenen Tatsachen [facts] und Werturteile [values] zueinander in Beziehung zu setzen.

Das Grundprinzip hinter der Erstellung einer EM besteht in der Verschränkung eines Sets an prima facie gültigen Prinzipien mittlerer Reichweite, die die Spalten der Matrix bilden, mit unterschiedlichen Stakeholdergruppen, die von einer fraglichen Technik, einer Maßnahme oder einer Problematik betroffen sind, und die die Zeilen der Matrix bilden (vgl. Mepham et al 2006: 8-11).

Die „Standardprinzipien“, die aber je nach Sachlage adaptiert oder abgeändert werden müssen, sind die Prinzipien des

“Respect for:

- Wellbeing is considered to be the best way of representing utilitarian concerns (,maximising the good‘)²⁵
- Autonomy represents deontological concerns (treating others as ,ends in themselves‘)
- Fairness represents respect for justice, an interpretation advocated by the philosopher Rawls“ (Mepham et al 2006: 9).

Für diese Prinzipien, die über diese knappen Angaben im “Manual” nicht weiter erläutert werden, berufen sich die AutorInnen ausdrücklich auf die „[...] common morality, i.e. the ethical code shared by most members of a society in the form of unreflective common sense and tradition [...]“ (Mepham et al 2006: 7), die als kleinster gemeinsamer moralische Nenner einem Reflexionsprozess unterzogen werden muss, um zu informiertem, bewusstem Dialog zu gelangen.

Die Auswahl der Stakeholder-Gruppen hängt zwar vom jeweiligen zu behandelnden Problem ab, in jedem Fall bedeutet der Einbezug aber auch, dass den Gruppen ein „moralischer Wert“ zugeschrieben wird: „A critical factor is that ,ethical standing‘ is claimed for all interest groups specified. In other

²⁵ Dieses stellt eine Zusammenlegung der Prinzipien des Nichtschadens und des Wohltuns von Beauchamp und Childress dar (vgl. Mepham 2000: 168), einerseits um die Matrix zu vereinfachen, andererseits weil die menschliche Verantwortung gegenüber Organismen im Bereich Landwirtschaft diese beiden Prinzipien untrennbar zusammenführe.

words, they are subjects of ethical consideration in their own right, and not just means to others' ends. Furthermore, the way they are affected by a novel technology should differ systematically from each other" (Mepham et al 2006: 9). In praktischer Hinsicht hat sich im Bereich von Nahrungsmittel- und Landwirtschafts-bezogenen Themen gezeigt, dass die Anzahl der zu berücksichtigen Gruppen meist auf vier beschränkt werden kann (etwa ProduzentInnen, KonsumentInnen, betroffene Nutztiere, Gesamtheit der Lebewesen [„Biota“]).²⁶

Eine EM enthält in den Zellen bzw. Feldern, die sich durch das Auskreuzen von Prinzipien und betroffener Stakeholdergruppe ergeben, entweder die Spezifizierung des jeweiligen Prinzips vor dem Hintergrund der betroffenen Stakeholdergruppe (vgl. exemplarisch Mepham 2000: 170) oder aber eine Auflistung der Implikationen bzw. Konsequenzen für die jeweilige Gruppe, welche im Rahmen gemeinsamer ethischer Abwägungen (etwa in einem Gruppensetting) auch mit Zahlenwerten [scores] oder positiven/negativen Vorzeichen, je nach Art der Betroffenheit, versehen werden können (vgl. exemplarisch Mepham et al 2006: 42).

Entsprechend obiger Ausführung lässt sich nach Ansicht des Verfassers der vorliegenden Arbeit der generische Aufbau einer EM folgendermaßen darstellen, entweder als EM von Spezifizierungen

Stakeholdergruppe	Prinzip 1	Prinzip 2	...
Stakeholdergruppe 1	Spezifizierung von Prinzip 1 für Stakeholdergruppe 1	Spezifizierung von Prinzip 2 für Stakeholdergruppe 1	...
Stakeholder Gruppe 2	Spezifizierung von Prinzip 1 für Stakeholdergruppe 2	Spezifizierung von Prinzip 2 für Stakeholdergruppe 2	...
...

Tabelle 1: EM Beispieltabelle 1

oder aber als EM für einen Überblick von Betroffenheitsdimensionen und Auswirkungen:

Stakeholdergruppe	Prinzip 1	Prinzip 2	...
Stakeholdergruppe 1	+ ethisch wünschenswerte Auswirkung(en) ersichtlich aus Prinzip 1 hinsichtlich Stakeholdergruppe 1 - ethisch problematische Auswirkung(en) ersichtlich aus Prinzip 1 hinsichtlich Stakeholdergruppe 1	+ ethisch wünschenswerte Auswirkung(en) ersichtlich aus Prinzip 2 hinsichtlich Stakeholdergruppe 1 - ethisch problematische Auswirkung(en) ersichtlich aus Prinzip 2 hinsichtlich Stakeholdergruppe 1	...
Stakeholder Gruppe 2	+ ethisch wünschenswerte Auswirkung(en) ersichtlich aus Prinzip 1 hinsichtlich Stakeholdergruppe 2 - ethisch problematische Auswirkung(en) ersichtlich aus Prinzip 1 hinsichtlich Stakeholdergruppe 2	+ ethisch wünschenswerte Auswirkung(en) ersichtlich aus Prinzip 2 hinsichtlich Stakeholdergruppe 2 - ethisch problematische Auswirkung(en) ersichtlich aus Prinzip 2 hinsichtlich Stakeholdergruppe 2	...
...

Tabelle 2: EM Beispieltabelle 2

²⁶ Wie weiter unten in der vorliegenden Arbeit noch ersichtlich wird, müssen für die Human+ bezogenen Analysen deutlich mehr als vier Stakeholdergruppen berücksichtigt werden.

In der praktischen Anwendung können zu Vergleichszwecken auch zwei EM konstruiert werden: eine die den Ist-Zustand beschreibt, und eine zweite die die Situation nach (potentieller) Einführung einer bestimmten (Bio-)Technologie oder Ausführung einer Maßnahme skizziert (vgl. Mepham et al 2006: 12f).²⁷ Die Veränderungen von Situation 1 zu Situation 2 können durch „semi-quantitative“ Werte [scores] verdeutlicht werden, je nachdem ob das Prinzip gegenüber der betroffenen Gruppe respektiert oder verletzt wird, z.B. verlaufend von -2 für starke Verletzung, -1 für schwache Verletzung, 0 für „irrelevante Auswirkungen“, +1 für schwache Beachtung und +2 für starke Beachtung. Das Prinzip verdeutlichen Mepham et al mit dem Beispiel, dass der Einsatz genmodifizierter herbizidresistenter Pflanzen sich zwar positiv auf den Ertrag von Landwirten auswirken kann (+1), aber durch den dann höheren Spritzmitteleinsatz die allgemeine Artenvielfalt leidet (-1). Die Werte dienen allein der Orientierung im ethischen Abwägungsprozess, und dürfen nach Ansicht der AutorInnen der EM nicht gegeneinander aufsummiert werden, weil die Werte in unterschiedlichen Feldern nicht identes Gewicht zueinander haben, etwa weil in manchen Fällen dafür argumentiert werden kann, dass ein Prinzip das andere überwiegt (was ja bereits im Konzept prima facie Prinzip angelegt ist).

Das Beispiel macht außerdem deutlich, dass sowohl Landwirten wie auch der Biota ein „ethical standing“ zuerkannt wurde, und dass die solcherart vorgenommene Bewertung der Auswirkungen hier ein utilitaristisches Prinzip (z.B. Wohlergehen oder Nichtschädigung) erforderlich macht. Es hebt außerdem besonders hervor, dass empirisches Wissen, etwa in Form der Ergebnisse von Forschungsarbeiten oder auch von „Erfahrungswissen“ von Stakeholdergruppen, in die Konstruktion einer EM einfließen muss. Eine EM erweist sich so in doppelter Hinsicht als Instrument der Angewandten Ethik, einmal um in Problemlagen mit angewandtem Bezug zu assistieren, und in zweiter Hinsicht weil sie selbst empirisch fundierte Inputs („Fakten“) benötigt. Vorausgeschickt werden kann an dieser Stelle bereits, dass für die Human+ bezogenen Analysen weiter unten, welche ebenfalls durch eine EM unterstützt werden wird, sowohl die Auswahl der zu berücksichtigenden Stakeholdergruppen wie auch der Einbezug von empirischem Wissen (etwa um Risiken auf Betroffene und Sichtweisen der BOS-MitarbeiterInnen zu berücksichtigen) zentralen Stellenwert einnehmen werden.

Der Inhalt der Felder kann also zum einen aus „**Fakten**“ [facts] bestehen (vgl. Mepham et al 2006: 12f). Während es auf den ersten Blick mitunter nicht weiter schwierig scheint, sich auf Forschungsarbeiten als gesichertes (oder zumindest intersubjektiv abgesichertes) Wissen zu beziehen, verweisen die AutorInnen auf die Probleme, die sich bei der Auswahl und Bewertung der Daten ergeben, etwa wenn

²⁷ Aus diesem Grund betrachten die AutorInnen die EM auch nicht als Werkzeug, um die ethische Zulässigkeit einer Technik oder Maßnahme „an sich“, sondern immer nur in Bezug zu den anderen, damit verglichenen Alternativen, zu bewerten. Damit setzen sie den Geltungsansprüchen ähnliche Grenzen wie sie weiter oben in Bezug auf TA erläutert wurde, wonach TA nie vollständig das sämtliche Spektrum von Konsequenzen auslösten kann, sondern jeweils reflektierte und begründete Schwerpunktsetzungen notwendig sind.

die Neutralität derjenigen, die dieses Wissen bereitstellen (z.B. ForscherInnen) angezweifelt werden kann. Bestehende Kontroversen, Zweifel an der Zuverlässigkeit oder Glaubwürdigkeit von Quellen sollten nicht nur in die ethische Abwägung einfließen, sondern können auch in den Zellen der EM protokolliert werden. Die andere Kategorie, mit denen die Zellen gefüllt werden können, sind, zum anderen, „**Werturteile**“ [values], die sich auf moralische Bewertungen der quantifizierbaren Fakten beziehen. So kann je nach eigener Bewertungsposition dasselbe quantitative Ergebnis anders beurteilt werden (etwa wenn man unterschiedlichen ökonomischen Theorien anhängt) oder den Konsequenzen ein anderes Gewicht beigemessen werden.

Beide exemplarische, abstrakte Darstellungen machen zudem deutlich: Eine EM dient als analytisches Werkzeug, das die Evaluation oder Abwägung unterstützt, sie aber nicht ersetzt oder automatisiert durchführt (vgl. Mepham 2000: 169). Während für die Analyse die AutorInnen recht ausführlich auf die Methode bei der Erstellung und Füllung der EM eingehen, wird dem Thema der ethischen Abwägung vergleichsweise nur knapp Raum gegeben und auf folgende Kurzformel gebracht: „The next step in the process, ethical evaluation, involves subjectively weighting the different impacts, which allows you to reach an ethical judgement on the acceptability of the technology in question“ (Mepham et al 2006: 14). Unterschiedliche Personen können dieselben Prinzipien bzw. Tatsachen unterschiedlich bewerten und gewichten. Aus diesem Grund ist der Verfasser der vorliegenden Arbeit der Ansicht, dass implizit der Schritt der ethischen Evaluation auch darin besteht, möglichst transparent den Abwägungsprozess zu verschriftlichen und in argumentativer Form nachvollziehbar zu machen, bzw. im Falle von Dissens von TeilnehmerInnen (bei kollektiver Bearbeitung der Matrix) die Gründe für unterschiedliche Beurteilungen festzuhalten. Dissens abseits davon, etwa wenn LeserInnen der Ergebnisse der ethischen Evaluierung zu anderen Schlüssen gelangen, könnte der Ausgangspunkt für eine gemeinsame weitere Deliberationsphase sein. In jedem Fall kann die EM als Hilfsmittel zur Strukturierung des für ihre Erstellung herangezogenen Ausgangsmaterial sehr hilfreich sein, vorausgesetzt dass entsprechende Sorgfalt in ihre Erstellung investiert wird, denn: „What you get out of the ethical matrix is totally dependent of what you put in“ (Mepham et al 2006: 14).

Mepham et al (2006) bieten nicht nur eine theoretische Erläuterung zur EM, sondern auch Überlegungen dazu, ob Prinzipien vorgegeben werden (im Sinne eines Top-Down-Vorgehens) oder ein EM-ähnlicher Aufbau zur Generierung von Prinzipien verwendet werden sollen (Bottom-Up-Vorgehen), weiters auch zu Ablaufschemata, zu ihrem Einsatz im Rahmen von Stakeholder-Workshops und welche Aspekte dabei beachtet werden müssen, etwa solche datenschutzrechtlicher Natur (vgl. dazu Mepham et al 2006: 15ff).

Vergleicht man die Ausführungen zur Ethical Matrix mit den Ausführungen zum Principlism bei Beauchamp und Childress (2009), dann fällt auf, dass im Rahmen der EM Spezifizierung (als Formulierung

konkreterer Normen auf Basis allgemeinerer Normen) eine untergeordnete Rolle spielt. Aus Sicht des Verfassers der vorliegenden Arbeit liegt das an der unterschiedlichen Zielsetzung: Während Beauchamp und Childress abstrakte Prinzipien der gemeinsamen Moral zu handlungsanleitenden, weniger allgemeinen Normen spezifizieren (z.B. zu informed consent), die praktizierendem medizinischen (Forschungs-)Personal konkrete moralische Orientierung in ethisch konfliktbehafteten Situationen geben sollen, scheint der Ethical Matrix Ansatz (und die hierfür aufgestellten methodologischen Empfehlungen) viel stärker als „reflexionsgenerierendes“ Suchwerkzeug angelegt zu sein – auch wenn eine der EM Beispieltabellen die Möglichkeit aufzeigte, in den Feldern auch die Spezifizierung von Prinzipien vor dem Hintergrund von betroffenen Stakeholdergruppen einzutragen. Diese letztere Art der Spezifizierung ist allerdings deutlich weniger formalisiert ausgearbeitet als wie im Ansatz von Beauchamp und Childress.

In den nachfolgenden Abschnitten dieses Kapitels wird auf andere Frameworks eingegangen, die in technikethischen Betrachtungen, v.a. bezogen auf IKTs, angewandt werden.

5.3 ETICA

ETICA (**ethical issues of emerging ICT applications**) ist der Name eines im 7. Forschungsrahmenprogramm der EU von April 2009 bis Mai 2011 abgewickelten Projekts, welches sich der Identifikation und Analyse von vom damaligen Standpunkt aus neu entwickelnden Technologien widmete und Policy-Empfehlungen für (forschungs-)politische Entscheidungsträger erarbeitete (vgl. Stahl 2011: 140f). Es versteht sich als “technology foresight project”, welches somit eine Reihe von möglichen „Zukünften“²⁸ erarbeitet, um Wege zu den als besonders wünschenswert erachteten aufzuzeigen. Technologie im ETICA Sinne „[...] is a high level system that effects the way humans interact with the world. This means that one technology in most cases can comprise numerous artefacts and be applied in many different situations”²⁹ (Stahl 2011: 142). Basierend auf dieser Definition wurden in einem ersten Projektschritt neu aufkommende Technologien [emerging technologies] identifiziert, indem eine Diskursanalyse von Dokumenten (publiziert z.B. von mit Forschungsförderung betrauten Stellen) angestellt wurde, und die Ergebnisse anschließend mit projektexternen Experten besprochen wurden, um insgesamt 11 neu aufkommende Technologien³⁰ mit Beispielanwendungen zu identifizieren. Die Details dieses Prozesses sind für die vorliegende Arbeit nicht weiter relevant³¹, das dabei verwendete

²⁸ Während etwa im Duden kein Plural von Zukunft vorgesehen ist, scheint sich im Bereich der TA ein Plural als Terminus Technicus etabliert zu haben, etwa bei Grunwalds Konzept der „Technikzukünfte“ (vgl. Grunwald 2012).

²⁹ An diesen drei Sphären „technology“, „application“, „artifact“ wird sich Brey (2012) orientieren und ein hierarchisches Ebenensystem entwickeln.

³⁰ Affective computing, ambient intelligence, artificial intelligence, bioelectronics, cloud computing, future internet, human-machine symbiosis, neuroelectronics, quantum computing, robotics, virtual/augmented reality (vgl. Stahl 2011: 144).

³¹ Für Details vgl. Stahl 2011: 142f.

analytische Raster bietet aber durchaus einen heuristischen Wert für die Beschreibung von Techniken, weshalb es hier wiedergegeben wird (vgl. Abbildung 1). Diese Zusammenstellung von Informationen wird auch „**technology description**“ genannt (vgl. ETICA D2.2: 12), und enthält jeweils fünf Anwendungsbeispiele der Technologien und ihre charakteristischen Eigenschaften.

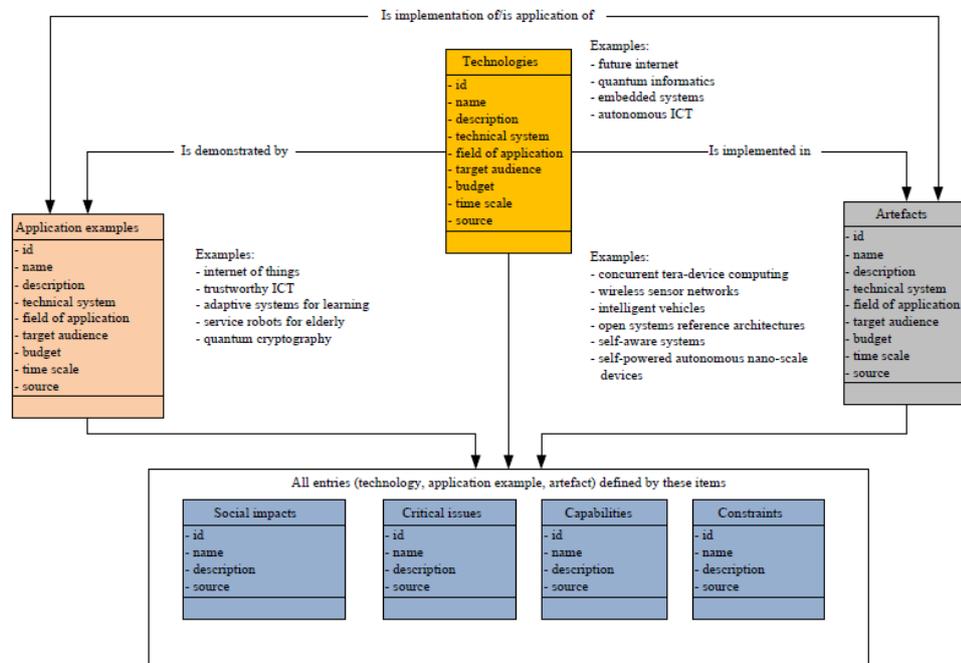


Abbildung 1: ETICA: „Analytical Grid for Discourse Analysis“ (Quelle: Stahl 2011: 143)

Die ethische Analyse zu den identifizierten, emergierenden Technologien, ist im ETICA Projekt der deskriptiven Ethik zuzurechnen, da sie auf eine quantitative Analyse der ethischen Fachliteratur hinausläuft³² (vgl. ETICA D.2.2: 9). Die ethische Analyse ist eingebettet in **vier Arbeitsschritte**. So wurde zunächst, **Schritt 1**, ein Datenset der relevanten akademischen Literatur aus dem Bereich „computer and information ethics“ zusammengestellt (wobei 1.038 Verweise, Abstracts und Schlagwörter von Artikeln aus Fachzeitschriften und Konferenzsammelbänden zusammengestellt wurden). Dieses Datenset wurde, **Schritt 2**, durch eine spezielle Software (VOSviewer) bibliometrisch analysiert, um semantische Distanzen und Näheverhältnisse zwischen Schlagwörtern aus dem Datenset zu ermitteln. Beispielsweise ergab dieses Vorgehen, dass die Begriffe „ambient intelligence“, „privacy“ und „autonomy“ häufig in Abstracts parallel vorkamen, was durch VOSviewer auch räumlich durch eine visuelle Landkarte abgebildet wurde. **Drittens** wurden die oben schon erwähnten „technology descriptions“ durch diesen bibliometrisch erzeugten Überblick „analysiert“. **Viertens** wurden durch Fokusgruppen mit Laien und durch die Befragung von Fachexperten die erzeugte „Analyse“ aus Schritt drei validiert.

³² „It is important to say that our method is descriptive in nature, because we are merely using existing literature on ethics of emerging ICTs to identify ethical issues. We are not applying a particular ethical framework ourselves, but are describing what other scholars have said on ethics of emerging ICTs“ (ETICA D2.2: 13).

Die ethische Analyse „im engeren Sinne“ (im oberen Absatz Schritt 3 genannt) beinhaltet folgende Ablaufschritte (vgl. ETICA D2.2: 12f):

- „Ethische Analyse“ der **charakteristischen Eigenschaften** der Technologien, wie sie sich aus den technology descriptions entnehmen lassen (z.B. waren im Falle der emergierenden Technologie „ambient intelligence“ solche Charakteristika u.a. „embeddedness and invisibility of the sensors and computational devices“). Das Vorgehen wird nicht genauer beschrieben, sondern umrissen als „All the defining features of each emerging ICT and their consequences are analyzed from an ethical point of view“ (ETICA D2.2: 13).
- „Ethische Analyse“ der Anwendungsbeispiele aus den technology descriptions. Auch diese Analyse bezieht sich auf die zusammengestellte Literatur: „When analyzing a technology, say, robotics, we did a search in the overview for abstracts on robotics [...]. If there were abstracts on this topic, we collected the accompanying article and extracted the relevant ethical issues from the article and incorporated it in our ethical analysis“ (ETICA D.2.2: 13).
- Verwendung der erzeugten visuellen Landkarte, um zu prüfen, ob in der „ethischen Analyse“ aus obigen zwei Punkten die wichtigsten ethischen Werte, Probleme und Erwägungen einbezogen wurden.
- Abschließende Diskussion durch Feedback von professionellen, Projekt-assozierten EthikerInnen.

Während Brey den ETICA Ansatz als „[...] possibly the most elaborate and promising ethical approach to emerging technologies that has been developed to date“ (Brey 2012: 309) bezeichnet, kritisiert er, dass der Ansatz sich als “future studies” proklamiert, aber hauptsächlich Texte als Referenz heranzieht, die nicht aus der Zukunftsforschung stammen, und dass zweitens viele der ethischen Analysen aus dem Projekt auf generische (charakteristische) Eigenschaften der Technologien abzielen, und somit nur eingeschränkt auf mögliche Anwendungen und darauf basierende (physische) Geräte bzw. Artefakte eingegangen wird.

Dem ist noch hinzuzufügen, dass ein Ansatz wie der in ETICA verfolgte erstens an den Zielen der vorliegenden Arbeit vorbeigeht (da es sich hierbei nicht um eine Arbeit im Rahmen deskriptiver Ethik handeln soll), und zweitens auch über die Möglichkeiten der vorliegenden Arbeit hinausgeht, weil ein ganzes Forschungsteam erforderlich ist, welches sich arbeitsteilig mit heterogenen Technologien beschäftigt, zudem spezialisierte Software anwendet, und über Fokusgruppen und assoziierte ExpertInnen projektexterne Gesichtspunkte hereinholt.

5.4 eTA – ETHICAL TECHNOLOGY ASSESSMENT

Als Ergänzung zu bestehenden Ansätzen der TA stellen Palm und Hansson ihren Checklist-basierten Ansatz vor, den sie als Ethical Technology Assessment (eTA) bezeichnen, und welcher in einem Frühstadium der technischen Entwicklung auf negative ethische Implikationen verweisen soll (vgl. Palm und Hansson 2006: 543f). Zentraler Bestandteil ist kontinuierlicher Austausch der mit dem ethischen Assessment befassten Personen mit den EntwicklerInnen der fraglichen Technologie und regelmäßige Assessments (vgl. Palm und Hansson 2006: 549ff). Dabei soll der gesamte Lebenszyklus einer technischen Entwicklung einbezogen werden, beginnend mit der Forschungs- und Entwicklungsphase, über die Patentierungs- und Marketingphasen, Tests und hin zur Überprüfung der Auswirkungen auf Individuen und Gesellschaft. Besonders heben sie hervor, dass in eTA unterschiedliche Meinungen und moralische Konflikte nicht aufgelöst, sondern im Gegenteil hervorgehoben werden sollen, da kein geteiltes moralisches Framework (im Projekt) vorausgesetzt werden kann, auf welchem Abwägungen und Entscheidungen basieren können, sondern (im Anschluss an Grunwald) verweisen sie darauf, dass dieses moralische Framework erst gemeinsam mit der technischen Entwicklung erstellt werden kann. Es sei die Aufgabe der Ethik, diese moralische Entwicklung zu unterstützen. Weitere Kennzeichen des Ansatzes sind auch hier der Anspruch, alle relevanten StakeholderInnen einzubeziehen, um nicht nur ein breites Spektrum der beteiligten Akteure, sondern auch der beteiligten Verantwortlichkeiten zu erfassen, und die Ergebnisse der eTA entsprechend weit auch zu kommunizieren. Der Ansatz erhebt den Anspruch der „Theorieunabhängigkeit“, d.h. sie berufen sich auf keine bestimmte ethische Theorie wie Utilitarismus oder deontologische Ansätze, sondern wollen (verschiedene) Konzepte aus der Moralphilosophie verwenden, um Probleme präziser zu charakterisieren und alternative Möglichkeiten und Policies zu entwickeln. Diese Offenheit soll auch größtmögliche Freiheiten für den Dialog mit den EntwicklerInnen und die Diskussion mit StakeholderInnen eröffnen. Für den Kontext der vorliegenden Arbeit wird daher diesem Appell gefolgt, ein Bewertungsframework zu entwerfen, wobei sehr wohl der Versuch unternommen wird, einen gewissen Grad an Generizität vorzusehen, damit das Framework auch auf andere Sicherheitsforschungsprojekte (und nicht nur auf Human+) anwendbar ist. Gleichzeitig lässt sich implizit eTA entnehmen, dass zwecks Dialogmöglichkeit mit TechnikentwicklerInnen und StakeholderInnen möglichst hohe Verständlichkeit auch für Nicht-EthikerInnen angestrebt werden sollte – die Zerlegung in Ablaufschritte im zu entwickelnden Framework wird ein Versuch sein, dies ein Stück weit einzulösen.

Im Herzstück von eTA steht eine „Checkliste“, die dabei helfen soll, ethische Probleme frühzeitig im Entwicklungsprozess einer Technik zu identifizieren (im Sinne eines Frühwarnsystems), welche folgende neun „kritische Aspekte“ umfasst (vgl. Palm und Hansson 2006: 551-555):

- “Dissemination and use of information”

- “Control, influence and power”
- “Impact on social contact patterns”
- “Privacy”
- „Sustainability“ (im Sinne von Nachhaltigkeit in Bezug auf nachkommende Generationen hinsichtlich des Einflusses auf ökonomische, soziale und ökologische Entwicklungen)
- “Human reproduction”
- “Gender, minorities and justice”
- “International relations”
- “Impact on human values”

Die AutorInnen definieren diese sehr breit gefassten Aspekte nicht näher, sondern geben Beispiele aus der Technikentwicklung, wie diese Aspekte vor dem Hintergrund einer Technik relevant sind. So verweisen sie etwa im Bereich „Impact on human values“ darauf, dass die zunehmende Verfügbarkeit von persönlichen Informationen unseren Respekt vor der Privatheit verändern kann (vgl. Palm und Hansson 2006: 555). Die genannten Aspekte sollen (offenbar) Gegenstand eines kontinuierlichen Dialogs mit TechnikentwicklerInnen und Stakeholdern sein, und damit frühzeitig Probleme erkennen helfen. Während die gewählten Kriterien sehr breit sind und sich auf unterschiedlichste Bereiche von Technologien (wenngleich in ungleichem Ausmaß) beziehen lassen, ist der Kritik von Brey rechtzugeben, dass der Ansatz methodisch recht unterbestimmt ist, und etwa keine Hilfestellungen für die Strukturierung des Dialoges zwischen EthikerInnen, EntwicklerInnen und StakeholderInnen, oder die Durchführung der ethischen Analyse basierend auf den genannten Kriterien, gibt (vgl. Brey 2012: 308). Weitere Kritikpunkte von Brey umfassen die begrenzte Checkliste und den offenbaren Fokus auf die nähere Zukunft. Ergänzt werden könnte an dieser Stelle, dass auch für die Human+ Analyse eTA nicht unmittelbar angewandt werden kann (was vielleicht, angesichts der Betonung der AutorInnen, dass ein Bewertungsframework auf Projekte zugeschnitten werden muss, nicht verwunderlich ist): erstens erscheint aus der Liste der neun kritischen Aspekte aus der eTA Checkliste nur ein Teil für die Human+ Analyse relevant, v.a. solche der Informationsverwendung (da sie den Anwendungszweck von Human+ in den Blickpunkt rückt), der Privatheit wie auch der Sensibilisierung gegenüber der Betroffenheit von Minderheiten bzw. vulnerablen Gruppen. Und zweitens werden diese Schlagworte bei eTA nicht im engeren Sinne definiert, sodass sie wohl eher als Anregungen für zu bedenkende Problemlagen betrachtet werden sollten. Nicht zufällig werden einige Aspekte, vor allem Privatheit, auch bei anderen der im Weiteren dargestellten Ansätze vorkommen.

5.5 ETHICAL IMPACT ASSESSMENT

Für den europäischen Raum (aber mit Übertragbarkeitsmöglichkeiten über den europäischen Kontext hinaus) entwickelte David Wright das Framework „Ethical Impact Assessment“, welches sich ausdrücklich auf die vier Prinzipien von Beauchamp und Childress bezieht, und diese mit weiteren Prinzipien bzw. Bewertungsdimensionen hinsichtlich Privatheit [privacy] und Datenschutz [data protection] anreichert (vgl. Wright 2011: 199). Die proklamierte Zielgruppe des Frameworks sind Personen, die sich mit neuen Technologien oder Diensten, Projekten oder diesbezüglichen Strategien befassen, und sicherstellen wollen, dass ethische Aspekte ausreichend (von beteiligten StakeholderInnen) berücksichtigt werden.

Wright argumentiert, dass ethische Zulässigkeit immer abhängig von Kontextfaktoren ist, weshalb (strikte) präskriptive ethische Vorschriften problematisch seien (vgl. Wright 2011: 200). Stattdessen plädiert er für die Verwendung eines Fragenkatalogs, ähnlich wie Forschungsprojekte für die Beantragung von Fördermitteln durch Forschungsrahmenprogramme der Europäischen Kommission ethische Fragen beantworten müssen, etwa ob die Forschung das Tracking oder die Beobachtung von Personen inkludiere. Besondere Bedeutung misst Wright der Unterscheidung zwischen **Werten** [values] und **Strategien** bzw. **Regulationen** [policies] zu, die sich beide gewissen Prinzipien (mittlerer Reichweite) zuweisen lassen (vgl. Wright 2011: 201): während etwa im Rahmen seines Frameworks Würde [dignity] als Wert dem Prinzip des Respekts gegenüber der Autonomie zugewiesen wird, wäre das Konzept der informierten Zustimmung [informed consent] als Strategie (zum Erreichen dieser) zu werten. Es sei jeweils eine Frage des diskursiven Austausches zwischen Stakeholdern, zu entscheiden, ob in bestimmten Situationen die Würde respektiert wird oder ob tatsächlich informierte Zustimmung vorliegt.

Zur Diskussion dieser Aspekte arbeitet Wright umfangreiche Fragenlisten aus, die sich an Werten orientieren, wie sie sich nicht nur bei Beauchamp und Childress finden lassen, sondern auch in unterschiedlichen grundlegenden Gesetzeswerken wie dem EU-Reform-Vertrag von Lissabon aus 2007 angeführt werden (z.B. Menschenwürde, Freiheit, Demokratie, Schutz der Menschenrechte, Toleranz, Solidarität, vgl. Wright 2011: 202). Für die Bewertung bzw. das „Assessment“ selbst beruft sich Wright auf unterschiedliche Praktiken, wie Technikfolgenabschätzung (TA), privacy impact assessment, risk assessment, und weitere, als deren gemeinsamer Bestandteil er jeweils den umfangreichen Einbezug von Stakeholdern identifiziert.

Die von Wright ausgearbeiteten Fragenlisten sind zu ausführlich, um sie hier in voller Länge wiederzugeben. Wie angeführt gruppiert Wright unter unterschiedliche Prinzipien eine Reihe von Werten bzw. zu reflektierenden Problemlagen, jeweils mit einem kurzen Erklärungstext und Hinweisen, woher diese

entnommen sind bzw. aus welchen anderen Regularien diese abgeleitet werden können, sowie eine Reihe von (Diskussions-)Fragen um deren Bedeutung in einem bestimmten Kontext und deren Beachtung gemeinsam mit Stakeholdern zu klären (vgl. Wright 2011: 204). Nachfolgende Tabelle zeigt die Struktur der von Wright verwendeten Prinzipien, und der unter diese Prinzipien subsumierten Werte, Problemlagen bzw. Strategien (vgl. Wright 2011: 204-215). Da Wright auch für die übergeordneten Prinzipien selbst Diskussionsfragen auflistet, werden sie hier auch in der „Wert“-Spalte noch einmal ergänzend angeführt. Da pro Wert/Strategie/Problemlage zwischen 3 bis 13 Diskussionsfragen formuliert werden, wird jeweils nur eine Frage beispielhaft angeführt. Um die nachfolgende Übersicht kurz zu halten, wird zudem auf eine Wiedergabe der zum Teil recht detaillierten Definitionen für die jeweiligen Prinzipien und Werte/Strategien/Problemlagen verzichtet, da der Bedeutungsumfang durch die Beispielfragen ausreichend illustriert wird.

Prinzip	Wert/Strategie/Problem-lage	Beispielfrage
Respect for autonomy (right to liberty)	Respect for autonomy	Does the technology or project curtail a person's right to liberty and security in any way? If so, what measures could be taken to avoid such curtailment? (Wright 2011: 204)
	Dignity	Does the project require citizens to use a technology that marks them in some way as cognitively or physically disabled? If so, can the technology be designed in a way so that it does not make them stand out in a crowd? (ebd.: 205)
	Informed Consent	Will the project obtain the free and informed consent of those persons to be involved in or affected by the project? If not, why not? (ebd.: 206)
Nonmaleficence (avoiding harm)	Nonmaleficence	(keine Beispielfragen angeführt)
	Safety	Is there any risk that the technology or project may cause any physical or psychological harm to consumers? If so, what measures can be adopted to avoid or mitigate the risk? (ebd.: 206)
	Social solidarity, inclusion and exclusion	Has the project taken any steps to reach out to the e-excluded (i.e., those excluded from use of the Internet)? If not, what steps (if any) could be taken? (ebd.: 207)
	Isolation and substitution of human contact	Will the project use a technology which could replace or substitute for human contact? What will be the impact on those affected? (ebd.: 207)
	Discrimination and social sorting	Could the project be perceived as discriminating against any groups? If so, what measures could be taken to ensure this does not happen? (ebd.: 208)
Beneficence	Beneficence	Who benefits from the project and in what way? (ebd.: 208)
	Universal service	Will training be provided to those who do not (yet) have computer skills or knowledge of the Internet? Who should provide the training and under what conditions? (ebd.: 208)
	Accessibility	Does the new technology or service or application expect a certain level of knowledge of computers and the Internet that some people may not have? (ebd.: 208)
	Value sensitive design	Is the project or technology or service being designed taking into account values such as human wellbeing, dignity, justice, welfare, human rights, trust, autonomy and privacy? (ebd.: 209)
	Sustainability	Is the project, technology or service economically or socially sustainable? If not, and if the technology or service or project appears to offer benefits, what could be done to make it sustainable? (ebd.: 209)
Justice	Justice	Has the project identified all vulnerable groups that may be affected by its undertaking? (ebd.: 210)
	Equality and fairness (social justice)	Does the project or policy apply to all people or only to those less powerful or unable to resist? (ebd.: 210)
Privacy and data protection	Collection limitation (data minimization) and retention	Is information collected in ways of which the data subject is unaware? (ebd.: 211)
	Data quality	If the information collected is not accurate, what consequences might ensue? (ebd.: 211)
	Purpose specification	Regarding the project, technology or service, are individuals aware that personal information is being (is to be) collected, who seeks it, and why? (ebd.: 212)
	Use limitation	Is the personal information used for the purposes given for its collection, and do the data stay with the original collector, or do they migrate elsewhere? (ebd.: 212)
	Confidentiality, security and protection of data	Who will have access to any personal data collected for the project or service? (ebd.: 212)
	Transparency (openness)	Has the data controller made known publicly that he has or intends to develop a new database, the purpose of the database, how the database will be used and what opportunities exist for persons to rectify inaccurate personal information? (ebd.: 213)
	Individual participation and access to data	Have measures been put in place to facilitate the person's access to his or her personal data? (ebd.: 213)
	Anonymity	Has the project taken steps to ensure that persons cannot be identified from the data to be collected? (ebd.: 214)
Privacy of personal communications: monitoring and location tracking	Privacy of personal communications: monitoring and location tracking	Does the project monitor or record a person's communications? If so, is it with the person's consent? (ebd.: 215)
	Privacy of the person	Does the project involve biometrics, e.g., taking fingerprints or eye scans? (ebd.: 215)
	Privacy of personal behaviour	Does the project involve surveillance of individuals or groups of people? If so, what is the legal basis of such surveillance? (ebd.: 215)

Tabelle 3: Ethical Impact Assessment: Prinzipien, Kategorisierungen und Beispielfragen

Wright versteht seine Auflistung nicht als vollständig, sondern eher als mit Hinweischarakter behaftet, sodass sie jeweils auf eine bestimmte Problemlage (begründet) weiter zugeschnitten werden müssen.

Auf Human+ bezogen zeigt sich, dass nicht alle Problemlagen gleichermaßen relevant scheinen. Als Technik für Fachanwender (und nicht für eine breite Schicht) wirken etwa Aspekte der sozialen In- und Exklusion weniger relevant als etwa Aspekte des Informed Consent (etwa bezogen auf diejenigen, deren Social-Media-Verhalten analysiert wird). Auch können zahlreiche Beispielfragen von Wright im Rahmen der Human+ Analyse nur tentativ beantwortet werden, etwa weil die Anwendungsbedingungen angesichts eines TRL 4 Projekts oft noch unklar sind. Dies trifft etwa für die oben angeführte Problemlage „use limitation“ zu. In manchen Fällen, soviel kann schon vorausgeschickt werden, lässt sich dieser Umstand für die Human+ Analyse dadurch entschärfen, dass verschiedene Szenarien für mögliche Anwendungsbedingungen analysiert und in ihren Konsequenzen untersucht werden.

Während obige Tabelle die Problemsphären illustriert, gibt Wright auch praktische Anleitungen dazu, in welchen Settings bzw. mit welchen Methoden (die er „**ethical tools**“ nennt) diese mit den StakeholderInnen diskutiert werden können (vgl. Wright 2011: 215-218). Er nennt und diskutiert hierbei u.a. Befragungen (im Sinne von wissenschaftlichen Surveys), Expertenworkshops, Fragechecklisten, Ethical Matrix, Ethical Delphi, Consensus Conferences und Citizen Panels. Ergänzt werden die ethical tools um sogenannte **Prozeduren** oder **Praktiken**, wobei auch hier Reflexionsfragen angeführt werden, die bei der Umsetzung oder Entscheidung helfen sollen (vgl. Wright 2011: 218-223). Ausgeführt werden dabei (das Etablieren eines Prozesses für) die Identifikation und den Einbezug von StakeholderInnen, die Durchführung einer Risikobewertung (für die Behandlung von Unsicherheiten und nicht-intendierten Konsequenzen), die Behandlung (und Klärung) des Themenfelds Verantwortung und Zurechenbarkeit, externer ethischer Evaluation bzw. Audits, der proaktiven Informationsversorgung von StakeholderInnen und dem Umgang mit Beschwerden seitens dieser, sowie Richtlinien für die Integration der vorhin genannten Aspekte hin zu einem strukturierten Ansatz für die Durchführung eines Ethical Impact Assessments (welches etwa Briefing Papers für Stakeholder impliziert und andere eher praktische Hinweise). Zwar können diese Hinweise für die praktische Berücksichtigung von ethischen Aspekten während der laufenden Projektarbeit durchaus wertvoll sein, an dieser Stelle wird aber im Rahmen der vorliegenden Arbeit nicht weiter darauf eingegangen, da Human+ ein bereits abgeschlossenes Projekt und die vorliegende Arbeit sich somit stärker auf diejenigen Aspekte fokussiert, die ohne den weiteren Einbezug von StakeholderInnen behandelbar sind.

Während Brey die detaillierte Checkliste des Frameworks lobt (vgl. Brey 2012: 308), kritisiert er die fehlenden methodischen Angaben dazu, wie die erwarteten Konsequenzen der Technik (auf Artefakt-, Anwendungs- oder sozialer Ebene) prognostiziert werden können, sondern es würde vorausgesetzt,

dass die EntwicklerInnen der Technik und andere StakeholderInnen irgendwie zu diesen Informationen gelangen können. Ähnliches gilt für die Beantwortung der Fragen, wie sie Wrights Checkliste enthält.

5.6 TECHNO-ETHISCHE ZUKUNFTSSZENARIEN

Als ein Hilfsmittel für politische EntscheidungsträgerInnen, um „**soft impacts**“ (also Auswirkungen auf soziale Beziehungen und das Moralverständnis) von technologischen Entwicklungen abzuschätzen, verstehen Boenink et al ihren Ansatz von techno-ethischen Zukunftsszenarien (vgl. Boenink et al 2010: 2).³³

Boenink et al unternehmen die Entwicklung eines „[...] framework that policy makers could use to anticipate soft impacts of technological development, in particular the mutual interaction between technology and morality“ (Boenink et al 2010: 3). Technologische Entwicklungen erweisen sich, so die AutorInnen, oft als bedeutsame Triebfeder auch des **Wandels der Moral** (definiert als Set von Werten und Normen, die eine bestimmte Gesellschaft für wichtig erachtet, weil sie sich auf berechnete Interessen, gegenseitige Verpflichtungen oder Perspektiven zum „guten Leben“ bezieht). Dieser Verknüpfung wird nach Ansicht der AutorInnen in keinem der bestehenden Ansätze zur Antizipation von „soft impacts“ von neuen Techniken Rechnung getragen (und sie erwähnen u.a. das oben angeführte eTA von Palm und Hansson [2006]). Diese Ansätze unterliegen nämlich einem „**moral presentism**“, wonach Bewertungen von Technikzukünften sich ausschließlich am gegenwärtigen Moralverständnis ausrichten, ohne zu berücksichtigen, dass sich mit der Technikentwicklung auch das Moralverständnis ändert (vgl. Boenink et al 2010: 2). Auch wenn etwa Palm und Hansson als eine ihrer Checklisten-Dimensionen den Einfluss auf moralische Werte nennen, würden sie letztendlich von Moral als einem stabilen, robusten Phänomen ausgehen, d.h. methodisch diesem Umstand nicht Rechnung tragen (vgl. ebd.: 5).

Ihren eigenen Ansatz verstehen sie als Beitrag aus dem Bereich von szenarienbasierten Methoden, dessen gemeinsamer Kern in der Beschreibung von unterschiedlichen, möglichen (und nicht zwangsläufig wahrscheinlichen) Zukünften in meist narrativer Form liege (vgl. ebd.: 6). Gegenüber kurzfristigen Ansätzen (wie eTA), die auf laufenden Austausch mit den beteiligten TechnikerInnen während der Entwicklung abzielen und somit in ihrem Horizont auf wenige Jahre beschränkt sind, würden sich szenarienbasierte Ansätze auf längere Zeiträume beziehen und somit Reflexionen über Langzeit-Veränderungen ermöglichen (zu welchen sie auch den moralischen Wandel zählen). An den bestehenden szenarienbasierten Ansätzen kritisieren Boenink et al allerdings ein ebenfalls nicht-dynamisches Moralverständnis, sodass etwa moralische Konflikte und Kontroversen bzw. moralische Verschiebungen

³³ „Hard impacts“ umfassen demgegenüber Auswirkungen von Technik auf Gesundheit und Sicherheit, wofür es nach Ansicht der AutorInnen ein wesentlich besser ausgebautes Instrumentarium gäbe als wie für die „soft impacts“.

nicht inkludiert werden, sondern eher von statischen, stabilen und sehr schematischen moralischen Einstellungen zur Erklärung von Ablehnung/Gutheiung von Techniken ausgegangen wrde.

Zur Entwicklung von **techno-ethischen „Zuknften“** entwerfen die AutorInnen somit ein eigenes Framework, welches ausdrcklich einen **ethischen Pluralismus** zulsst (und sich somit nicht im speziellen auf Konsequentialismus, deontologisches Denken etc. festlegt; vgl. ebd.: 8f). Da es auerdem um den **moralischen Wandel** geht (welcher sich etwa in einer radikal unterschiedlichen Interpretation von moralischen Prinzipien, einer Vernderung von Abwgungsgewichtungen ber die Zeit, oder auch einem Bedeutungsverlust jener manifestieren kann), beziehen sich diese Szenarien auf Zeitrume, die etwa 10-30 Jahre in der Zukunft liegen. Moralischer Wandel verlaufe auf unterschiedlichen Ebenen unterschiedlich schnell: whrend auf der **Makroebene** sich ein Wandel nur sehr langsam vollzieht (etwa die Zunahme der Gewichtung des Prinzips des Respekts vor der Autonomie whrend des 20. Jahrhunderts), wrde auf dem **Mesolevel** in Form von institutionalisierten Praktiken (Regeln und Prozeduren, die diese Prinzipien der Makroebene in Form von „**moralischen Regimes**“ konkretisieren, z.B. in das Konzept der informierten Einwilligung) der Wandel schneller vor sich gehen. Auf der **Mikroebene** dagegen geht es um spezifische moralische Probleme innerhalb konkreter lokaler Rahmenbedingungen, sodass sie hier von „Nischen“ reden, in denen moralische Probleme verhandelt werden und Vernderungen vergleichsweise hufig vor sich gehen.

Um Zukunftsszenarien historisch fundiert zu entwerfen, ist nicht nur ein Kenntnis der gegenwrtigen ethischen Debatte notwendig, sondern die AutorInnen verweisen auch auf die Erfordernis der Kenntnis von Argumentationsmustern und Tropen, die sich auf Basis bisheriger Debatten erschlieen lassen, und von denen auch in zuknftigen Debatten ausgegangen werden darf (vgl. ebd.: 9). Zum Entwurf der **dichten Beschreibung** eines Zukunftsszenarios muss ber historisches Wissen auch abgeschtzt werden, welche Teile der Moralitt eher vernderungsoffen und welche tendenziell stabil sind, sowie welche anderen Entwicklungen auf Basis aktueller sozialer Trends plausibel sind.

Folgende drei Schritte sind ntig, um ein **techno-ethisches Szenario** zu konstruieren (vgl. Boenink et al 2010: 11-14):

- a) **Landkarte aktueller moralischer Debatten entwerfen:** Zunchst wird ein berblick ber die aktuellen moralischen berzeugungen, Praktiken und Regulierungen (inkl. damit verbundener Kontroversen, etc.) hinsichtlich einer bestimmten fraglichen technischen Entwicklung (bzw. relevanten Vorgngertechnologien) skizziert, vorzugsweise mit Hinweisen zu ihrer Genese. Dies ist der Ausgangspunkt fr den nchsten Schritt, nmlich:

- b) **Erstellung potentieller moralischer Kontroversen durch Bezug auf NEST-ethics:** NEST-ethics steht für „ethics of New and Emerging Science and Technology“³⁴, und umfasst ein Inventarium an Tropen, Argumentationsmustern und formalen Strukturen, wie sie üblicherweise in ethischen Kontroversen über Technologien vorkommen. Auf Basis dessen wird erstens skizziert, was die aktuellen Versprechungen und Erwartungen gegenüber der neuen Technik sind, zweitens welche kritischen Einwände dagegen erhoben werden könnten (auf Basis üblicher ethischer Argumentationsmuster in unterschiedlichen Ansätzen), und drittens Argumentationsketten (Einwände auf diese Einwände und dergleichen; z.B. durch Skizzieren möglicher Hinweise auf die Nachteile der Technik, unvorteilhafter Verhältnisse von Vorteilen und Risiken, oder auch durch Spannungen mit Vorstellungen vom „guten Leben“ und dergleichen).
- c) **Konstruktion von möglichen diskursiven Schließungen bzw. Abschlüssen dieser solcherart konstruierten Debatten:** Die in Schritt 2 erzeugten Argumentationsketten und Debatten werden reduziert auf mögliche Abschlüsse, d.h. welche Auflösungen der Debatten und der Entscheidungsfindung als wahrscheinlich und sinnvoll gelten, vor dem Hintergrund von zu erwartenden sozialen und moralischen Trends (d.h. es geht hier ausdrücklich nicht um bloße rationale Auflösung, sondern um wahrscheinliche Ausgänge vor dem Hintergrund von skizzierten, zukünftig-fiktiven Settings, z.B. politischen Kräfteverhältnissen, möglichen Vorfällen, die die Debatten beeinflussen können, und dergleichen). Zu beachten ist hierbei, dass unterschiedliche Bereiche der Moral unterschiedlich stabil gegenüber Wandel sind (v.a. auf der Makroebene, d.h. bezogen auf grundsätzliche Prinzipien, gibt es üblicherweise nur langsame Veränderungen), und der Einbezug von bekannten sozialen Langzeittrends diesen Skizzen zusätzliche Plausibilität verleihen kann. Diese hypothetischen Schließungsszenarien sind kontingent, weshalb die AutorInnen empfehlen, unter Umständen auch zwei oder mehr mögliche Schließungsszenarien zu entwickeln.

Die AutorInnen verdeutlichen ihren Szenarienansatz, indem sie ein hypothetisches, mehrere historische Entwicklungsphasen aufgeteiltes Beispiel-Zukunftsszenario zur Bionanotechnologie und Experimenten mit Menschen entwerfen, bei welchen sie zeitliche Entwicklungsschritte, fiktive Zeitungsmeldungen (um die Narration dichter zu gestalten), politische Kräfteverhältnisse, kontingente Einzelfälle (z.B. Selbstmorde die mit der Technologie in Verbindung stehen, aber die öffentliche Diskussion anheizen können) und dergleichen beschreiben (vgl. ebd.: 16-32). Zu bemerken ist hierbei, dass die Konstruktion eines solchen techno-ethischen Szenarios häufig nur unter Einbezug von externem Expertenwissen erfolgen kann. So veranstalteten die AutorInnen für das Beispielszenario auch einen ExpertIn-

³⁴ Vgl. hierzu Swierstra und Rip (2007).

nenworkshop mit politischen EntscheidungsträgerInnen und AkteurInnen aus dem Feld der TA. Außerdem werden sie in der Regel nicht von Einzelpersonen, sondern von einem ganzen Forschungsteam entworfen. Sie sind Produkt von internen Debatten von mehreren in die Thematik eingearbeiteten KollegInnen, und erfahren mehrere Überarbeitungsdurchläufe mit zunehmender Plausibilisierung, in welchen Schritt 2 (Konstruktion von Argumentationsketten) und Schritt 3 (diskursive Schließung) eng miteinander verzahnt sind. Auf Grund der anwachsenden Breite des Feldes kann es auch nötig sein, bestimmte zu erwartende Kontroversen auszusparen und begründet sich auf bestimmte Aspekte zu beschränken (vgl. Boenink et al 2010: 14ff).

Im Unterschied zu den meisten anderen der hier vorgestellten Ansätzen bezieht sich dieser Ansatz somit nicht auf Checklisten und hat auch keine festen Bewertungsdimensionen, sondern diese werden erst im Rahmen der Skizzierung der moralischen bzw. diskursiven „Landschaft“ bezüglich der fraglichen Techniken erarbeitet. Brey sieht als einen limitierenden Faktor allerdings den Umstand, dass der Ansatz deskriptiv und prädiktiv, anstatt normativ und vorschreibend sei (vgl. Brey 2012: 309). Das Framework zeige moralische Probleme auf, die wahrscheinlich im Rahmen des technologischen Prozesses auftauchen und in die öffentliche Wahrnehmung geraten, aber nicht solche, die aus ethischer Perspektive thematisiert werden sollten. Ähnlich deskriptiv wird der Lösungsprozess betrachtet: der Ansatz betrachtet wie diese Probleme wahrscheinlich gelöst werden, aber nicht notwendigerweise wie sie aus ethischer Sicht gelöst werden sollten. Zudem werden unter Umständen solche moralische, öffentliche Kontroversen identifiziert, die auf missverstandenen Annahmen (z.B. der Technik oder ihrer sozialen Konsequenzen) beruhen. Für den Kontext der vorliegenden Arbeit erscheint der gerade vorgestellte Ansatz auf Grund der Erfordernisse an ein Forschungsteam sowie des avisierten zeitlichen Horizonts (10 – 30 Jahre) und des Erkenntnisinteresses weniger relevant. Die Analyse von Human+ kann zudem ausdrücklich als Forschungsarbeit im Rahmen des oben angeführten „moral presentism“ gesehen werden, da mögliche ethische Problemlagen aus gegenwärtiger Perspektive betrachtet werden, und Prognosen zum Wandel der Moral in den nächsten Jahrzehnten hier nicht angestellt werden.

5.7 ATE-FRAMEWORK

Ein weiterer Ansatz für technikethische Betrachtungen im Bereich von emergierenden Informationstechnologien der Ansatz „anticipatory technology ethics“ (ATE) nach Brey (vgl. Brey 2012). Brey versteht den Fokus seines Ansatzes besonders auf der Forschungs- und Entwicklungsphase einer neuen Technik, d.h. noch bevor eine Technik marktreif ist und zuerst von einigen wenigen Personen angewandt wird. Da in der Forschungs- und Entwicklungsphase (wie auch noch in der Einführungsphase) zahlreiche Geräte, Nutzungsmuster und soziale Konsequenzen noch nicht empirisch gegeben wären, müsse die ethische Beurteilung sich auf bestimmte Entwicklungsaktivitäten fokussieren und dann

mögliche Geräte und Nutzungsmuster (spekulativ) in die Zukunft projizieren, die dann ethisch bewertet werden (vgl. Brey 2012: 306).

Vom bereits dargestellten ETICA-Framework übernimmt Brey drei Ebenen, mit jeweils spezifischen Themenbereichen für die ethische Analyse, die entweder allesamt einbezogen werden können, oder im Sinne von Schwerpunktsetzungen gezielt behandelt werden (vgl. ebd.: 310-314):

- **Technologie** [technology] als Sammlung von Abläufen bzw. Prozeduren, die zusammengehören, weil sie einem gemeinsamen Zweck oder Bereich angehören, oder gemeinsame formale oder funktionale Eigenschaften aufweisen (z.B. Atomtechnik), wobei dies in weitere Sub-Technologien unterteilt sein kann. Auf dieser Ebene bezieht sich die ethische Analyse entweder auf generische ethische Probleme, **die den Eigenschaften der Technologien inhärent sind**, oder aber die sich aus daraus ableitbaren **Folgen** der Anwendung in fast allen Artefakt- oder Anwendungs-Instanzen dieser Technologie ergeben, oder auf **bestimmte Risiken** (z.B. dass als moralisch problematisch erachtete Artefakte oder Anwendungen entwickelt werden könnten).
- **Artefakte** [artifacts] als physikalische Konfigurationen, die bei ordnungsgemäßer Bedienung einen bestimmten erwünschten Zweck erreichen, aber auch (erneut) Prozeduren als definierte Sequenz von Handlungen, die bei ordnungsgemäßer Durchführung (mit den vorgesehenen Werkzeugen etc.) ebenfalls ein bestimmtes Ergebnis bewusst herbeiführen. Auf dieser Ebene fokussiere die ethische Analyse erneut entweder auf **inhärenten** Problematiken von Artefakten, auf moralisch kontroversielle **Folgen**, entweder weil diese typischerweise mit (fast) allen Anwendungen einhergehen, oder weil ganz bestimmte Anwendungen spezifischer Artefakte **risikenbehaftet** oder moralisch kontroversiell sind.
- **Anwendungen** [applications], als konkrete Verwendung von technologischen Artefakten oder Prozeduren für bestimmte Zwecke oder in einem bestimmten Kontext, bzw. als konkrete Art der Verwendung von Artefakten oder Prozeduren (z.B. industriell versus häusliche Anwendungen, langfristige oder kurzfristige Anwendungen, Unterschiede nach Eigenschaften der AnwenderInnen, Aspekte der sozialen Situation, und dergleichen). Auf dieser Ebene kann sich die ethische Reflexion entweder auf die Legitimität der **intendierten** Anwendung für ganz bestimmte Zwecke ausrichten, oder auf moralische Probleme die sich als Nebenfolgen oder **nicht-intendierte Konsequenzen** dieser Anwendungen ergeben, und zwar entweder für die AnwenderInnen selbst oder für nicht direkt in der Anwendung beteiligten StakeholderInnen.

Versucht man die eingangs skizzierte Human+ Technik einem dieser drei Bereiche zuzuordnen, zeigt sich, dass Human+ auf alle drei Arten betrachtet werden kann. So kann Human+ als **Technologie** im Sinne einer Sammlung von Abläufen und Prozeduren (nämlich der grundlegenden Algorithmen, mittels derer Migrationsprognosen angestellt werden sollen, sowie in Form der abstrakten Konzepte dahinter)

verstanden werden. Außerdem kann Human+ als **Artefakt** betrachtet werden, wenn etwa der Demonstrator als Webanwendung in den Blick genommen wird. Schließlich kann die hypothetische Human+ Verwendung analysiert werden, im Sinne der **Anwendung** von Artefakten und Prozeduren in bestimmten Kontexten (etwa im Rahmen einer sogenannten besonderen Aufbauorganisation, vgl. vorausgreifend hierzu S. 117). Aufgrund dessen wird im Abschnitt zu den Analysen zu Human+ auf einen breiten Querschnitt der „typischerweise“ diesen Bereichen zugeordneten ethischen Problemfelder eingegangen werden: auf moralisch kontroverse Folgen der Anwendung und damit verbundenen Risiken, sowie die Legitimität der Anwendung für bestimmte Zwecke (genauer gesagt: die Legitimität dieser Zwecke selbst).

Für die ethische Analyse obiger Bereiche ist nach Brey nicht nur der Einbezug von Zukunftsprognosen [forecasts] nötig (entweder aus der Literatur oder ergänzt mit eigens durchgeführten ExpertInnenbefragungen und ExpertInnendiskussionen, z.B. aus dem Bereich TA, Soziologie, Geschichte), sondern auch die Zusammenarbeit der EthikerInnen mit den bei der Entwicklung beteiligten IngenieurInnen (vgl. ebd.: 312). Dies dient dem Zusammenstellen von **Beschreibungen** zu den fraglichen Techniken, die dann zwei notwendigen weiteren Phasen unterzogen werden (vgl. ebd.: 312-314):

a) **Identifizierungsphase**: Hier werden die Beschreibungen in Beziehung gesetzt zu ethischen Werten und Prinzipien, um zu prüfen, ob Eigenschaften einer Technik sich negativ auf diese Werte oder Prinzipien auswirken. Brey schlägt hierzu ein Verfahren vor, das er „**Operationalisierung**“ der Werte oder Prinzipien nennt; gemeint ist damit „[...] a description of it that specifies real-world conditions for its realization or frustration“ (Brey 2012: 313), womit Brey dem weiter oben bereits erläuterten Konzept der Spezifizierung einen anderen Namen gibt. Die relevanten Werte oder Prinzipien können einer „ethical checklist“ entnommen sein, um zu verhindern, dass bestimmte Aspekte übersehen werden. Brey selbst schlägt (ähnlich zu anderen schon geschilderten Ansätzen) eine „vorläufige“ Checkliste vor, die gesellschaftlich und ethisch weithin akzeptierte Dimensionen umfasst, wie sie sich in technikethischen Publikationen bewährt haben (vgl. Brey 2012: 313):

Oberkategorie	Unterkategorie	Unterkategorie Detaillierung
Harms and risks	Health and bodily harm	
	Pain and suffering	
	Psychological Harm	
	Harm to human capabilities	
	Environmental harm	
	Harms to society	
Rights	Freedom	...of movement
		... of speech and expression
		... of assembly
	Autonomy	Ability to think one's own thoughts and form one's own opinions
		Ability to make one's own choices
		Responsibility and accountability
		Informed consent
	Human dignity	
	Privacy	Information privacy
		Bodily privacy
		Relational privacy
	Property	Right to property
		Intellectual property rights
		Other basic human rights as specified in human rights declarations (e.g., to life, to have a fair trial, [...])
	Animal rights and animal Welfare	
Justice (distributive)	Just distribution of primary goods, capabilities, risks and hazards	
	Nondiscrimination and equal treatment relative to age, gender, sexual orientation, social class, race, ethnicity, religion, disability, etc.	
	North-south justice	
	Intergenerational justice	
	Social inclusion	
Well-being and the common good	Supportive of happiness, health, knowledge, wisdom, virtue, friendship, trust, achievement, desire-fulfillment, and transcending meaning	
	Supportive of vital social institutions and structures	
	Supportive of democracy and democratic institutions	
	Supportive of culture and cultural diversity	

Tabelle 4: Anticipatory Technology Ethics (ATE): Checkliste

Um der Unvollständigkeit von Ethik-Checklisten zu begegnen, empfiehlt auch Brey eine Analyse der Ethikliteratur zu bestimmten Techniken oder Bereichen, um sie auf spezifische Themen anzupassen, bzw. die Befragung unterschiedlicher StakeholderInnen, um im Sinne eines Bottom-up-Vorgehens zusätzliche Werte oder Prinzipien einbeziehen zu können.

b) **Evaluierungsphase:** Diese nach der Identifizierungsphase erfolgende Phase dient der Bewertung der Schwere der Probleme bzw. ihrer Gewichtung (inkl. Abwägungen im Falle möglicher Wertekonflikte) bzw. der Beurteilung der Wahrscheinlichkeit, dass diese Probleme in der Gesellschaft wirksam werden.

Anschließend an diese beiden notwendigen Phasen können in einer optionalen **Design-Feedback-Phase** die Ergebnisse der Evaluierungsphase in die Entwicklung zurückgespielt werden, bzw. in einer **Verantwortungszuteilungs-Phase** die moralische Verantwortung auf unterschiedliche im Entwicklungsprozess von Artefakten und Anwendungen beteiligte Akteure zugeteilt werden, oder in einer ETICA-angelegten **Governance-Phase** (Policy-)Empfehlungen für EntscheidungsträgerInnen ausgearbeitet werden (vgl. Brey 2012: 314). Der Einbezug dieser drei optionalen Phasen würde nach Brey ATE zu einem vollständigen Responsible-Research-and-Innovation-Ansatz ausbauen.

5.8 MENLO REPORT

Als IKT-spezifische Zuspitzung des Principlism lässt sich der Menlo Report (vgl. Menlo 2012 und Menlo 2013) bezeichnen: er fußt auf dem nach einem Konferenz-Center benannten Belmont Report, welcher im Nachgang an unethische Humanexperimente wie die Tuskegee-Syphilis-Studien³⁵ von der „National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research“ verfasst wurde, um ethische Erfordernisse für die Forschung am Menschen aufzustellen, die auch in die US-amerikanische Gesetzgebung und etwa Auflagen für den Erhalt von Fördermitteln eingeflossen sind (vgl. Menlo 2013: 3).³⁶ Der Belmont-Report stellt drei ethische Prinzipien auf, die sich in drei konkreten Erfordernissen an die Forschungspraxis niederschlagen (vgl. für die Prinzipien Belmont 1978a: 4-10, für die forschungspraktischen Auswirkungen Belmont 1978a: 10-20):

- **Respekt vor Personen** [respect for persons], dies umfasst u.a. die Behandlung von Individuen als autonom Handelnde, und den Schutz von Personen mit verminderter Handlungsfähigkeit, was sich forschungspraktisch in der Notwendigkeit informierter Zustimmung niederschlägt
- **Wohltun** [beneficence], dies umfasst erstens die Regel der Nichtschädigung, und zweitens die Regel der Maximierung vorteilhafter Folgen mit der Minimierung von schädlichen Folgen, mit der forschungspraktischen Erfordernis einer systematischen Risiken-/Nutzen-Bewertung, welche auch das Verhältnis von Nutzen für die Gesellschaft und Risiken für die beteiligten Personen betrachtet;
- **Gerechtigkeit** [justice], im Sinne einer „fairen“ Verteilung von Risiken und Nutzen der Forschung bzw. der Gleichbehandlung von gleich gelagerten Fällen (wobei im Report auf unterschiedliche Arten von fairer Verteilung eingegangen wird: Gleichverteilung, Verteilung nach Bedürfnis, bzw. Varianten von leistungsabhängiger Verteilung), was sich praktisch in Anforderungen an die (nicht von Diskriminierungsmechanismen beeinflusste) Auswahl von Forschungssubjekten und Stakeholdern niederschlägt

Die Ähnlichkeiten zu den weiter oben bereits erläuterten Hauptprinzipien des Principlism nach Beauchamp und Childress sind nicht zufällig: sowohl Beauchamp wie auch Childress haben in der Kommission

³⁵ Dabei handelt es sich um humanmedizinische Experimente in den USA ab 1932, welche erst 1972 der Öffentlichkeit bekannt gemacht wurden (vgl. Menlo 2013: 3). Dabei wurden gezielt afroamerikanische Männer aus niedrigen Einkommenschichten entweder mit der Syphilis infiziert, oder aber ihnen wurde die Behandlung von bereits vorhandenen Syphilis-Infektionen zugesagt, die sie letztendlich aber nicht erhielten. Denn: Die dabei involvierten MedizinerInnen wollten die Langzeit-Effekt von (unbehandelter) Syphilis untersuchen. Die Probanden dieser Experimente wurden über den wahren Zweck im Unwissen gelassen, und erhielten letztendlich keine wirksame Behandlung, sondern ihnen wurde sogar aktiv Schaden zur Generierung von Wissen zugefügt, sodass diese Experimentreihe ein vielzitiertes Beispiel für unethische Forschungspraxis im humanmedizinischen Bereich ist.

³⁶ Auf US-gesetzliche Bestimmungen oder Auflagen wie das Aufsetzen eines „Institutional Review Board“ bzw. eines „Research Ethics Board“ (vgl. Menlo 2013: 4) wird in diesem Abschnitt nicht weiter eingegangen.

zum Belmont-Report fachliche Beiträge geleistet, wie etwa aus dem Anhang zum Belmont Report ersichtlich wird, dem sie in schriftlicher Form beigelegt sind (vgl. Belmont 1978b). Der augenscheinlichste Unterschied liegt darin, dass im Principlism von Beauchamp und Childress die Prinzipien der Nichtschädigung und des Wohltuns nicht wie im Belmont-Report in einem einzigen Prinzip zusammengefasst sind.

Der Menlo Report versucht nun, die auf die direkte Forschung am Menschen bezogenen Prinzipien und forschungspraktischen Maßnahmen (informierte Zustimmung, Risiken-/Nutzen-Analyse, Auswahl von involvierten Akteuren) für den oft nur **indirekt** auf Menschen bezogenen IKT-Forschungs-Kontext neu auszulegen und zu ergänzen, da sich Forschung im IKT-Bereich hinsichtlich des Ausmaßes an involvierten Personen, der Geschwindigkeit, der Kopplung von bis dahin getrennten Systemen (z.B. können Smartphones sensible Daten unterschiedlichster Quellen enthalten), Dezentralisierung, zeitlicher und räumlicher Verteilung und hinsichtlich der Komplexität der involvierten Technik(en) stark von klassischer Forschung **direkt** an Menschen unterscheidet (vgl. Menlo 2013: 6f). Stattdessen kann im IKT-Kontext oft auf im Vorfeld erstellte Datenbanken rekuriert werden, oder IKT selbst (und nicht spezifisch Menschen) als Forschungsobjekte in den Blick genommen werden.

Die oben genannten Prinzipien behalten auch im IKT-Kontext ihre Gültigkeit, werden im Menlo-Report aber auf neue Schwierigkeiten hin reflektiert:

- **Respekt vor Personen** [respect for persons]: Das Einholen einer **informierten Zustimmung** wird im IKT-Kontext erschwert durch die je nach Forschungsprojekt unklare Frage, von wem überhaupt eine Zustimmung einzuholen ist, wie die betroffenen Gruppen (welche nunmehr neben natürlichen Personen auch Firmen oder Organisationen umfassen können) kontaktiert werden können (die entweder räumlich disloziert oder hinter unterschiedlichen technischen „Service“-Schichten oder Applikationen verborgen sein können). Erschwerend wirkt auch die Tatsache, dass Forschungsprojekte im IKT-Kontext durch die einbezogene Infrastruktur und Algorithmen (z.B. Forschung an infiltrierenden Botnetzen im Falle von Cyberkriminalität oder Social-Media-Auswertungen) schwieriger greifbar und damit weniger leicht zu erklären sind, v.a. wenn nicht auf Aspekte wie graphisch anschauliche Benutzeroberflächen rekuriert werden kann (vgl. Menlo 2013: 8).
- **Wohltun** [beneficence]: In die notwendige strukturierte Risiko-/Nutzen-Analyse müssen im IKT-Kontext neue Formen von Risiken einbezogen werden, z.B. Privatheits-Risiken die sich auf die Gefahr der Veröffentlichung von persönlichen Daten beziehen (was erschwert wird durch kontextabhängige und sich stetig entwickelnde Definitionen davon, was etwa ein Datum zu einem persönlichen Datum macht) bzw. anderer Risiken wie z.B. Tracking oder die Zusammenführung von bis dahin getrennten „virtuellen Identitäten“ durch neue Forschungsprojekte und

Applikationen. In diesem Rahmen wird auch die Vorbereitung eines Reaktionsplans für vernünftigerweise vorhersehbare Schäden, sowie ein allgemeiner Notfallplan für schwerwiegende Risiken mit niedriger Eintrittswahrscheinlichkeit vorgeschlagen.

- **Gerechtigkeit** [justice]: Während im IKT-Kontext die häufige (Teil-)Anonymisierung von betroffenen Personen es schwieriger macht, basierend auf sensiblen Eigenschaften wie Geschlecht, Religionszugehörigkeit, Rasse/Ethnie/Hautfarbe etc. (klassisch) zu diskriminieren, können die Charakteristika von IKT zu neuen Gerechtigkeitsproblemen führen, z.B. dass die Natur und das Ausmaß von Forschungsrisiken falsch eingeschätzt werden. Ein Beispiel ist etwa, wenn die **Veröffentlichung von Ergebnissen** zu einer Analyse von Schwachstellen in einem Sicherheitssystem schneller von potentiellen AngreiferInnen rezipiert werden als von den Personen, die für deren Verbesserung zuständig wären.

Ergänzt werden die drei klassischen Prinzipien aus dem Belmont Report im Menlo Report um ein viertes Prinzip: **das Prinzip des Respekts vor dem Gesetz und dem öffentlichen Interesse** [respect for law and public interest], welches implizit im Prinzip des Wohltuns inkludiert gewesen sei, nun aber auf Grund der Vielzahl an relevanten Gesetzesmaterien im IKT-Kontext, konfligierenden und uneindeutigen Gesetzeslagen über mehrere geopolitische Gesetzgeber, Schwierigkeiten in der Identifizierung von Stakeholdern, und Inkongruenzen zwischen Gesetz und öffentlichem Interesse notwendig geworden wäre (vgl. Menlo 2013: 10). Forschungspraktisch wirkt sich dieses Prinzip zum einen in „**Compliance**“ aus (d.h. Beachtung von relevanter Gesetzesmaterien und anderer Regularien, wie etwa organisationsbezogene Auflagen, Best-Practices oder auch „culturally appropriate procedures or other guidelines“ [Menlo 2013: 17], sowohl in der Risikobewertung wie in der Durchführung der Forschung selbst). Zum anderen kann darunter auch „**Transparenz und Zurechenbarkeit von Verantwortung**“ [transparency and accountability] verstanden werden, was externe Bewertungsorgane (z.B. der wissenschaftlichen Gemeinschaft oder Ethics Review Boards) es erleichtert, die Einhaltung ethischer oder fachlicher Maßstäbe zu überprüfen. Gleichzeitig nimmt das die beteiligten ForscherInnen in die Pflicht, sich dieser Verantwortung bewusst zu sein und sie zu berücksichtigen. Für Transparenzzwecke schlagen die AutorInnen des Menlo Reports vor, eine Art Forschungssteckbrief zu entwerfen, der Auskunft über die Beteiligten, die Sponsoren, die Forschungsziele, die Forschungsdauer, die Zielpopulation, die Methoden, das Ausmaß des Einbezugs von Menschen, möglicher Schäden, Risiken und positiver Faktoren, Zeiträume, Veröffentlichungen, Hintergrundannahmen und dergleichen gibt (vgl. Menlo 2013: 18).

Um die Berücksichtigung der vier Prinzipien des Menlo Reports bzw. der daraus folgenden forschungspraktischen Anwendungen zu unterstützen, führen die AutorInnen eine Reihe von Hilfsfragen an, wovon exemplarisch und zur Verdeutlichung der Struktur in der nachfolgenden Tabelle jeweils eine angeführt ist:

Prinzip	Forschungspraktische Anwendung	Beispielfrage
Respekt vor Personen [respect for persons]	Identifikation von Stakeholdern [identification of stakeholders]	“Who owns, controls or authorizes the use of ICT resources, or the collection, use and disclosure of data related to those resources?” (Menlo 2013: 11)
	Informierte Zustimmung [informed consent]	“Are the following justifications for not obtaining informed consent present: (a) foregoing consent is truly necessary to accomplish research goals (b) all known risks are minimal (c) there is an adequate plan for debriefing subjects, when appropriate, and (d) obtaining consent truly impacts research validity and is not just an inconvenience to the researcher?” (ebd.: 12)
Wohltun [beneficence]	Identifikation möglicher Schäden/Leid [identification of potential harms], ...	
	... davon Integritätsrisiken [integrity risks]	“Are risks to the integrity of not only information, but also the information systems used to store and process information, considered?” (ebd.: 13)
	... davon Verfügbarkeitsrisiken [availability risks]	“Have all risks to availability of information and information systems considered, including disruption due to overloading links, corruption of communication or routing pathways, or exhaustion of resources?” (ebd.: 13)
 davon Vertraulichkeitsrisiken [confidentiality risks]	“Research has exposed limitations and inaccurate assumptions about the efficacy of data anonymization techniques. If harm from re-identification is foreseeable or plausible, have the risks of re-identification been considered? Is the sole means of protection based on anonymization of a defined set of identifiers [...]? How accessible are secondary data sources that can be combined with published data to re-identify individuals?” (ebd.: 13)
	... davon Risiken der Geheimhaltung bzw. mangelnder Transparenz [secrecy and lack of transparency]	“Does research involve recording or monitoring individuals’ behavior or location across time and place, resulting in harms related to surveillance? Direct harms may include: identity theft, revelation of embarrassing information, government persecution, costs associated with evading surveillance.” (ebd.: 13)
	... davon allgemeine Risiken [general risks]	„What is the severity of potential harms to all persons who may be affected by research activities (e.g., collection, use or disclosure of data, publication of research results, or the interaction with ICT)? Certain people may be at a higher level of risk than others.” (ebd.: 14)
	Identifikation möglicher Vorteile [identification of potential benefits]	“Does the research activity clearly benefit society?” (ebd.: 14)
	Abwägung von Schäden und Vorteilen [balancing risks and benefits]	“Does the ICTR consider only collecting and maintaining personal data that are adequate, relevant and not excessive in relation to the research purposes for which they are collected and/or further processed? Could the research be conducted without the collection, use, or disclosure of the data? If ICTR involves human subjects surveillance, are minimization techniques and processes used? E.g., limited collection, purpose specification, limited data use, limited data retention, etc.?” (ebd.: 15)
	Eindämmung von eingetretenen Schäden [mitigation of realized harms]	“Does the mitigation protocol consider the cause(s) and extent of the risk exposure? Is exposed sensitive data a systemic problem or an isolated incident? Did data exposure result from external malicious behavior or an untargeted exposure? Was the data lost or stolen? Will a targeted theft result in further harmful activity?” (ebd.: 16)
Gerechtigkeit [justice]	Fairness und Gleichheit [fairness and equity]	“Does the research target certain groups by selecting research subjects based on race, sex, religious affiliation, or other legally protected attribute? If so, how does this satisfy Justice concerns?” (ebd.: 17)
Respekt vor dem Gesetz und dem öffentlichen Interesse [respect for law and public interest]	Compliance [compliance]	„Does ICTR violate federal or state criminal laws, civil laws or regulations, or other nation’s laws? If the ICTR conflicts with a law or regulation, is there an exception or agreement that permits research?” (ebd.: 17)
	Transparenz und Zurechenbarkeit von Verantwortung [transparency and accountability]	“What are the attributes of the environment being studied that justify the use of the proposed ICTR methodology to achieve the stated goal(s)? Have the goals of the research been adequately described?” (ebd.: 18)

Tabelle 5: Prinzipien und Beispielfragen im Menlo-Report

Für die Analyse von Human+ werden einerseits einige der IKT-Spezifizierungen der obigen Prinzipien berücksichtigt werden, etwa dass das (Unterlassen eines) Einholen(s) von „informed consent“ von betroffenen Personen nicht nur im humanmedizinischen, sondern auch auf im IKT-Kontext einen moralischen Unterschied machen kann. Andererseits sensibilisiert obige Auflistung für die unterschiedlichsten Arten von Risiken, die beim Einsatz von Techniken auftreten können.

Der Menlo-Report illustriert, ähnlich wie der oben schon erläuterte Ansatz des Ethical Impact Assessments (Wright 2011), wie Prinzipien mittlerer Reichweite dediziert auch in IKT-Kontexten Anwendungen finden können, und dass diese Prinzipien mitunter für solche Kontexte um eigene Prinzipien ergänzt werden können. Zudem illustrieren sie, dass die im sehr ausgefeilten Principlism-Ansatz von Beauchamp und Childress argumentierte Fundierung dieser Prinzipien in der gemeinsamen Moral in Zuspitzungen dieses Ansatzes häufig unterlassen wird: weder im Menlo Report noch im Artikel zum Ethical Impact Assessment von Wright (2011) findet sich etwa der Terminus „common morality“. Auch erfolgt die „Spezifizierung“ eher in Form von Erläuterungen, welche Unterdimensionen man im Rahmen der Reflexion zu den verschiedenen Prinzipien im IKT-Kontext beachten könnte. Ein stark formalisierter Prozess, wie er für die „Principles of Biomedical Ethics“ ausgearbeitet wurde, findet sich im Menlo Report oder bei Wright dagegen nicht.

5.9 DESSI

Nicht begrenzt auf den Bereich von informationsbezogenen Entwicklungen ist das Bewertungsframework „Decision Support on Security Investments“ (Kurzform: **DESSI**). DESSI dient der Unterstützung für die Entscheidungsfindung hinsichtlich verschiedener Alternativen, wie mit bestimmten Bedrohungen [threats] umgegangen werden kann (vgl. Kreissl und Čas 2012: 8ff). Das Grundprinzip hinter DESSI besteht in einem breiten partizipativen Beratungsprozess, in welchem VertreterInnen unterschiedlichster Perspektiven (z.B. fachspezifische ExpertInnen und Laien, VertreterInnen unterschiedlichster Stakeholder-Gruppen) gleichberechtigt in einem Diskurs die Problemlage, und/oder mögliche Lösungsvorschläge anhand von mehreren Dimensionen strukturiert diskutieren und erarbeiten. Die Lösungsvorschläge begrenzen sich nicht im Vorhinein auf technische Lösungen, sondern sollen auch soziale oder organisatorische Lösungen für Sicherheitsprobleme in den Blick bekommen helfen. Die Dimensionen, anhand derer diskutiert wird, dienen als Referenzrahmen, die nacheinander durchlaufen werden, um unterschiedliche Aspekte der Sicherheitsinvestition aus den unterschiedlichen Perspektiven der beteiligten Akteure zu durchleuchten, sodass DESSI von seinen ErstellerInnen auch als „diskursiver Algorithmus“ bezeichnet wird. In der Liste der Dimensionen finden sich folgende Punkte, wobei für jede Dimension zur Unterstützung des diskursiven Prozesses sechs Kriterien bzw. unterstützende Fragestellungen ausgearbeitet wurden (vgl. Čas und Kaufmann 2012: 13):

- Security gain or loss
- Fundamental Rights and Ethics
- Legal Framework
- Social Implications
- Acceptability

- Political Significance
- Economy

Lediglich die ersten zwei Dimensionen haben einen die anderen Dimensionen überstimmenden Charakter: wenn eine Investition keinen Sicherheitsgewinn bringt, ist sie überflüssig. Maßnahmen, die gegen Grundrechte verstoßen, müssen die Verhältnismäßigkeit der Maßnahme und ihre Notwendigkeit in einer demokratischen Gesellschaft beweisen, da ansonsten eine Betrachtung der anderen Dimensionen sich ebenfalls erübrigt (vgl. hierzu DESSI 2011).

DESSI ist primär kein ethisches Bewertungsframework, sondern ist eher einem umfassenderen Ansatz von partizipativer TA zuzurechnen. Dennoch ist es im Kontext der vorliegenden Arbeit naheliegend, in diesem Überblick die Beachtung ethischer Aspekte im DESSI-Framework näher zu beleuchten. Die Bezeichnung der ethisch relevanten Dimension als „**Fundamental Rights and Ethics**“ verweist bereits auf das Kontinuum zwischen (z.B. in der Charta der Grundrechte der EU) verbrieften Grundrechten einerseits, stärker der gruppenspezifischen Aushandlung bedürftigen Werten andererseits, und ethischen Normen in der Mittelposition (vgl. Kaufmann 2012: 21). Da Sicherheitsmaßnahmen häufig in einer Grauzone zwischen Beachtung und Überschreitung von Rechten arbeiten (z.B. im Falle von Antiterrormaßnahmen), versuchten die AutorInnen für DESSI einen strukturierten Ansatz für die Behandlung dieser Dimension zu erarbeiten. Dies soll geleistet werden, indem folgende inhaltliche Aspekte in der gemeinsamen Diskussion behandelt werden, von denen die ersten drei der aus der Grundrechtecharta der EU entnommen wurden, und die anderen teilweise einem Katalog ethischer Normen entstammen (Kaufmann 2012: 24ff):

Kriterium	Beschreibung	Erklärung/Beispiel
Private life	Does the security investment respect private zones, the right to private data, the right to your own picture etc.?	The right to privacy is often discussed in relation to technological solutions, surveillance measures and data-retention.
Freedoms	Does the security investment respect freedoms of thought, conscience, religion, expression and information?	The debate on censorship classically relates to freedoms of expression and information.
Non-discrimination	Does the measure discriminate against any societal group? Does it allow for diversity, equality and value-pluralism?	Profiling measures are generally known for infringing upon the right to non-discrimination.
Transparency	Are the security investment itself and its effects easy to understand? Is it clearly communicated what the measure includes and entails?	For political reasons and potential referendums on security measures a clear communication of advantageous and detrimental consequences is necessary.
Social Cohesion	Does the security investment further trust and confidence in the government and within society (between citizens)?	If security investments openly infringe upon fundamental rights they might not only alienate the citizen from the government, but also create a culture of distrust, separation and fear among society.
Proportionality	Is the security investment suitable, necessary and proportional to the problem?	Should the measure not follow these assessments of suitability, necessity and proportionality, the investment should be re-planned. The assessments of criteria 1-5, as well as Security gain or loss and Social Implications, can also help to assess proportionality.

Tabelle 6: Bewertungsdimensionen im DESSI-Framework zu „Fundamental Rights and Ethics“

Die Begrenzung der zu betrachtenden Kriterien für diese Dimension auf 6 Items entspringt ausdrücklich der Erfordernis einer praktischen Handhabbarkeit für diskursive Prozesse, andererseits wurden Doppelungen vermieden, wenn relevante Aspekte bereits in anderen Wertedimensionen abgedeckt wurden, z.B. Übereinstimmung mit Gesetzen in der Dimension „Legal Framework“ oder „Innovativeness and Constructiveness“ in den Dimensionen „Economy“ und „Political Significance“ (vgl. Kaufmann 2012: 23).

Im Rahmen der vorliegenden Arbeit wird kein partizipativer Ansatz verfolgt, sodass die große Stärke von DESSI, nämlich die Anwendbarkeit in Diskurssettings mit verschiedenen StakeholderInnen, hier keine Anwendung findet. Allerdings werden einige der Kriterien, die der ethischen Dimension von DESSI zugeordnet werden, auch in der Human+ Analyse beleuchtet, etwa Aspekte und Probleme bezüglich der Transparenz der Human+ Anwendung (etwa beim verdeckten oder offenen Einsatz von Überwachungskamera für das Generieren von In-Situ-Daten) wie auch von Privatheit (z.B. bei der Datenauswertung von eigentlich zu anderen Zwecken gedachten Social-Media-Daten).

5.10 VALUE SENSITIVE DESIGN

Seinem Selbstverständnis nach ist Value Sensitive Design (VSD)

„[...] a theoretically grounded approach to the design of technology that accounts for human values in a principled and comprehensive manner throughout the design process. It employs an integrative and iterative tripartite methodology, consisting of conceptual, empirical and technical investigations“ (Friedman et al 2006: 348).

VSD erhebt somit den Anspruch, Werten [values] im Design von Technologie bzw. Technik Rechnung zu tragen. „**Wert**“ wird dabei definiert als „[...] what a person or group of people consider important in life“ (Friedman et al 2006: 349). Betont wird dabei die Trennung zwischen Werten und Fakten: Fakten käme nicht automatisch auch Wert zu, sodass gemäß des sogenannten „Naturalistischen Fehlschlusses“ (allein) aus Sein nicht Sollen folgen könne. Deshalb reichen empirische Beschreibungen der Welt nicht für die Berücksichtigung der Wertdimensionen aus, sondern Interessen und Wünsche von Menschen innerhalb bestimmter kultureller Milieus müssen ausdrücklich berücksichtigt werden. Zudem betonen die AutorInnen, dass Werte nicht unbedingt moralische Werte sein müssen (wie etwa das Recht auf Privatheit), sondern auch nichtmoralischer Natur sein können (z.B. ästhetische Präferenzen). Eine der Grundannahmen von VSD ist zudem, dass Werte nicht fest in Technologie bzw. Technik eingeschrieben sind oder aber nur durch soziale Kräfte transportiert werden, sondern die tatsächliche Anwendung in sozialen Kontexten abhängt von den **Zielen der AnwenderInnen**, auch wenn Eigenschaften der Technik durch das Design dazu führen können, dass manche Werte tendenziell eher befördert und andere behindert werden (vgl. Friedman et al 2006: 361). Während zudem **bestimmte** Werte (bei abstrakter Definition) praktisch universell angenommen werden können, kann die konkrete Ausgestaltung sich nach sozialem, zeitlichem und kulturellem Kontext unterscheiden (z.B. Privatheit bei Bewohnern von Einfamilienhäusern vs. Privatheit bei Bewohnern von Iglus). Je handlungsfokussierter die Konzeptualisierung von Werten ist, desto sensibler für kulturelle Variationen wird die Definition sein.

Die Autoren geben eine übersichtliche Auflistung von Werten, wie sie typischerweise oftmals beim Design von Systemen impliziert sind (Friedman et al 2006: 364f):

Human Value	Definition
Human Welfare	Refers to people's physical, material, and psychological well-being
Ownership and Property	Refers to a right to possess an object (or information), use it, manage it, derive income from it, and bequeath it
Privacy	Refers to a claim, an entitlement, or a right of an individual to determine what information about himself or herself can be communicated to others
Freedom from Bias	Refers to systematic unfairness perpetrated on individuals or groups, including preexisting social bias, technical bias, and emergent social bias
Universal Usability	Refers to making all people successful users of information technology
Trust	Refers to expectations that exist between people who can experience goodwill, extend goodwill toward others, feel vulnerable, and experience betrayal
Autonomy	Refers to people's ability to decide, plan, and act in ways that they believe will help them achieve their goals
Informed Consent	Refers to garnering people's agreement, encompassing criteria of disclosure and comprehension (for "informed") and voluntariness, competence, and agreement (for "consent")
Accountability	Refers to the properties that ensures that the actions of a person, people, or institution may be traced uniquely to the person, people, or institution
Courtesy	Refers to treating people with politeness and consideration
Identity	Refers to people's understanding of who they are over time, embracing both continuity and discontinuity over time
Calmness	Refers to a peaceful and composed psychological state
Environmental Sustainability	Refers to sustaining ecosystems such that they meet the needs of the present without compromising future generations

Tabelle 7: Value Sensitive Design - Wertetabelle

Im Prozess des VSD werden drei Untersuchungsphasen durchlaufen:

a) **Konzeptuelle Untersuchungen** [conceptual investigations] (vgl. Friedman et al 2006: 350f): In dieser Phase geht es um die Beantwortung von Fragen wie: Welche Stakeholder-Gruppen sind direkt und indirekt betroffen durch das fragliche Design, und wie? Welche Werte werden impliziert? Wie kann zwischen konkurrierenden Werten abgewogen werden im Design, der Implementierung und der Verwendung von Informationssystemen (z.B. Autonomie vs. Sicherheit, bzw. Autonomie vs. Vertrauen)? Und sollen moralische Werte höheres Gewicht als nichtmoralische Werte bekommen? Zudem werden in dieser Phase **philosophische Definitionen** [working conceptualizations] von spezifischen Werten erarbeitet, die Aufklärung im Bereich grundlegender projektbezogener Themenfelder bieten sollen, und für Vergleichbarkeit über unterschiedliche Forschungsteams hinweg sorgen. Friedmann et al illustrieren dies am Begriff „Vertrauen“, welcher neben einer grundlegenden Definition auch eine Zuspitzung auf den Bereich „Vertrauen im Onlinebereich“ erfährt. Dieser Definitionsprozess dient dazu, im Vergleich unterschiedlicher Forschungsarbeiten transparenter zu machen, wenn unterschiedliche Bedeutungen nuanciert werden.

b) **Empirische Untersuchungen** [empirical investigations] (vgl. Friedman et al 2006: 351): Diese sind in vielen Fällen zur Ergänzung der vorherigen Phase notwendig. Hier geht es um die Klärung des menschlichen Kontextes, in welchem ein technisches Artefakt verortet ist, und fokussiert nicht auf die Technik selbst, sondern auf die Individuen, Gruppen oder sozialen Systeme die durch eine Technik betroffen sind. In diesen Bereich fällt auch die Bewertung des Erfolgs von bestimmten Designs. Da empirische Untersuchungen auf jede beobachtbare oder messbare menschliche Aktivität angewandt werden können, fallen unterschiedliche Methoden in diesen Bereich (quantitative und qualitative Methoden der

empirischen Sozialforschung, wie Beobachtungen, Interviews, Befragungen, Experimente, und dergleichen). Diese Phase dient u.a. der Beantwortung von Fragen wie StakeholderInnen bestimmte Werte in interaktiven Kontexten auffassen, wie sie konfligierende Werte bzw. Werte und Benutzerfreundlichkeits-Faktoren abwägen, ob es Unterschiede zwischen berichteten und gelebten Wertedimensionen gibt, und wie beispielsweise Organisationen in der Entwicklung neuer Technologien und Techniken Werte-Dimensionen in den Design-Prozess einfließen lassen (z.B. durch etablierte Trainingsmethoden, Belohnungsstrukturen, ökonomische Anreize, und dergleichen).

c) **Technische Untersuchungen** [technical investigations] (vgl. Friedman et al 2006: 351): In dieser Phase geht es einerseits um die Beantwortung der Frage, wie bestimmte technische Eigenschaften und zugrunde liegende Mechanismen bestimmte (in der ersten Phase erarbeitete) Werte unterstützen oder behindern, sowie andererseits wie durch proaktives Design von Systemen die Werte befördert werden können, die in den konzeptuellen Untersuchungen erarbeitet wurden. Hier steht somit anders als in den empirischen Untersuchungen die Technik selbst im Vordergrund.

Die verdeutlichenden Fallbeispiele, die die AutorInnen anschließend anführen, zeigen, dass es je nach fraglichem Projekt schon im Vorhinein um einen ganz zentralen Wert gehen kann, z.B. Privatheit, oder aber bis hin zu einer fast unüberschaubaren Vielfalt von Werten, deren Bedeutung erst im Untersuchungsprozess durch Stakeholderbefragungen erarbeitet wird. Die konkrete Anwendung des VSD muss daher flexibel und bis zu einem gewissen Grad theoriegeleitet erfolgen, sodass sich wie weiter oben im Bereich der TA jeweils Begründungen für Entscheidungen als notwendig erweisen.

Diese drei Phasen können jeweils iterativ erfolgen, sodass etwa Ergebnisse aus den empirischen Untersuchungen nicht nur das technische Design beeinflussen (proaktives Design im Sinne der technischen Untersuchungen), sondern auch neue Bestandteile für in den konzeptuellen Untersuchungen ausgearbeitete Definitionen hinzufügen. Der Ablauf der drei Phasen wird daher auch als „**iterative and integrative**“ bezeichnet (vgl. Friedman et al 2006: 353).

Neben diesen Grundsätzen geben die AutorInnen eine Reihe von (teilweise detailliert ausgeführten) Anwendungshinweisen, welche grob zusammengefasst folgende sind (vgl. Friedman et al 2006: 362-367):

- Beginne mit dem/der (für das Projekt zentralsten) Wert / Technik / Anwendungskontext, und arbeite dich zu den Implikationen für die übrigen weiter
- Identifiziere direkte und indirekte Stakeholder
- Identifiziere Gefahren/Schäden/Risiken und Vorteile für jede Stakeholdergruppe (und berücksichtige Unterschiede der technischen, kognitiven oder physischen Leistungsfähigkeit)
- Weise (wenn möglich) Vorteile und Gefahren/Schäden/Risiken korrespondierende Werte zu

- Führe eine konzeptuelle Untersuchung von Schlüsselwerten durch (unterstützt durch die relevante Literatur)
- Identifiziere potentielle Wertekonflikte (und wie sie sich einschränkend auf den Möglichkeitsraums des Designs auswirken, eventuell unterstützt durch Diskussionen mit StakeholderInnen)
- Integriere die Berücksichtigung von Werten in die Organisationsstruktur (z.B. der Firma oder staatlichen Agentur, die eine Technik entwickelt)
- Orientiere dich (als Ausgangspunkt) an einer Liste von moralisch relevanten Werten
- Verwende bei Interviews mit StakeholderInnen die „Warum“-Frage, und Frage nach Werten nicht direkt, sondern indirekt, und unterziehe sie im Vorfeld einer Konzeptualisierung und Abgrenzung durch formelle Kriterien
- Skizziere, wie durch Wertekonflikte im Design von Techniken notwendig gemachte Alternativen von Design-Entscheidungen sich jeweils unterschiedlich auf StakeholderInnen auswirken
- Designe wenn möglich die technische Infrastruktur auf eine Weise, das auch im Nachhinein noch nicht-antizipierte Werte und Wertekonflikte berücksichtigt werden können

Mit VSD liegt somit ein Ansatz vor, welcher sich auf kein Principlism-Fundament bezieht, sondern stattdessen Werte in das Zentrum der Aufmerksamkeit holt. Welchen Unterschied macht das? Werte wurden eingangs in diesem Unterkapitel definiert als intersubjektiv geteilte oder auch nur persönliche Vorstellungen davon, was im Leben wichtig ist. Mit Ropohl kann ergänzt werden, dass Werte üblicherweise aber nicht bloß subjektive Wunschkomplexe sind, sondern einen überindividuellen Geltungsanspruch mit zumindest teilweiser gesellschaftlicher Zustimmung haben (vgl. Ropohl 2016: 323). In jedem Fall sind Vorstellungen davon, was „wichtig“ ist, für sich genommen noch keine moralischen Prinzipien. Ein moralisches Prinzip ist eine moralische Norm (vgl. Beauchamp und Childress 2009: 13). Ein Wert dagegen ist noch keine moralische Norm, kann aber nach Ansicht des Verfassers der vorliegenden Arbeit als Grundlage für die Formulierung eines Prinzips dienen. Wenn etwa, wie oben dargestellt, im VSD-Ansatz Privatheit als Wert gehandhabt wird, dann könnte darauf bezugnehmend ein Prinzip formuliert werden, welches besagt, dass auf die Privatheit von moralischen Subjekten Rücksicht zu nehmen ist.

Ob die Verwendung von Prinzipien statt Werten tatsächlich forschungspraktisch einen Unterschied macht, hängt nach Ansicht des Verfassers der vorliegenden Arbeit von der konkreten Ausformung der zugrunde gelegten „Wertelehre“ ab. Betrachtet man die Wertetabelle des VSD-Ansatzes, so fällt die Vielfalt möglicher Werte auf, und Friedman et al verweisen darauf, dass dies „typischerweise“ wichtige Werte sind, wie sie im Design von Systemen impliziert sein können, und verstehen sie als nichtabgeschlossene, beispielhafte Liste. Dies rückt den VSD-Ansatz in die Nähe von Checklisten-basierten An-

sätzen, deren Vollständigkeit jeweils vor dem Hintergrund einer bestimmten Aufgabenstellung zu hinterfragen und gegebenenfalls zu verbessern ist. Die Liste von Friedman et al enthält sowohl „Informed Consent“ wie auch „Autonomy“, und zwar auf derselben „Ordnungsebene“ prima facie gleichberechtigt nebeneinandergestellt. Demgegenüber wollen Beauchamp und Childress, wie schon dargestellt, „Informed Consent“ als Spezifikation des Prinzips des Respekts vor der Autonomie verstanden haben. Im Principlism von Beauchamp und Childress gibt es daher eine Hierarchie von Ableitungen bzw. Spezifikationen von konkreteren Normen („Regeln“) aus abstrakteren Normen, eben den „Prinzipien“ (die aber prima facie dann ebenfalls auf einer gleichberechtigten Ebene liegen), welche von Beauchamp und Childress auch nicht als lediglich „beispielhaft“ betrachtet werden, sondern als weitgehend sogar kulturinvariant geteiltes, gemeinsames Fundament von moralischen Akteuren, welchen daher auch nicht ohne weiteres zusätzliche Prinzipien dieses hohen Abstraktionsgrades hinzugefügt werden können. Demgegenüber scheint der VSD-Ansatz zumindest in der Darstellung von Friedman et al keine unterschiedlichen Hierarchieebenen von Werten vorzusehen. Diese Unterschiede suggerieren zunächst, dass Arbeiten, die sich auf Werte statt auf Prinzipien berufen, grundlegend andere Vorgangsweisen erfordern.

Aber: Principlism-Ansätze haben demgegenüber die Notwendigkeit, durch Spezifikation ihre Geltung vor einer je gegebenen Problemlage anzupassen, und wie weiter unten noch sichtbar werden wird, stellt sich hier das Problem, dass dieselbe Spezifikation unter mehrere Prinzipien subsumiert werden könnte, und daher womöglich eine begründete, aber bis zu einem gewissen Grad kontingente, Setzung der damit befassten EthikerInnen notwendig wird. Die „klare“ Hierarchie von Regeln, die vor dem Hintergrund von Prinzipien spezifiziert werden, wird so zumindest etwas verwischt.

Auch findet sich bei Ropohl der Versuch, einen „ethischen Minimalkonsens“ auf Basis von Werten zu konstruieren. Er argumentiert, dass für die Bewertung einer Technik oder von Folgen einer Technik nicht lediglich tagesaktuelle Kriterien herangezogen werden dürfen, sondern „alle Werte menschlicher Lebensqualität“ einzubeziehen, sodass eher von „**Wertsystemen**“ (also von Werten und ihren systematischen Beziehungen zueinander) die Rede sein sollte (vgl. Ropohl 2016: 223ff). Der Terminus „Wertsystem“ lässt zumindest die Möglichkeiten einer Strukturiertheit erahnen, wie sie bei einem bloßen Katalog von nebeneinander gestellten Werte-Checklisten so noch nicht erkennbar ist.

Werte sind für Ropohl außerdem wichtig, um (bei Ropohl: teleologisch-utilitaristisch) die Qualität von Handlungsfolgen zu beurteilen (vgl. Ropohl 2016: 139). Forschungspraktisch kann Ähnliches auch von Prinzipien mittlerer Reichweite gesagt werden, wenn diese etwa zur Konstruktion einer Ethical Matrix herangezogen werden, welche positive und negative Betroffenheiten unterschiedlicher Stakeholdergruppen vor dem Hintergrund der Prinzipien aufzeigen.

Weiter oben wurde bereits das Framework „Ethical Impact Assessment“ dargestellt (vgl. Wright 2011). Bei jenem Ansatz wurden den Prinzipien mittlerer Reichweite nicht nur moralische Regeln in einem Spezifikationsprozess zugeordnet, sondern ausdrücklich verwendet Wright in seiner Terminologie auch den Begriff „Werte“, um die abstrakten Prinzipien zu spezifizieren.

Beiden Ansätzen ist außerdem gemein, dass Anschlussstellen für interdisziplinäre Forschung bestehen. Im VSD können etwa besonders in der Phase „empirische Untersuchungen“ auch sozialwissenschaftliche Methoden angewandt werden, z.B. um zu eruieren, wie StakeholderInnen zwischen unterschiedlichen Werten abwägen, oder wie unterschiedliche Werte verstanden werden. Auch in Principlism-basierten Ansätzen, wie z.B. unter Bezug auf die Ethical Matrix dargestellt wurde, können „Fakten“ als Ergebnisse von interdisziplinärer Forschung eingebunden werden, z.B. wenn es um die möglichen Betroffenheiten von Stakeholdergruppen durch eine Technik geht. Im VSD-Ansatz scheint allerdings ein insgesamt stärker deskriptiver Zugang vorzuliegen, da zumindest in manchen Varianten weniger Vorannahmen über die zu berücksichtigenden Werte gemacht werden, während im Principlism die einbezogenen Prinzipien mittlerer Reichweite präskriptiven Charakter haben, und somit klarer über ethisch zulässige oder problematische Aspekte Auskunft geben.

Ob man mit einem Werte-basierten Zugang andere Aspekte in der ethischen Analyse in den Blick bekommt als bei einem Principlism-basierten Zugang, lässt sich daher pauschal nicht beantworten, sondern hängt von den obigen Faktoren ab: ob Werte eher deskriptiv genutzt werden, oder stärker als Unterdimension oder „Wurzel“ von ethischen Prinzipien mittlerer Reichweite in den Blick genommen und damit einen präskriptiveren Charakter annehmen. Zumindest auf Basis der in der vorliegenden Arbeit betrachteten Literatur scheinen Prinzipien-basierte Ansätze in der Technikethik häufiger vorzukommen, was auch eine wechselseitige Übernahme von verschiedenen Komponenten dieser Principlism-basierten Ansätzen erleichtern dürfte – schon allein aufgrund der anschlussfähigen Terminologie. Demgegenüber dürfte zumindest ein stärker deskriptiver Zugang wie beim VSD hier noch eine zusätzliche Übersetzungsleistung erfordern, um etwa die Bedeutung von „Werten“ im begrifflichen Repertoire des Frameworks und in Relation zu anderen Begriffen wie „Prinzipien“ und „Regeln“ klarzustellen.

5.11 5-SCHRITTE-SCHEMA ETHISCHER ENTSCHEIDUNGSFINDUNG

Bleisch und Huppenbauer erarbeiteten ein „Schema ethischer Entscheidungsfindung“, welches allgemein für Projekte und Maßnahmen geeignet sein soll, und sich nicht auf eine bestimmte Anwendungssphäre (z.B. IKT) beschränkt. Es ist gedacht als „Werkzeugkasten“, um „[...] moralische Konflikte einer stichhaltigen und ethisch vertretbaren Lösung zuzuführen“ (Bleisch und Huppenbauer 2014: 14). Aufgrund seines universellen Zuschnitts, des klaren Ablaufes und seiner augenscheinlichen Ergänzenbarkeit

durch Elemente anderer Ansätze soll es an dieser Stelle ebenfalls dargestellt werden. Das Schema gestaltet sich folgendermaßen (ebd.: 15):

Schema ethischer Entscheidungsfindung nach Bleisch und Huppenbauer	
1. Schritt	
Analyse des Ist-Zustandes	<ul style="list-style-type: none"> • Harte Fakten auflisten • Geltendes Recht berücksichtigen • Stakeholder identifizieren • Kontextsensibilität entwickeln
2. Schritt	
Die moralische Frage benennen	<ul style="list-style-type: none"> • Moralisch relevante Frage(n) und Konflikte identifizieren • Strittige Frage formulieren • Nichtmoralische Aspekte ausscheiden
3. Schritt	
Analyse der Argumente	<ul style="list-style-type: none"> • Argumente pro und kontra aufführen • Moralische Normen und Werte rekonstruieren • Argumente mit normativen Hintergrundtheorien abgleichen
4. Schritt	
Evaluation und Entscheidung	<ul style="list-style-type: none"> • Standpunkt der Moral einnehmen • Argumente beurteilen und gewichten • Einen Entscheid fällen
5. Schritt	
Implementierung	<ul style="list-style-type: none"> • Möglichkeiten der Implementierung abschätzen • Maßnahmen zur erfolgreichen Implementierung ergreifen • Alternative Perspektiven ausloten

Tabelle 8: Schritte im Schema ethischer Entscheidungsfindung nach Bleisch und Huppenbauer

Während das Schema selbst klar und übersichtlich ist, verwenden die AutorInnen auf jeden der einzelnen Schritte ein eigenes Kapitel, um Hintergründe und Anwendungsaspekte weiter zu vertiefen und auch epistemologische Aspekte zu erläutern (z.B. was die Schwierigkeit bei der Benennung „harter Fakten“ betrifft, etwa die Unvermeidbarkeit selektiver Beschreibungen und Werturteilen, Aspekte der Unsicherheit des Wissens oder der Verlässlichkeit von Quellen, vgl. ebd.: 18f). Auf diese Zusatzausführungen kann hier nur begrenzt eingegangen werden.

Für die vorliegende Arbeit relevant (und in den bisherigen Abschnitten thematisch so auch noch nicht angesprochen) sind die Ausführungen zur Benennung der moralischen Frage. „**Moralische“ Fragen** (die etwa so eingeleitet werden können: „Ist es moralisch legitim, ...“) weisen demnach folgende Charakteristika auf:

- Es sind normative Fragen mit Handlungsbezug, im Sinne von: Was soll ich (man) tun, welche Handlungsoptionen sind zu wählen? Nicht im Zentrum stehen somit Aspekte der Machbarkeit (Was „kann“ man tun?), sondern sie weisen einen Zusammenhang mit den drei „**deontischen Operatoren**“ auf: Gebot/Pflicht, Verbot, Erlaubnis (vgl. ebd.: 37). Die Normativität muss im Fall des Ablaufschemas sich allerdings auf moralische Normativität beziehen, sodass Fragen von Anstand (Sitte/n/Bräuchen), Klugheit, Geschmack oder Recht nicht gemeint sind.

- Ihre Lösung ist allgemein verbindlich, d.h. ethische Lösungen verlangen danach, dass sie von allen Menschen in der gleichen Situation berücksichtigt werden sollen (vgl. ebd.: 39). Sie soll „**universalisierbar**“ sein.³⁷ Ihre Formulierung muss daher auch unparteilich erfolgen, und sie muss argumentativ rechtfertigbar sein, d.h. an Vernunft und Empfinden anderer appellieren.³⁸
- Sie betreffen zentrale Güter und Werte von Menschen und Gesellschaften (vgl. ebd.: 41f), die durch die Befolgung von moralischen Regeln und Normen geschützt werden sollen. Moralische Konflikte lassen sich daher auch als Wertekonflikte oder Güterabwägungen beschreiben. Moralische Werte definieren sie als „[...] jene Werte, die den fundamentalen Bedürfnissen sowie legitimen Erwartungen und Interessen entsprechen, die Menschen (und allenfalls einige Tiere) haben.“ (ebd.: 42).

Für die Formulierung der „strittigen“ (moralischen) Frage empfehlen die AutorInnen folgendes (vgl. ebd.: 46f):

- Analyse der Konfliktfelder: Was genau ist strittig? Dabei können (fiktiv) identifizierte Stakeholder um einen runden Tisch versammelt werden, um zu überlegen, welche Argumente sie vorbringen würden, und welche Positionen/Forderungen welche Gruppe empören würden.
- Ordnung des Konfliktmaterials: Welche moralische(n) Frage(n) verbergen sich nun hinter dem Konfliktmaterial (hinter Anschuldigungen/Vorwürfen/Rückfragen)? Diese Fragen sollen schriftlich in ein paar Sätzen festgehalten werden. Außermoralische Fragen (juristische/ökonomische/politische Aspekte) sollten hier ausgeschieden werden, um nicht in Sachzwängen zu denken (sie kommen erst in Schritt 5 des Schemas wieder herein).
- Welche Entscheidungs- oder Handlungsoptionen und Alternativen existieren? Hier wird empfohlen, Optionen als strenge Alternativen hinsichtlich von Ja/Nein-Positionen zu formulieren, z.B. „Haben Eltern behinderter Kinder Anspruch auf externe Kinderbetreuung, der rechtlich verbrieft werden sollte?“.

³⁷ Um moralische Legitimität beanspruchen zu können, muss eine Entscheidung vom „Standpunkt der Moral“ aus getroffen werden. Das bedeutet dreierlei (vgl. Bleisch und Huppenbauer 2014: 78ff): 1.) Es wird ein „universaler Standpunkt“ eingenommen, welcher an die Vernunft als allen Betroffenen gleichermaßen zugängliche Instanz appelliert. (**Universalität**) 2.) Die Entscheidung muss „informiert“ getroffen werden und Unvoreingenommenheit durch kritische Selbstdistanzierung erreicht werden. (**Informiertheit und Selbstdistanzierung**, d.h. **Unvoreingenommenheit**) 3.) Sie muss unparteilich getroffen werden (**Unparteilichkeit**). Die AutorInnen verweisen darauf, dass dieser universelle Standpunkt der Moral häufig auch kritisiert wurde, z.B. dass dieser kein wertfreier, sondern ein liberaler Standpunkt sei. Eine Lösung sehen sie in der Idee der Überlegungsgleichgewichts (in der vorliegenden Arbeit auch als reflektives oder Reflexionsgleichgewicht bezeichnet) von John Rawls, welches darin besteht, Gerechtigkeitsgrundsätze mit weit verbreiteten Intuitionen und Wertvorstellungen in ein Gleichgewicht bzw. kohärentes System zu bringen.

³⁸ Bleisch und Huppenbauer verweisen etwa auf den Empörungstest: „Moralische Empörung kann als Indiz dafür gewertet werden, dass wir eine moralische Norm verletzt sehen bzw. dass wir bei jemandem die Aufmerksamkeit vermissen, die moralische Fragen gebühren würde“ (ebd.: 40). Moralische Empörung **beweist** noch keinen moralischen Missstand (z.B. weil das Bauchgefühl irren kann), aber es ist ein Indiz.

Für die Identifizierung von Argumenten in Schritt 3 des Schemas wird empfohlen, eine Liste der StakeholderInnen zu erstellen, und vorgebrachte Argumente für oder gegen eine bestimmte Position diesen StakeholderInnen zuzuweisen (vgl. ebd.: 51). Manchmal verwenden StakeholderInnen dieselben oder analogen Argumente, um ihre Interessen zu verteidigen, und in der weiteren Analyse kann so von den konkreten StakeholderInnen abstrahiert werden, und der sachliche Gehalt der Argumente betrachtet werden. Außerdem sei die Aufgabe der bei der Problemlösung involvierten EthikerInnen, zusätzliche Argumente zu finden, die von den involvierten StakeholderInnen eventuell nicht angedacht wurden, um nicht „[...] bloß zufällige und – unter Umständen – lautstark geäußerte Argumente zu analysieren und zu evaluieren, sondern herauszufinden, welche Argumente eine Position besonders gut stützen oder entkräften“ (ebd.: 53). Dafür schlagen sie die Beleuchtung des Problems explizit aus dem Blickwinkel mehrerer Moraltheorien vor (z.B. Deontologie, Konsequentialismus, oder Tugendethik), d.h. sie beziehen sich auf eine Vielfalt von Moraltheorien, um wie durch „Linsen“ ein moralisches Problem zu betrachten.

In der Beurteilung der Argumente wird, wie schon angeführt, eine Tabellierung von Stakeholdern und den jeweiligen pro-/kontra-Argumenten vorgeschlagen (vgl. ebd.: 83ff). Diese Argumente werden hinsichtlich Gültigkeit und Qualität beurteilt, ungültige werden ausgeschieden (oder aber korrigiert). Im Anschluss an die „Aufbereitung“ der (gültigen) Argumente, werden diese gewichtet, indem sie in kurzen schriftlichen Sätzen zusammengefasst oder nach Stichwortlisten diskutiert werden. Dabei wird gefragt. „Welche Argumente könnten aus welchen Gründen gewichtiger als andere sein?“ (ebd.: 84). Das **Gewichten der Argumente** erfolgt seinerseits erneut durch Argumente, sodass Argumentationszusammenhänge erstellt werden. In vielen Fällen kann Argumentieren hier auch Begriffsklärung bedeuten, um Streitfragen zu lösen. Falls keine eindeutige Lösung gefunden werden kann (d.h. ein Konflikt besteht, in dem zwei oder mehrere gleichwertige Güter nicht gleichzeitig gewählt werden können), kommt es zur **Güterabwägung**. „Güter“ meint die zu diskutierenden Optionen, und kann sowohl Positives wie auch Übel beinhalten. Dies kann erschwert werden, wenn bestimmte Güter als „unantastbar“, d.h. nicht abwägbar betrachtet werden (z.B. in vielen Positionen wäre die Menschenwürde ein solches Gut). Argumente, die **unantastbare Güter** in Frage stellen, wären in dem Falle nicht ethisch zulässige Argumente. Dies betont die Bedeutung einer Begründung der Güterabwägung als Nachweis darüber, weshalb ein Interesse als höher zu gewichten eingestuft wird als ein anderes (bzw. warum ein Recht als unverletzbar betrachtet wird) – was oft den Rückgriff auf Moraltheorien erfordere. Die Entscheidungen aus der Güterabwägung müssen erneut schriftlich in Form von Argumenten festgehalten werden. Dabei ist auf innere Kohärenz (innere Widerspruchsfreiheit) und äußere Kohärenz (als harmonisches Einfügen einer getroffenen Entscheidung oder einer postulierten Norm in ein System anderer verbindlich anerkannter Normen und Werte) zu achten (vgl. ebd.: 89). Beim Kriterium der

äußeren Kohärenz muss abgewogen werden, ob bei einem Verstoß gegen dieses Kriterium eventuell nicht die postulierte Norm, sondern andere bestehende Normen und Regeln einer Korrektur bedürfen.

Mit dem Ablaufschema und Bleisch und Huppenbauer liegt ein Gerüst für die strukturierte Behandlung von ethischen Problemen vor, welches sich nach Ansicht des Verfassers der vorliegenden Arbeit gut mit Aspekten anderer Frameworks ergänzen lässt. Für die Analyse von Human+ wird daher nicht nur der grundlegende Aufbau des Frameworks übernommen und erweitert, sondern auch die Strukturierung von Pro- und Kontra-Überlegungen in Form von bausteinartigen, abgrenzbaren und Stakeholdergruppen zuweisbaren „Argumenten“ (welche dann auch in eine EM übernommen werden können), wird für den nächsten Abschnitt der vorliegenden Arbeit adaptiert.

6 BEWERTUNG VON HUMAN+

6.1 EINLEITUNG

Im vorhergehenden Abschnitt wurde eine Reihe von in der Ethik verwendeten bzw. verwendbaren Ansätzen für die ethische Analyse von technischen und IKT-Projekten vorgestellt. Dabei wurden trotz Unterschieden im Detail zwischen zahlreichen Ansätzen Gemeinsamkeiten auffällig, z.B.

- die formale Separierung von Faktensammlung, darauf aufbauender ethischer Analyse und ethischer Bewertung bzw. Evaluation in unterschiedliche Phasen,
- die Identifikation von StakeholderInnen,
- die Orientierung an (Check-)Listen von Prinzipien mittlerer Reichweite bzw. Werten, wobei beim Vergleich unterschiedlicher Listen Redundanzen und weitgehende Überlappungen auffallen (z.B. die Ergänzung von „Privacy“ im Bereich von Informationstechnologien, auch wenn diese mitunter unterschiedlichen Prinzipien zugeordnet wird oder gar als eigenes Prinzip formuliert wird), in jedem Fall also ein moralischer Pluralismus,
- die Erarbeitung von Umsetzungs- oder Regelungsvorschlägen.

In diesem Kapitel werden die zusammengetragenen Elemente zu einem Bewertungsframework vereinheitlicht, und die Analyse der bereits skizzierten Human+ Technik durchgeführt, um die in der Einleitung aufgeworfenen drei „Hauptfragen“, und damit die Forschungsfrage nach der ethischen Zulässigkeit von Human+, zu beantworten.

Damit das Bewertungsframework auch über den Kontext der vorliegenden Arbeit in der beruflichen Tätigkeit des Verfassers Anwendung finden kann (das „methodische Ziel“ der Arbeit), ist besonders ein übersichtlicher und transparenter Aufbau des Bewertungsschemas wichtig. Dies verbessert die Nachvollziehbarkeit der erzielten Beurteilungen auch für Nicht-EthikerInnen, die den überwiegenden Anteil an ProjektpartnerInnen in der genannten beruflichen Tätigkeit stellen. Aufgrund seines klaren und strukturierten Aufbaus wird daher das Ablaufschema von Bleisch und Huppenbauer (2014) herangezogen, welches für die Beantwortung der drei Hauptfragen erweitert und mit einem Principlism-Ansatz für die inhaltlichen Analysen ergänzt wird. Der genaue Aufbau des Bewertungsframeworks wird im nächsten Unterkapitel noch detaillierter betrachtet. Zusammengeführt wird das Ablaufschema mit einem Principlism-Fundament, da ein solcher Ansatz einerseits im Hintergrund vieler der weiter oben dargestellten Ansätze erprobt wurde, sondern wie Forschungen mit der Ethical Matrix verdeutlichen, auch für die interdisziplinäre Zusammenarbeit, etwa im Rahmen von Workshops mit Betroffenen einer Technik, anschlussfähig ist. Dies wurde zwar im Rahmen der Human+ Analysen nicht unternommen,

stellt aber für die weitere Verwendbarkeit ein weiteres Qualitätsmerkmal dar. Da die Prinzipien mittlerer Reichweite außerdem in der „gemeinsamen Moral“ fundiert sind, wird dem Verständnis von Fachfremden zusätzlich entgegengekommen. Ausdrücklich wird im vorliegenden Ansatz ein „moral presentism“ gewählt (vgl. S. 68), und somit kein zukunftszenarienbasierter Ansatz gewählt. Der Grund hierfür ist, dass einerseits für die Arbeit an solchen Szenarien ein ganzes Forschungsteam notwendig ist, um in iterativen Schritten mögliche „Schließungen“ von Zukunftsentwicklungen gemeinsam diskursiv zu plausibilisieren, und andererseits erfolgt bei einer möglichen Weiterentwicklung von Human+ eine solche innerhalb der nächsten Jahre, sodass weniger der Wertewandel, als eher die aktuelle „moralische Landschaft“ relevant scheint. Das in der vorliegenden Arbeit formulierte Metaprinzip des Negativen Utilitarismus fungiert als Korrektiv, um bei Entscheidungssituationen zu wählen, in denen sich die Alternativen besonders hinsichtlich ihrer negativen Auswirkungen deutlich unterscheiden.

6.2 HARMONISIERUNG DES BEWERTUNGSFRAMEWORKS

Bevor in den nachfolgenden Unterkapiteln jeder Schritt der Human+ Bewertung durchlaufen wird, soll schon vorab ein kurzer Gesamtüberblick über das zur Anwendung kommende Framework gegeben werden, und in welcher Weise Anleihen bei den in Kapitel „5 Ansätze und Frameworks in der Technikethik“ dargestellten Vorgehensweisen genommen wurden. Abbildung 2 verdeutlicht den grundsätzlichen Aufbau des Human+ Bewertungsschemas.

Zur Strukturierung des grundsätzlichen Ablaufes wird das 5-Schritte-Schema von Bleisch und Huppenbauer (2014) übernommen, da es sehr anschlussfähig für Ergänzungen aus den anderen Ansätzen ist. So lässt es etwa offen, auf welche Weise die jeweiligen Argumente für die Beantwortung der moralischen Frage(n) gefunden, formuliert und strukturiert werden, auch wenn es hierzu Vorschläge gibt. Grundsätzlich werden im Schema, wie es zudem auch bei der Ethical Matrix (vgl. Mephram et al 2006), bei ETICA (Stahl 2011), ATE (Brey 2011) und einigen Ansätzen der TA vorgeschlagen wird, Faktensammlung, Aufbereitung der ethischen Frage(n), Argumenten, und Abwägung bzw. Evaluation getrennt.

Klassisch dem Ablaufschema von Bleisch und Huppenbauer entsprechend wird zunächst der „Ist-Zustand“ analysiert, v.a. um Kontextsensibilität zu erreichen. Beispiele der Tätigkeiten in diesem Schritt sind eine **Begriffsanalyse** von Termini im Flucht- und Migrationsbereich, der historische Kontext der „Flüchtlingskrise 2015/2016“ (als initialer Impuls der Human+ Forschung), und die Analyse der StakeholderInnen. Dabei wird v.a. auf die einschlägige Forschungsliteratur zurückgegriffen werden. Auf dieser Ebene werden noch keine ethischen Implikationen erarbeitet. Außerdem erfolgt eine geraffte Wiedergabe von Ergebnissen aus der Anforderungsanalyse des Projekts.

Adaptiert wird der „Mittelteil“ des Ablaufschemas (Schritte 2 bis 4 bei Bleisch und Huppenbauer), denn die **drei Hauptfragen** von Fenner, als Zerlegung der übergeordneten Forschungsfrage: „**Inwiefern sind**

die angestrebte Human+ Technik bzw. ihre Teilkomponenten ethisch zulässig?“, werden hiermit zusammengeführt. Das bedeutet, dass Schritt 2 bis 4 laut Ablaufschema von Bleisch und Huppenbauer nun jeweils für die Beantwortung der Hauptfrage 1 und der Hauptfrage 3 durchlaufen werden. In der Beantwortung von Hauptfrage 2 wird von diesem Schema insofern abgewichen, als dass ein verkürztes Verfahren gewählt wird, in welchem tabellarisch einerseits außerplanmäßige Handlungen identifiziert werden, diese grob hinsichtlich ihrer moralischen Bedenklichkeit kurz charakterisiert werden, und sodann die Möglichkeiten von Gegenmaßnahmen skizziert werden. Der Versuch, zunächst auch für die Beantwortung von Hauptfrage 2 den „Dreischritt“ von „Benennung der Frage“, „Analyse der Argumente“ und „Evaluation und Entscheidung“ durchzuführen, hat sich auf Grund der Heterogenität der möglichen außerplanmäßigen Handlungen nicht bewährt, und hätte zu einer Myriade von hierzu nötigen „Dreischritten“ geführt, nämlich zu jeder potentiellen außerplanmäßigen Handlungsweise.

Im Hintergrund der Analysen für die drei Hauptfragen stehen Prinzipien mittlerer Reichweite, welche im Anschluss an die Charakterisierung des Human+ Ablaufschemas noch einmal wiedergegeben und hinsichtlich ihrer Adaption kommentiert werden. Außerdem wird für die Beantwortung von Hauptfrage 3 eine Ethical Matrix konstruiert, in welcher diese Prinzipien die Spalten bilden und die Ergebnisse der Stakeholderanalyse aus Schritt 1 die Zeilen.

Am Ende des Kapitels zur Beantwortung der Hauptfragen werden diese Ergebnisse zusammengeführt und damit die übergeordnete Forschungsfrage hinsichtlich der ethischen Legitimität von Human+ beantwortet.

Im abschließenden fünften Schritt werden die während der Beantwortung der Hauptfragen gesammelten Vorschläge für die technische Weiterentwicklung von Human+ noch einmal gerafft wiedergegeben, um Empfehlungen für die Entwicklung von Human+ in Richtung eines höheren TRL und für die Festlegung von Anwendungsbedingungen zur Verbesserung der ethischen Verträglichkeit zu bieten.

Human+ Bewertungsframework		
1. Schritt		
Analyse des Ist-Zustandes	<ul style="list-style-type: none"> • Harte Fakten auflisten • (Geltendes Recht berücksichtigen) • Stakeholder identifizieren • Kontextsensibilität entwickeln 	
Beantwortung der drei Hauptfragen (Schritt 2 bis 4)		
Beantwortung von Hauptfrage 1: Sind die Handlungsmöglichkeiten, die durch diese Technik eröffnet oder verbessert werden, an sich ethisch legitim? (primärer Zweck)	2. Schritt für Hauptfrage 1	
	Die moralische Frage benennen	<ul style="list-style-type: none"> • Moralisch relevante Frage(n) und Konflikte identifizieren • Strittige Frage formulieren • Nichtmoralische Aspekte ausscheiden
	3. Schritt für Hauptfrage 1	
	Analyse der Argumente	<ul style="list-style-type: none"> • Argumente pro und kontra aufführen • Moralische Normen und Werte rekonstruieren • Argumente mit normativen Hintergrundtheorien abgleichen
	4. Schritt für Hauptfrage 1	
Evaluation und Entscheidung	<ul style="list-style-type: none"> • Standpunkt der Moral einnehmen • Argumente beurteilen und gewichten • Einen Entscheid fällen 	
Beantwortung von Hauptfrage 2: Werden ausreichend Maßnahmen ergriffen, um alle moralisch bedenklichen außerplanmäßigen Handlungsweisen auszuschließen, die sich damit abseits vom primären Zweck durchführen lassen? Unterfrage 2a: Welche außerplanmäßigen Handlungsweisen abseits vom primären Zweck lassen sich durchführen? Unterfrage 2b: Welche davon sind moralisch bedenklich?	2. Schritt für Hauptfrage 2	
	Identifizierung außerplanmäßiger Handlungen	
	3. Schritt für Hauptfrage 2	
	Bewertung der moralischen Bedenklichkeit	
	4. Schritt für Hauptfrage 2	
Einschätzung von ergriffenen Gegenmaßnahmen bzw. Einschätzung des Potentials von möglichen Gegenmaßnahmen		
Beantwortung von Hauptfrage 3: Gefährden die Folgen und Nebenwirkungen der planmäßigen Techniknutzung die Lebensbedingungen oder die Lebensqualität gegenwärtiger oder zukünftiger Menschen?	2. Schritt für Hauptfrage 3	
	Die moralische Frage benennen	<ul style="list-style-type: none"> • Moralisch relevante Frage(n) und Konflikte identifizieren • Strittige Frage formulieren • Nichtmoralische Aspekte ausscheiden
	3. Schritt für Hauptfrage 3	
	Analyse der Argumente	<ul style="list-style-type: none"> • Argumente pro und kontra aufführen • Moralische Normen und Werte rekonstruieren • Argumente mit normativen Hintergrundtheorien abgleichen
	4. Schritt für Hauptfrage 3	
Evaluation und Entscheidung	<ul style="list-style-type: none"> • Standpunkt der Moral einnehmen • Argumente beurteilen und gewichten • Einen Entscheid fällen 	
5. Schritt		
Implementierung	<ul style="list-style-type: none"> • Möglichkeiten der Implementierung abschätzen • Maßnahmen zur erfolgreichen Implementierung ergreifen • Alternative Perspektiven ausloten 	

Abbildung 2: Human+ Bewertungsframework: Ablaufschema

6.3 PRINZIPIEN MITTLERER REICHWEITE DES HUMAN+ BEWERTUNGSFRAMEWORKS

In Abschnitt 5 „Ansätze und Frameworks in der Technikethik“ wurde ein Überblick über die Grundlagen des Principlism sowie zu einer breiten Auswahl an in der Technikethik üblichen Bewertungsframeworks gegeben. Dabei hat sich herausgestellt, dass zahlreiche der Frameworks ein Principlism-Fundament beanspruchen oder sich zumindest implizit auf Prinzipien mittlerer Reichweite beziehen.

Angesichts des dargestellten Spektrums stellt sich die Frage, an welchen Prinzipien und Dimensionen sich die Bewertung von Human+ orientieren soll. Im Rahmen der vorliegenden Arbeit wird der Bezug auf ein Principlism-Fundament beansprucht. Das bedeutet, dass im Hintergrund die Überlegung steht, dass sich aus einer „gemeinsam geteilten Moral“ zunächst eine Reihe von abstrakten prima facie gültigen Prinzipien rekonstruieren lassen, denen jedes Mitglied der moralischen Gemeinschaft zunächst zustimmen würde (vgl. Kapitel 5.2.2 Der Vier-Prinzipien-Ansatz von Beauchamp und Childress auf Seite 46 der vorliegenden Arbeit). Dabei werden die vier Prinzipien von Beauchamp und Childress in ihrer allgemeinen Formulierung übernommen und noch einmal pointiert dargestellt (unter der Bezeichnung „abstrakter Inhalt“).

Da diese Prinzipien oft für die konkrete Bewertung zu abstrakt sind, werden sie für den vorliegenden IKT-Kontext weiter spezifiziert. Anders als beim Vorgehen von Wright (2011) bzw. dem Menlo-Report (2012 und 2013) wird davon Abstand genommen, den vier Prinzipien weitere Prinzipien hinzuzufügen, z.B. ein Prinzip der Beachtung der Privatheit [privacy] oder der Beachtung von Rechtsvorschriften. Stattdessen werden diese Dimensionen unter die bereits bestehenden Prinzipien subsumiert und ihr Stellenwert im Rahmen der Spezifizierung berücksichtigt. „Spezifizierung“ wird hierbei in einem etwas loseren Sinne als wie bei Beauchamp und Childress verwendet, beinhaltet aber ebenfalls die Anreicherung einer Norm mit zusätzlichen Informationen. Anders als bei Beauchamp und Childress dient die Spezifizierung nicht so sehr dem Zweck, eine abstrakte Norm handlungsanleitend für konkrete Situationen zu machen, sondern als heuristisches Mittel zu dienen (z.B. für Ethical Matrices). Zusätzliche Orientierung bieten zudem Elemente der weiter oben im Rahmen unterschiedlicherer Frameworks vorgestellten Checklisten, welche ebenfalls als „IKT-Spezifizierung“ den Prinzipien zugeordnet werden (berücksichtigt wurden dabei Brey 2012, Wright 2011, Menlo 2013, und Aspekte aus dem DESSI-Framework). Dabei wird nicht zwangsläufig der Zuteilung aus den Originalquellen gefolgt. Zum Beispiel ordnet Wright (2011: 209) die „Nachhaltigkeit“ von Projektentwicklungen dem Prinzip des Wohltuns zu, während in nachfolgender Systematik diese eher dem Prinzip der Gerechtigkeit zugewiesen wird – weil es sich bei Human+ wie auch anderen KIRAS-Forschungsprojekten um Forschungsarbeiten mit öffentlichen Geldern handelt, und daher eine nachhaltige Entwicklung auch einen positiven Beitrag für die faire Verteilung von Lasten und Pflichten im Sinne einer auch langfristig effizienten Mittelverwendung leistet.³⁹ Dies hat zudem den Nebeneffekt, dass bei der Arbeit mit Ethical Matrices die dabei erzeugten Tabellen übersichtlicher bleiben.

³⁹ Zahlreiche Dimensionen werden in der Literatur uneinheitlich unter verschiedene Prinzipien verteilt. Nach Ansicht des Verfassers der vorliegenden Arbeit lässt es sich nicht vermeiden, dass dieselbe spezifizierte Thematik, z.B. Datenschutz, unter verschiedenen Prinzipien subsumiert werden könnte, z.B. Respekt vor der Autonomie oder auch Nichtschädigung. Im vorliegenden Fall wird Datenschutz, da er eine Voraussetzung auch für die Achtung der Privatheit ist (im Sinne von: Betroffene sollen selbst entscheiden können, welchen Teil ihrer Privatheit sie aufgeben möchten oder nicht) dem Respekt vor der Autonomie

Abschließend sei noch darauf hingewiesen, dass die nachfolgenden Prinzipien so zusammengestellt wurden, dass sie in der weiteren beruflichen Praxis des Verfassers der vorliegenden Arbeit auch für die Bewertung anderer Sicherheitsforschungsprojekte herangezogen werden können. Das bedeutet, dass nicht jedes Element der nachfolgenden Ausführungen sich auch in der ethischen Analyse von Human+ wiederfinden wird – ganz einfach aus dem Grund, weil sie für Human+ wenig Spielraum zur Anwendung gefunden haben. Zum Beispiel findet sich in der IKT-Spezifizierung des Prinzips der Fürsorge auch die Förderung von Demokratie und demokratischen Institutionen, was bei anderen Projekten eine wichtige Dimension sein kann, im Rahmen von Human+ aber „neutral“, also ohne Auswirkungen, ist.

zugewiesen. Zentral ist in jedem Fall, die Zuteilung transparent zu machen, und möglichst stringent in der darauffolgenden Analyse durchzuhalten.

Prinzip	Beschreibung und Spezifizierung
Respekt vor der Autonomie (Prinzip der Selbstbestimmung)	<p>Abstrakter Inhalt: Berücksichtigung der Autonomie anderer, d.h. Ernstnehmen der Präferenzen anderer, und Förderung anderer in ihrer Fähigkeit, autonom Entscheidungen zu treffen und Wünsche formulieren zu können (Hauptvoraussetzungen dafür: Handlungsfähigkeit und Fehlen von kontrollierenden Einflüssen)</p> <p>IKT-Spezifizierung:</p> <ul style="list-style-type: none"> • Informierte Zustimmung • Beachtung von Datenschutz und Privatheit⁴⁰ (z.B. Minimierung von Datensammlungen, Schutz von erhobenen Daten, Spezifizierung des Zwecks der Technik bzw. der Datensammlung, Transparenz zum Datenumgang und zum Technikeinsatz, Erschwerung von Cyberkriminalität) • Menschenwürde⁴¹ • Erhalt grundlegende Freiheiten (z.B. Bewegungsfreiheit, Meinungsfreiheit, Versammlungsfreiheit) • Stakeholder-Identifikation
Prinzip der Nichtschädigung	<p>Abstrakter Inhalt: Unterlassung von Handlungen, die andere in dem Sinne schädigen, dass sie ihren Interessen zuwiderlaufen, oder aber andere dem Risiko aussetzen, Schädigungen zu erleiden</p> <p>IKT-Spezifizierung:</p> <ul style="list-style-type: none"> • Zu berücksichtigende Schadensarten und Risiken: Gesundheit und körperliche Schadensarten/Risiken, Schmerz und Leiden, psychologische Schäden/Risiken, Schädigungen von Fertigkeiten und Leistungsvermögen, gesellschaftliche Schäden/Risiken, Integrität (von Informationen), Verfügbarkeit • Vertraulichkeit und Geheimhaltung • Sicherheit [safety] (der Projektentwicklungen) • Eindämmung von eingetretenen Schäden
Prinzip der Fürsorge	<p>Abstrakter Inhalt: Aktives Fördern des Wohlbefindens anderer. Beinhaltet sowohl aktive bzw. „positive“ Fürsorge als auch die Abwägung zwischen Nutzen, Risiken und Kosten um das allgemein beste Ergebnis zu erzielen.</p> <p>IKT-Spezifizierung:</p> <ul style="list-style-type: none"> • Bedienbarkeit/Anwendbarkeit der Projektentwicklungen für die AnwenderInnen • Förderung von persönlichen Aspekten wie Glück, Gesundheit, Wissen, Vertrauen, Sinnerfüllung • Förderung von sozialen Institutionen und Strukturen • Förderung von Demokratie und demokratischen Institutionen • Förderung von kultureller Diversität • Proportionalität von Risiken (Eingriffen bzw. Verletzung von anderen prima facie Prinzipien oder ihren Spezifizierungen) und Nutzen
Prinzip(ien) der Gerechtigkeit	<p>Abstrakter Inhalt: Faire, angemessene Behandlung im Lichte dessen, was einer Person geschuldet ist bzw. ihr zusteht. Gleiche Behandlung von Gleichem und ungleiche Behandlung von Ungleichem [formales Prinzip der Gerechtigkeit], ergänzt durch unterschiedliche materielle Prinzipien der Gerechtigkeit, z.B. der Verteilungsgerechtigkeit als faire Verteilung von Kosten und Risiken auf alle Betroffenen (mit je unterschiedlichen Verteilungsschlüsseln, z.B. Gleichverteilung, nach Bedarf, nach Bemühen, nach Beitrag, nach Verdienst, oder marktwirtschaftlich), bzw. noch breiter gefasst: als Verteilung von Rechten und Verantwortlichkeiten in einer Gesellschaft.</p> <p>IKT-Spezifizierung:</p> <ul style="list-style-type: none"> • Ergebnisgerechtigkeit und prozedurale bzw. Verfahrensgerechtigkeit⁴² • Verteilungsgerechtigkeit: als faire Verteilung von Gütern, Risiken und Belastungen • Nachhaltigkeit der Entwicklung • Vermeidung von Diskriminierung (durch Algorithmen) und Social Sorting⁴³ • Begründeter (und nicht willkürlicher) Einbezug von Personengruppen als Zielgruppen (Auswahl mit Bezug zum Forschungsziel) • Einhaltung von Rechtsvorschriften (Compliance) und Zurechenbarkeit von Verantwortung • Proportionalität, Angemessenheit, Notwendigkeit der angestrebten Maßnahme/Technik gegenüber dem bestehenden Problem

Tabelle 9: Human+ Bewertungsframework: verwendete Prinzipien mittlerer Reichweite

⁴⁰ Vgl. auch Charta der Grundrechte der Europäischen Union, Artikel 7: „Jede Person hat das Recht auf Achtung ihres Privat- und Familienlebens, ihrer Wohnung sowie ihrer Kommunikation“ (EU 2012, Artikel 7).

⁴¹ Nach Wright (2011: 205) besteht die Berücksichtigung der Menschenwürde darin, dass staatliche Behörden nicht die Privatsphäre von Menschen beeinträchtigen dürfen, sondern Schritte ergriffen werden, die es erlauben, dass Individuen in Würde leben können. Leben in Würde bedeutet u.a. ein Leben frei von Ausbeutung und physischer oder psychischer Miss-handlung.

⁴² Krüger (2018: 45) interpretiert dieses Prinzip im Rahmen von Überwachungstechnologien dahingehend, dass „Ergebnisgerechtigkeit“ dann vorliegt, wenn die Technik das „richtige“ Ergebnis liefert, und „verfahrensgerecht“ ist, wenn das Ergebnis nicht auf Grund einer Aneinanderreihung von Fehlern „zufällig“ richtig ist. Bei Verwendung von Algorithmen bestünde demnach „notwendigerweise“ eine unvollkommene Verfahrensgerechtigkeit, weil Fehler nie ausgeschlossen sein können, und daher die Qualität in Relation zu alternativen Möglichkeiten bewertet werden müssen.

⁴³ Diskriminierung wird auch untersagt durch Artikel 21 der Charta der Europäischen Union, welche besagt: „Diskriminierungen insbesondere wegen des Geschlechts, der Rasse, der Hautfarbe, der ethnischen oder sozialen Herkunft, der genetischen Merkmale, der Sprache, der Religion, oder der Weltanschauung, der politischen oder sonstigen Anschauung, der Zugehörigkeit zu einer nationalen Minderheit, des Vermögens, der Geburt, einer Behinderung, des Alters oder der sexuellen Ausrichtung sind verboten“ (EU 2012: Artikel 21).

Noch eine Ergänzung zum Begriff der informierten Zustimmung:

Eine forschungspraktische Möglichkeit, Individuen in ihrer Selbstbestimmungsfähigkeit ernst zu nehmen, besteht im Einholen einer „informierten Zustimmung“ [informed consent] (vgl. Wright 2011: 205f). Während das Konzept v.a. im Bereich der (direkt) menschenbezogenen Forschung in der Medizin entwickelt wurde, kann es auch in der Forschung zu Informationstechnologien angewendet werden. Die Anforderungen an eine gültige informierte Zustimmung finden sich etwa in einer entsprechenden EU-Verordnung (vgl. EU 2014 für Auflagen zum klinischen Bereich). Vorgeschrieben wird u.a., dass eine Person, die an einer Versuchsreihe teilnimmt, die Entscheidung teilzunehmen auch frei trifft, nachdem sie umfassend (in Form eines Gesprächs mit Mitgliedern des Forschungsteams) über die Art des Versuchs, der Implikationen und Risiken informiert wurde und diese Zustimmung auch entweder schriftlich oder in Gegenwart von Zeugen gegeben wird, jederzeit sich ohne Nachteile aus dem Versuch zurückziehen kann (und über dieses Recht auch informiert wurde), und eine Anlaufstelle für weitere Informationen über den Versuch genannt wurde. Im Menlo-Report wird (allerdings teilweise unter Bezug auf die US-amerikanische Gesetzeslage) darauf verwiesen, dass in Fällen, in denen das Einholen einer informierten Zustimmung dem Forschungsziel entgegenstehen würde (z.B. weil sie die Validität der Daten beeinträchtigt), Ausnahmen geltend gemacht werden können, sofern gezeigt werden kann, dass das Risiko für die Forschungssubjekte minimal ist, Schutzmaßnahmen für die Betroffenen getroffen werden, und diese im Anschluss an die Forschung nachträglich darüber informiert werden.

In den nächsten Abschnitten dieses Kapitels wird das entworfene Framework mit den erarbeiteten Prinzipien benutzt, um die Human+ Technik einer ethischen Beurteilung zu unterziehen.

6.4 ANALYSE DES IST-ZUSTANDES (SCHRITT 1)

6.4.1 Harte Fakten und Kontextsensibilität

Der Human+ Projektantrag sowie die Erfassung der Anforderungen an ein entsprechendes System nahm direkt Bezug auf die „Flüchtlingkrise 2015/2016“. Daher sollen zuerst einige relevante Hintergrundinformationen und Eckpunkt der Ereignisse des Kontexts von 2015/2016 skizziert werden.

6.4.1.1 Begriffliche Vielfalt zum Themenfeld Geflüchtete/Flüchtling/MigrantIn

Eine grundsätzliche Schwierigkeit bei der Bewertung von Human+ besteht in den oft uneinheitlichen und unklaren Begriffsverwendungen, die sich im Themenfeld von Migration und Flucht auftun. Folgende Tabelle gibt einen Überblick über Begriffe und ihre Bedeutung, wie sie häufig verwendet werden, um hier schon eingangs für eine Klärung zu sorgen.

Begriff	Bedeutung
AsylantIn	Alltagssprachliche Verwendung des bedeutungsgleichen Begriffs AsylwerberIn bzw. Asylsuchende/r, aber häufig mit negativem Beigeschmack (vgl. UNHCR o.D.: 13).
AsylwerberIn (Asylsuchende/r)	Personen, die außerhalb ihres Heimatlandes sind und um Asyl (Aufnahme und Schutz vor Verfolgung) ansuchen und deren Asylverfahren noch nicht abgeschlossen ist. Haben für die Dauer des Asylverfahrens Aufenthaltsberechtigung und sind legal in Österreich. Werden mit positivem Abschluss des Verfahrens Asylberechtigte bzw. anerkannte Flüchtlinge bzw. Konventionsflüchtlinge (vgl. UNHCR o.D.: 13).
Besonders schutzbedürftige Person	Minderjährige, unbegleitete Minderjährige, Menschen mit Behinderung, Menschen mit psychischer oder physischer Erkrankung, ältere Menschen, Schwangere, Alleinerziehende mit minderjährigen Kindern, Opfer von Menschenhandel und Personen, die Folter, Vergewaltigung oder sonstige schwere Formen psychischer, physischer oder sexueller Gewalt erlitten haben; benötigen spezielle Unterstützung und Schutz (vgl. UNHCR o.D.: 15).
Flüchtling (auch: Geflüchtete/r)	Laut Genfer Flüchtlingskonvention sind Personen Flüchtlinge, die sich aus begründeter Furcht vor Verfolgung wegen ihrer Religion, Nationalität, Zugehörigkeit zu einer bestimmten ethnischen oder sozialen Gruppe oder ihrer politischen Überzeugung außerhalb ihres Landes befinden und den Schutz dieses Landes nicht in Anspruch nehmen können oder wegen dieser Befürchtungen nicht in Anspruch nehmen wollen. Bei positivem Abschluss des Asylverfahrens bekommen sie den Status des/der Asylberechtigte/n bzw. anerkannten Flüchtlings bzw. Konventionsflüchtlings (vgl. UNHCR o.D.: 27). Refugee (mandate): „A person who qualifies for the protection of the United Nations provided by the High commissioner for Refugees (UNHCR), in accordance with UNHCR’s Statute and, notably, subsequent General Assembly’s resolutions clarifying the scope of UNHCR’s competency, regardless of whether or not he or she is in a country that is a party to the 1951 Convention or the 1967 Protocol – or a relevant regional refugee instrument – or whether or not he or she has been recognized by his or her host country as a refugee under either of these instruments” (IOM 2019: 170). Refugee (1951 Convention): “A person who, owing to a well-founded fear of persecution for reasons of race, religion, nationality, membership of a particular social group or political opinion, is outside the country of his nationality and is unable or, owing to such fear, is unwilling to avail himself of the protection of that country; or who, not having a nationality and being outside the country of his former habitual residence as a result of such events, is unable or, owing to such fear, is unwilling to return to it.” (IOM 2019: 171). Hinweis: Personen sind dann Flüchtlinge, sobald sie diese Kriterien erfüllen, d.h. notwendigerweise schon bevor sie formell als Flüchtlinge anerkannt werden. Die Anerkennung macht sie nicht zu Flüchtlingen, sondern erklärt sie als solche (vgl. IOM 2019: 171).
illegale/r MigrantIn	„Refers to migrants in an irregular situation and undocumented migrants” (IOM 2019: 102). Hinweis: von der Bezeichnung als “illegal” in Bezug auf MigrantInnen wird oft abgeraten, weil der Begriff eine kriminelle Konnotation habe, gegen die Würde der MigrantInnen sei und den Respekt gegenüber den Menschenrechten der MigrantInnen unterminierte. Wie alle Menschen können auch MigrantInnen als Personen nicht „illegal” sein (vgl. IOM 2019: 102).
irreguläre/r MigrantIn	Personen, die im Wohnland nicht über den Einwanderungsstatus verfügen, den sie benötigen. Kommen entweder illegal über die Grenze, oder mit einem gültigen Visum in das Zielland und bleiben dort über die Gültigkeitsdauer hinaus (vgl. Naujoks 2020: 35). „Movement of persons that takes place outside the laws, regulations, or international agreements governing the entry into or exit from the State of origin, transit or destination” (IOM 2019: 116).
MigrantIn	Menschen, die von einem Wohnsitz bzw. Land zu anderen dauerhaften Wohnsitzen bzw. Ländern aus unterschiedlichen Gründen wandern. Der Unterschied von MigrantInnen und Flüchtlingen besteht darin, dass MigrantInnen in ihrem Herkunftsland keine Verfolgung droht und sie jederzeit in ihr Heimatland zurückkehren können (vgl. UNHCR o.D.: 41).
refugee	siehe Flüchtling

reguläre/r MigrantIn	„A person who moves or has moved across an international border is authorized to enter or to stay in a State pursuant to the law of that State and to international agreements to which that State is a party“ (IOM 2019: 134).
Schutzsuchende/r	Begriff des Deutschen Statistischen Bundesamtes zur Quantifizierung des Bestands an AusländerInnen, die sich aus humanitären Gründen in Deutschland aufhalten (vgl. Eberle 2019: 19). Bezieht sowohl AusländerInnen mit anerkanntem Schutzstatus ein, als auch solche, die sich im laufenden Asylverfahren befinden oder mit abgelehntem Schutzstatus in Deutschland leben (d.h. 3 Gruppen: Antrag auf humanitären Schutz entweder offen, anerkannt, abgelehnt [bzw. ausgelaufen oder widerrufen]).
Unbegleitete minderjährige Flüchtlinge (UMF)	Kinder und Jugendliche (unter 18 Jahren), die ohne Eltern oder andere obsorgeberechtigte Erwachsene nach Österreich flüchten. Für sie gelten während des Asylverfahrens besondere Regelungen, z.B. Anspruch auf Rechtsvertretung oder Unterbringung in einer kindgerechten Unterkunft (vgl. UNHCR o.D.: 61).
undokumentierte/r MigrantIn	„A non-national who enters or stays in a country without the appropriate documentation“ (IOM 2019: 223). Wenn diese Personen zwar ein Identitätsdokument haben, aber kein Dokument das ihr Aufenthaltsrecht klarlegt im Land, oder dieses gefälscht und nicht länger gültig ist, dann ist der Begriff synonym mit irreguläre/r MigrantIn (vgl. IOM 2019: 223).
Wirtschaftsflüchtling	Fälschliche Bezeichnung für Wirtschaftsmigrant/in (vgl. UNHCR o.D.: 67).
WirtschaftsmigrantIn	Menschen, die aus wirtschaftlichen Gründen und üblicherweise freiwillig ihre Heimat verlassen. Können im Unterschied zu Flüchtlingen jederzeit wieder in ihre Heimat zurückkehren (vgl. UNHCR o.D.: 67).

Tabelle 10: Verwendete Begriffe im Themenfeld Flucht und Migration

Für die vorliegende Arbeit bedeutet diese Vielfalt an unterschiedlichen Termini die Schwierigkeit, dass bei der Stakeholder-Analyse keine 100% passende Bezeichnung für die Gruppe derjenigen Menschen, deren Bewegungen prognostiziert werden sollen, bereitsteht. Der Projektantrag verwendet etwa den Begriff „MigrantInnen“, allerdings ist aus dem übrigen Antrag klar zu entnehmen, dass es nicht um die Voraussage etwa von TouristInnen oder ArbeitsmigrantInnen geht, sondern man sich auf den Bereich von Flucht und zumindest irregulärer Migration bezieht. Gleichzeitig kann auch der Begriff „irreguläre/r MigrantIn“ nicht verwendet werden, weil dieser auch solche Personen inkludiert, die auf völlig regulärem Wege (etwa über ein Visum) ein Land betreten haben, dort aber solange bleiben, bis es ausläuft (und erst dann zählen sie zu den irregulären MigrantInnen). Während gemäß obigen Definitionen ein „Flüchtling“ (bzw. ein/e Geflüchtete/r) nicht erst zu einem solchen erklärt werden muss (sondern es reicht, wenn sie sich aus begründeter Furcht vor Verfolgung aus bestimmten Gründen außerhalb ihres Landes befinden), ist nicht auszuschließen, dass ein Teil der Personen, die während der „Flüchtlingskrise 2015/2016“ als Teil von medial als „Flüchtlinge“ bezeichneten Gruppen nach Österreich kamen, von diesen Gründen nicht betroffen ist. Der Begriff „Schutzsuchende“ des Deutschen Statistischen Bundesamtes wiederum bezieht sich streng genommen auf lediglich in Deutschland aufhältige Personen, die entweder um Schutzstatus ansuchen, ihn erhalten haben, oder dahingehend abgelehnt wurden, und würde somit nicht diejenigen Personen umfassen, die sich etwa außerhalb von Deutschland aufhalten, aber deren Bewegungen genauso von Human+ prognostiziert werden sollen. Daher muss an dieser Stelle eine Setzung vorgenommen werden. Begrifflich wird in weiterer Folge als Bezeichnung für die Gruppe derjenigen, deren Bewegungen prognostiziert werden sollen, der Begriff

„Geflüchtete/r“ verwendet, trotz obiger Unschärfen. Warum nicht „Flüchtling“? Die Deutsche Gesellschaft für Sprache e.V. hat den Terminus „Flüchtling“ angesichts der zentralen Themenlage im Jahr 2015 zum Wort des Jahres 2015 gekürt, verweist aber auch darauf, dass es durch das Ableitungssuffix „-ling“ für „sprachensible Ohren“ auch tendenziell abschätzig klingen kann. Der Grund hierfür ist, dass analoge Bildungen wie Eindringling oder Emporkömmling negativ konnotiert sind, und andere wie z.B. Prüfling eine passive Konnotation hätten (vgl. GFDS 2015). Alternative Möglichkeiten wie „Geflüchtete/r“ haben nicht diese Konnotation, und werden daher für die meisten Zwecke dieser Arbeit bevorzugt. In ihrer neutralen Bedeutung sollen beide Begriffe aber als deckungsgleich betrachtet werden. Wie in einer Fußnote zur Einleitung bereits vorausgeschickt, wird weiterhin dann der Begriff „Flüchtling“ verwendet, wenn der Begriff Teil eines Terminus Technicus ist, z.B. hat sich während des Projekts gezeigt, dass auch innerhalb der bei der Bewältigung der Einsatzlage von 2015/2016 beteiligten BOS häufig von „Flüchtlingslage“ die Rede war. Ähnlich wird zwecks einfacheren Verständnisses verfahren mit „Flüchtlingslager“, „Flüchtlingsbewegungen“, „Flüchtlingskrise“.

Zudem ist darauf hinzuweisen, dass bei den nachfolgenden Ausführungen zur Kontextsensibilität dann der Begriff Geflüchtete/r verwendet wird, wenn im Original auch von Geflüchteten oder Flüchtlingen die Rede ist. Wenn in einer Publikation dagegen allgemeiner von „MigrantIn“ die Rede ist, dann wird jener Begriff beibehalten.

6.4.1.2 Historischer Kontext – „Flüchtlingskrise 2015/2016“ als Initiator der Human+ Forschung

Kurz soll an dieser Stelle auf die Ereignisse um 2015/2016 eingegangen werden, um den historischen Kontext für die Human+ Initiierung zu skizzieren. Dabei wird kein Anspruch auf chronologische Vollständigkeit erhoben – vielmehr steht im Vordergrund, dass es bereits Vorzeichen für größere Wanderbewegungen und eine Zunahme von Geflüchteten außerhalb von Europa gegeben hatte, bevor diese dann zeitversetzt auch Europa erreicht hatten. Denn: Während 2014 die Anzahl an weltweiten Geflüchteten, bei allen Schwierigkeiten einer Quantifizierung, ein Höchstmaß erreichte, wurde in Europa erst in der zweiten Hälfte des Jahres 2015 ein großer Zustrom bemerkbar, welcher sich in hunderttausenden Geflüchteten manifestierte, die jedes Jahr versuchten, das Mittelmeer auf einer gefährlichen Reise zu durchqueren, wobei die meisten Geflüchteten aus der Syrischen Arabischen Republik, Afghanistan und Irak kamen (vgl. für diese Ausführungen Vrgovic und Kapor-Stanulovic 2019: 12ff). Verglichen mit 2014 (in welchem mehr als 200.000 Ankömmlinge über das Meer verzeichnet wurden) waren es im Jahr 2015 über eine Million. Die häufigste Route verlief aus der Türkei auf die benachbarten griechischen Inseln, bis im März 2016 ein Abkommen zwischen der EU und der Türkei geschlossen wurde, woraufhin Italien Griechenland als das Land mit den häufigsten Ankömmlingen (wieder) ablöste. Während nicht jede der irregulär in Europa eingereisten Personen auch Geflüchtete im engeren

Sinn waren (nämlich weil sie zwar Asyl suchten, aber nicht aus Regionen kamen, die offiziell als besonders gefährlich eingestuft wurden), lasse sich nach Vrgovic und Kapor-Stanulovic die Kategorie „MigrantInnen“ insgesamt nur schwer in Gruppen wie „wirtschaftliche“ und „ums Überleben kämpfende“ MigrantInnen einteilen, weil es hier häufig Überschneidungen gäbe. Stattdessen sprechen sie von einem Kontinuum unterschiedlichster Bedingungen in unterschiedlichen Ländern und Regionen, und jede/r MigrantIn habe eine eigene Position auf diesem Kontinuum. Während die einen dem Militärdienst, politischer Verfolgung, religiöser Unterdrückung, ethnischer Gewalt, Bürgerkrieg, Menschenhandel oder Organhandel zu entgehen versuchen, geht es bei den anderen um Armut und Hunger, während wiederum andere, so die AutorInnen, sich einfach anderswo eine „bessere Existenz“ erhoffen.

Die österreichischen und deutschen Infrastrukturen und Behörden waren jedenfalls besonders in der Anfangsphase der Ereignisse auf die Anzahl an Ankommenden nicht vorbereitet. So war beispielsweise im Juli 2015 das Flüchtlingslager bzw. die Betreuungsstelle Traiskirchen, welche auf rund 1.800 fixe Plätze und weitere 480 Zeltplätze ausgelegt war, mit über 4.500 Asylsuchenden weit überbelegt (vgl. UNHCR 2015). Die weiter wachsende Zahl an Ankömmlingen führte dazu, dass am 5. August 2015 keine weiteren Geflüchtete in Traiskirchen aufgenommen wurden, wobei die in den Ländern laufend neu geschaffenen Quartiere nicht mit dem steigenden Bedarf Schritt halten konnten (vgl. DerStandard 2015a). Am Freitag, dem 11. September 2015, als der Grenzübergang Nickelsdorf nach Ungarn gesperrt wurde, nachdem gemeinsam mit dem Donnerstag davor mehr als 11.000 Geflüchtete über den Grenzübergang nach Österreich gekommen waren, beklagte der damalige burgenländische Landespolizeidirektor Doskozil die Informationspolitik der ungarischen Behörden, wonach man nicht ausreichend früh informiert worden wäre, wieviele Geflüchtete in der Nacht auf Freitag zu erwarten gewesen waren (vgl. DerStandard 2015b). Das Hauptziel der meisten Geflüchteten war Deutschland, und auf ihrem Weg dorthin, so ein österreichischer Sicherheitsbeamter in einem Interview, würden die Geflüchteten „in keinem Land“ registriert werden, sondern „durchgewinkt“, also in Empfang genommen und ans Nachbarland weitergegeben werden (vgl. DerStandard 2015c).

Es ist diesen Ereignissen geschuldet, dass gegenüber dem KIRAS Ausschreibungsleitfaden von 2015 (vgl. KIRAS 2015), in welchem das Wort „Migration“ kein einziges Mal vorkam, in der darauf folgenden Version Migration als inhaltlicher Schwerpunkt für eine bilaterale Österreich-Deutschland-Kooperation aufgenommen wurde (vgl. KIRAS 2016: 15).

6.4.1.3 Objektiver Bedarf an Human+ und Einschätzungen von BOS

Der eigene Migrations-bezogene Themenschwerpunkt für das österreichische Förderungsprogramm für Sicherheitsforschung KIRAS im Anschluss an die Ereignisse von 2015/2016 illustriert den wahrgenommenen Bedarf nach Forschung in diesem Bereich auf einer strategischen, bundesweiten Ebene.

Innerhalb des in diesem Rahmenprogramm durchgeführten Human+ Projekts wurden umfangreiche Tätigkeiten im Arbeitspaket „Bedarfserfassung“ unternommen, um fachliche Anforderungen an ein wie in der Einleitung skizziertes „Informations- und Darstellungssystem“ zu eruieren. Dabei wurden unterschiedlichste Methoden der qualitativen und quantitativen Sozialforschung angewandt, z.B. qualitative Leitfadeninterviews oder quantitative Onlinebefragungen mit Personen, die in der Bewältigung der „Flüchtlingslage 2015/2016“ in unterschiedliche Aufgabengebieten⁴⁴ in unterschiedlichsten staatlichen und nicht-staatlichen Organisationen eingebunden waren.⁴⁵ Dies umfasste z.B. auf österreichischer Seite Personen aus den Bereichen Rettungsdienste (z.B. Rotes Kreuz und Johanniter), Feuerwehr, Polizei/BMI, BMLV, NGOs, und dergleichen, insbesondere MitarbeiterInnen in Krisenstäben unterschiedlichster Ebenen. Ergänzt wurden diese Ansätze mit nichtreaktiven Methoden, z.B. der Analyse von, soweit zugänglich, organisationsinternen Einsatztagebüchern. Insgesamt wurde auf eine breite Streuung sowohl auf österreichischer wie deutscher Seite geachtet. Die Erhebungen wurden 2018 durchgeführt, was den Effekt hatte, dass zahlreiche Details zu den Tätigkeiten und Ereignissen von 2015/2016 den Befragten nicht mehr vollständig in Erinnerung waren, das Thema der „Flüchtlingslage 2015/2016“ aber, wie sich zeigte, von den RespondentInnen in den Gesprächen nach wie vor als wichtig eingestuft wurde. Neben den Erhebungen der fachlichen Anforderungen (die hier im Detail nicht weiter im Vordergrund stehen) wurden allerdings auch eine Reihe von kontextbezogenen Nebenaspekten beleuchtet, z.B. wie sich die Kommunikation unter Einsatzkräften und beteiligten Organisationen in dieser Krisenphase von eingespielten Abläufen unterschied, und was geholfen hätte, die Arbeit in diesem Bereich besser zu erfüllen. Folgende Aspekte scheinen für die Bewertung von Human+ in der vorliegenden Arbeit relevant (vgl. für nachfolgende Aspekte Human+ 2018a)⁴⁶:

- Die „Chaosphase“ (d.h. die Frühphase einer Krise, in welcher sich beteiligte AkteurInnen noch nicht darauf eingestellt haben und noch kein „Lage-Alltag“ entwickelt wurde) brachte zahlreiche Schwierigkeiten, z.B. notwendige Neuordnungen von Meldewegen, wenn von den althergebrachten auf Grund von nun notwendiger, größerer interorganisationaler Zusammenarbeit abgewichen werden musste, oder falls diese nicht allen Beteiligten bekannt waren (vgl. Human+ 2018a: 3).

⁴⁴ z.B. Organisation/Koordination (strategischer Ebene), Logistik und Einsatzorganisation (taktischen Ebene), oder die Umsetzung von humanitärer Sicherheit (operative Ebene).

⁴⁵ Die Identifizierung der bei einer Flüchtlingslage relevanten Beteiligten war im Human+ Projekt Gegenstand eines eigenen Tasks im Arbeitspaket „Bedarfserfassung“, und mündete in einen „Katalog der bei einer Flüchtlingslage relevanten Bedarfsträger“ (vgl. Human+ 2018b)

⁴⁶ Polizeiliche Bedarfe (aus deutscher Sicht) wurden in einem eigenen Deliverable erfasst (vgl. Human+ 2018c). Der Bericht enthält keine Angaben zu den Informationen, die während der Flüchtlingslage gefehlt hatten – formuliert aber Informationsbedarfe, die weitgehend den aus Human+ 2018a bekannten, während 2015/2016 episodenhaft fehlenden Informationen entsprechen – ergänzt um stärker sicherheitspolizeilich geprägte Informationen, wie z.B. zu Gewaltpotential und gefährdungsrelevanten Sachverhalten (vgl. Human+ 2018c: 15).

- Die beteiligten Organisationen hatten insgesamt häufig mit fehlenden Informationen zu kämpfen. So gaben 54% der Auskunftspersonen auf deutscher Seite (Human+ 2018a: 9) und 73% auf österreichischer Seite (Human+ 2018a: 28) an, dass benötigte Informationen nicht zur Verfügung gestanden hätten.⁴⁷ Wesentliche Informationen, die demnach (zumindest episodenhaft)⁴⁸ gefehlt hatten umfassen u.a.:
 - Angaben zur Anzahl von einreisenden Geflüchteten und der von ihnen verwendeten Transportmitteln, und zahlenmäßige Einschätzungen zur Flüchtlingsbewegung insgesamt, welche als Entscheidungsgrundlage und für die bessere Orientierung und Planung wichtig gewesen wären
 - Zeitpunkte der Ankunft an unterschiedlichen Orten
 - Einschätzungen zur Anzahl von Menschen an bestimmten Orten
 - Informationen zu den Ankommenden selbst (Alleinreisende? Familien? Minderjährige? Nationalitäten und Kulturen? Gesundheitlicher Zustand vor der Ankunft? Grad der Hilfsbedürftigkeit oder Vorliegen besonderer Hilfsbedürftigkeit wie im Falle von Schwangeren, Erkrankten, Menschen mit Behinderung, Säuglingen?)
 - Versorgungszustand der Ankommenden, zur Eruierung der Notwendigkeit der Bereitstellung von Kleidung, Nahrung, Hygieneartikeln, usw.
 - Informationen zu Geschlecht, Alter, Nationalität/Herkunft/Ethnie, Religion, Sprache und Sprachkenntnissen
 - Zuständigkeiten und Kontaktmöglichkeiten von involvierten Personen und Organisationen
 - Informationen/Überblick zu Einsatzorten und Aufträgen anderer Organisationen
 - Informationen/Datensätze zu verfügbaren Ressourcen (Aufnahmekapazitäten von Ländern/Warteräumen, verfügbare Container und Feldbetten, etc.) und dazu, ob diese Datensätze verifiziert wurden/verifizierbar sind
 - Informationen zur Situation an Grenzen bzw. Grenzübertritten (Slowenien – Österreich, Italien – Österreich)
 - bessere Absprache und (behördliche) Informationsweitergabe aus Nachbarländern (v.a. Slowenien und Ungarn)
 - Informationen zum Flotten- bzw. Transportmanagement und zu den Standorten der Transportmittel

⁴⁷ Aufgrund des Erhebungsmodus und der kleinen Fallzahlen sind diese Angaben nicht repräsentativ – aber es handelt sich dabei um eine Erhebung unter einem ExpertInnenpublikum, weshalb den Zahlen dennoch ein starker Hinweischarakter zugeschrieben werden sollte.

⁴⁸ „Episodenhaft“ deshalb, weil die befragten Personen oft nicht während der ganzen Dauer der Flüchtlingslage in deren Bewältigung eingebunden waren.

- politisch eine höhere Transparenz bezüglich Flüchtlingsbewegungen und Strategien der Bewältigung (z.B. durch eine gemeinsam befüllte Plattform mit aktuellen Zahlen)
- Bestandteil der Bedarfserfassung war auch die (allerdings sehr knapp ausfallende) Erhebung von „praktischen und ethischen Bedenken“. Folgende Aspekte wurden von den RespondentInnen genannt (vgl. Human+ 2018a: 17 und 34):
 - Verwendung des Systems nach den Grundsätzen der Einsatzorganisationen⁴⁹ solle oberstes Prinzip sein (umfasst z.B. Neutralität, Vermeidung von Diskriminierung und politischer Involvierung). Zu vermeiden seien außerdem die Weitergabe und Verwendung medizinischer Daten wie Patientendaten, racial profiling, und dass sich rein subjektive Meinungen in einem solchen System widerspiegeln.
 - In einem der durchgeführten Interviews gab ein Interviewpartner zudem zu verstehen, dass die verschiedenen Arten von BOS sich unterschiedlichen übergeordneten Zielsetzungen verpflichtet sehen würden: während etwa für klassische Rettungsorganisationen wie das Rote Kreuz die Hilfe ohne Ansehung von Nationalität und anderen persönlichen Kriterien im Vordergrund stünde, würde die Polizei Sicherheit als obersten Wert hochhalten, die ÖBB wiederum (die ebenfalls in das Krisenmanagement einbezogen waren, weil sie den Weitertransport in Richtung Deutschland organisiert hatten) hätten als obersten Wert Pünktlichkeit.
 - Volles Umfangtragen von Daten- und Geheimschutzaufgaben notwendig (v.a. weil in Einsatzorganisationen häufig nicht nur Haupt- sondern auch Ehrenamtliche MitarbeiterInnen beschäftigt sind, was im Zugang zu Informationen, dem Rechte- und Rollenkonzept berücksichtigt werden muss)
 - Entsprechung von IT-Sicherheitsstandards und Bundesamt-für-Sicherheit-und-Informationstechnik-Regularien (da unerlaubte Zugriffe auf das skizzierte System als Risiko für die Sicherheit sowohl von Geflüchteten wie auch Einsatzkräften betrachtet wurden, und eine vertrauenswürdige Sicherheitszertifizierung auch die Bereitschaft zur Nutzung des Systems steigern würde)
 - Anwendung der hypothetischen Technik auf der richtigen Ebene (d.h. Informationsversorgung der unteren Ebenen durch obere Ebenen, damit an unteren Ebenen kein Ressourcenabzug durch Bedienung eines solch umfangreichen Systems entsteht)

⁴⁹ Dies wurde v.a. von MitarbeiterInnen von Rettungsdiensten genannt. Aus dem Bericht zu polizeilichen Anforderungen wurden keine ethischen, lediglich praktische, Bedenken und Anmerkungen geäußert. Es findet sich aber der Hinweis, dass ein gemeinsames Lagebild nicht politisch bewertet werden solle, sondern als das, was „es ist“, nämlich (als Tool) zur Information der „entsprechenden Stellen“ (wie der Polizei), damit diese die damit verbundenen oder bevorstehenden Aufgaben zielgerichtet und unchaotisch bewältigen können. Die Person befürchtet zudem, dass mit solchen Lagebildern vorschnell Ressentiments des Erstellers oder Empfängers gegenüber Migranten assoziiert werden, was aber nicht der Fall sei (vgl. Human+ 2018c: 21f).

- Wengleich auf den ersten Blick zahlreiche diese Aspekte Großteils nicht ethische Bedenken im moralphilosophischen Sinne darstellen, zeigen sich neben manifesteren Nennungen wie Neutralität auch indirekt eine Reihe von „moralischen Werten“, wie etwa die Wichtigkeit der Einhaltung von Datenschutz und Sicherheitsstandards.

An dieser Stelle sei noch erwähnt, dass diese Innenperspektive von BOS sich nicht notwendigerweise mit der Außenwahrnehmung, auch von ExpertInnen, deckt. So hat der Verfasser der vorliegenden Arbeit im Zuge seiner Mitarbeit bei Human+ auch ein ExpertInnen-Interview mit einem Forscher vom Institut für Technikfolgenabschätzung in Wien durchgeführt, um einen Blick von außen auf das Human+ Projekt zu erhalten. Neben kritischen Einschätzungen was v.a. den Modus und das Ausmaß der Datensammlung im Social-Media-Bereich betraf (und welche zumindest nach der für das Interview möglichen Kurzcharakterisierung des Projekts als „unverhältnismäßig“ beurteilt wurde), wurde auch ersichtlich, dass die Bewältigung der Ereignisse von 2015/2016 als recht erfolgreich wahrgenommen wurde, denn: „Die Einschätzung, dass die Flüchtlingswelle von 2015 in Österreich sehr gut bewältigt wurde, ist eigentlich konsensual“ (Human+ 2019b: 3). Kommunikationsketten und Krisenmanagement hätten gut funktioniert, und Quartiere wären schnell aufgeboten gewesen, so der Forscher. Dies war nach Auskunft der RespondentInnen aus der Bedarfserhebung aber erst nach Bewältigung der Chaosphase der Fall.

6.4.1.4 Einsatzkontext von Human+ - Migrationsmanagement und Fehlen einer konkreten AnwenderInnen-Gruppe

Da sich der Projektantrag von Human+ sehr allgemein auf die Verbesserung der Informationslage für in der Bewältigung von Flüchtlingslagen tätigen BOS und Einsatzkräfte bezog, wurde darin keine Behörde oder Organisation im Besonderen als spezielle AnwenderInnen-Gruppe definiert (was angesichts des TRL 4, welches sich noch in einer deutlichen Entfernung von einem marktreifen Produkt befindet, nicht überrascht). Diese Vagheit macht es aber notwendig, im Sinne einer Kontextsensibilität sich mit dem allgemeineren Human+ Anwendungskontext auseinanderzusetzen: Migrationsmanagement (welches unter der Bezeichnung „Migrationsmanagement und Integration“ ein Themenschwerpunkt im KIRAS Ausschreibungsleitfaden von 2016 war, neben „Migrationsbewegungen und Lagebild-darstellung“ sowie „Sichere Identifizierung/Registrierung“, vgl. KIRAS 2016: 15).

Migrationsmanagement lässt sich laut International Organization for Migration (IOM) folgendermaßen definieren:

„The management and implementation of the whole set of activities primarily by States within national systems or through bilateral and multilateral cooperation, concerning all aspects of migration and the mainstreaming of migration considerations into public policies. The term refers to planned approaches to the implementation and operationalization of policy, legislative and administrative frameworks, developed by the institutions in charge of migration“ (IOM 2019: 139).

Außerdem wird festgestellt, dass sich das Konzept auf die öffentliche Administration im Bereich Migration bezieht, und somit auch auf das Management und die Implementierung von Gesetzen und Regularien sowie Entscheidungen der Regierung und die Bereitstellung von damit verknüpften öffentlichen Dienstleistungen bezieht.

Den Grundgedanken bringt Bimal Ghosh, der als Berater der IOM den Begriff geprägt hat, folgendermaßen auf den Punkt: „Central to managed migration is the establishment of a regime that is capable of ensuring that movement of people becomes more orderly, predictable and productive, and thus more manageable“ (Ghosh 2007: 107, zitiert nach de Jong et al 2017: 5). Während Migrationsmanagement, so der Anspruch derjenigen, die sich positiv für das Konzept starkmachen, für „alle Beteiligten“ Vorteile haben soll (z.B. durch Auswahl derjenigen Zuwanderung, die einem Industriestaat besonders nützt), besteht die Kehrseite in der verstärkten Rückführung (irregulärer) MigrantInnen bzw. abgewiesenen Geflüchteten, sodass der Ausschluss bestimmter Personengruppen einkalkuliert wird (vgl. de Jong et al 2017: 4). So lautet ein Vorwurf aus dem Bereich der „Kritischen Migrationsforschung“, dass sich Migrationsmanagement zwar als objektives Konzept präsentiert, tatsächlich aber voller ideologischer und institutioneller Voraussetzungen sei, z.B. die Kategorisierung von MigrantInnen nach volkswirtschaftlichem Nutzen, die Aufrechterhaltung strikter Grenzregime und die Aushöhlung eines damit vermeintlich nicht kompatiblen Asylsystems (vgl. de Jong et al 2017: 5).⁵⁰ Jong et al verweisen nicht nur darauf, dass in der österreichischen Migrations- und Integrations-Politik seit den 1990er Jahren Migrationsmanagement sich fest verwurzelt hat, sondern auch auf die Ausweitung des Konzepts auf die Fluchtmigration (vgl. de Jong et al 2017: 11f). Ein Beispiel davon gebe die Einladung der österreichischen Regierung an VertreterInnen ausgewählter Balkanstaaten im Jahr 2016 zur Konferenz „Managing Migration Together“, deren Ziel ein effizientes Migrationsmanagement sei, und letztendlich zur Schließung der „Balkanroute“ für Geflüchtete nach Europa geführt hatte.

Während in der vorliegenden Arbeit keine tiefergehende Analyse von Migrationsmanagement gegeben werden kann, sollen diese Ausführungen dafür sensibilisieren, dass im Kontext des Managements von Migration ein häufiger Fokus auf Nutzenkalküle gelegt wird, welcher sich mittlerweile auch in den

⁵⁰ Als ein weiterer Nebeneffekt könnte außerdem, aufgrund des Zuschnitts nach volkswirtschaftlicher Verwertbarkeit, die „Verdinglichung“ der MigrantInnen genannt werden.

Fluchtkontext fortgesetzt hat, worauf in der Beantwortung der Hauptfrage nach der Legitimität des primären Zwecks von Human+ noch zurückzukommen ist.

6.4.2 Stakeholder-Identifizierung

Unter Stakeholder bzw. Stakeholderin (sinngemäß auch oft übersetzt als „gesellschaftliche Akteure“) werden InteressensvertreterInnen (von Gruppen) verstanden, die von der Entwicklung oder Nutzung einer Technik betroffen sein können, wobei der Begriff üblicherweise eine Art von Organisation (oder besser: Organisiertheit) impliziert, um Interessen überhaupt artikulieren zu können (z.B. Bürgervertretungen für die Anliegen von BürgerInnen; vgl. Grunwald 2010: 94). In der vorliegenden Arbeit werden als Stakeholder auch Gruppen von Betroffenen inkludiert, die nicht als kollektiver Akteur auftreten oder auftreten können, da in der vorliegenden Arbeit kein diskursiver bzw. partizipativer Ansatz verfolgt wird und dies sonst zum Ausschluss von „unorganisierten“ Betroffenen aus den Erwägungen führen könnte. Dementsprechend wird später der Ansatz von Bleisch und Huppenbauer verfolgt, diese StakeholderInnen „fiktiv“ zu befragen, um mögliche von ihnen vorgebrachte Argumente zu sammeln (vgl. Bleisch und Huppenbauer 2014: 46f).

Wer in ethischen Untersuchungen als StakeholderIn inkludiert wird, der/die wird auch als moralisch zu berücksichtigen definiert (Mepham et al 2006: 9). Gleichzeitig wird bei vielen Problemlagen die Frage aufgeworfen, an welches Kriterium sich die Berücksichtigung als Stakeholder knüpfen soll. Bei der Analyse des Ist-Zustandes moralischer Probleme stellen sich zwei Fragen (vgl. Bleisch und Huppenbauer 2014: 26): Wer meldet Interessen an? Welche Interessen werden angemeldet? Gemäß dieser Charakterisierung wäre die Berücksichtigung als StakeholderIn an die Möglichkeit, Interessen zu haben, geknüpft, und daran wiederum hängt der sogenannte „**moralische Status**“, und die Folgefrage, wozu dieser Status im Detail verpflichtet. Die skizzierte Human+ Technik betrifft in ihren Auswirkungen in erster Linie Menschen, sodass Problemlagen wie die Berücksichtigung der Interessen von nicht-menschlichen Lebewesen (wie sie mitunter bei ökologischen Projekten zentral werden können) hier nicht auftreten. Für den Kontext der vorliegenden Arbeit wird jedenfalls davon ausgegangen, dass mit dem Zuschreiben von moralischem Status an betroffene Gruppen auch der Umstand einhergeht, dass die formulierten Prinzipien mittlerer Reichweite auf diese Gruppen anzuwenden sind.

Mitunter finden sich in der Literatur Differenzierungen für Kategorien von Stakeholdern, z.B. in direkte Stakeholder („aktive AnwenderInnen“) und indirekte Stakeholder („passiv Betroffene“; vgl. Friedman et al 2006: 356 und 361). Eine andere Einteilung wäre nach primären Stakeholdern (diejenigen, die letztendlich betroffen sind, im IKT-Kontext z.B. AnwenderInnen der Technik oder diejenigen, die Informationen von Informationssystemen verwenden), sekundäre Stakeholder (Vermittler bzw. Zwischenglieder in der Vermittlung von Vorteilen oder Nachteilen, z.B. zwischengeschaltete Service-Provider),

und Schlüsselstakeholder (als diejenigen, die maßgeblich den Erfolg eines Projekts beeinflussen können, z.B. Forschungsteam, oder auch Kriminelle, die das System angreifen könnten; vgl. Menlo 2013: 10f). Entsprechend dem letzten Beispiel ließe sich auch noch zwischen wohlwollend eingestellten Stakeholder und negativ eingestellten Stakeholder unterscheiden.

Eine Liste von StakeholderInnen könnte bei der Berücksichtigung von ausreichend weitverzweigten Sekundärfolgen der Technik beliebig lange fortgesetzt werden. Wie weiter oben bei den Ausführungen zur TA bereits klargestellt, muss aber aus praktischen Überlegungen eine begründete Auswahl getroffen werden. Folgende Stakeholdergruppen müssen nach Ansicht des Verfassers der vorliegenden Arbeit berücksichtigt werden für eine ethische Analyse von Human+, da sie recht unmittelbar entweder mit den Auswirkungen der Human+ Technik oder dem Forschungsprojekt in Zusammenhang stehen:

Stakeholder-Gruppe	Beschreibung und Rechtfertigung für den Einbezug
Geflüchtete	Gruppen oder Individuen, deren Bewegungen über die Human+ Technik prognostiziert werden sollen (vgl. Abschnitt 6.4.1.1 auf Seite 99 zu den definitiven Problemen)
Zivilbevölkerung	Allgemeine Zivilbevölkerung in Österreich und Deutschland
In-Situ-Betroffene	BewohnerInnen von und unbeteiligt aufhältige Personen in Gebieten, die von den In-Situ/-Fernerkundungs-Komponenten von Human+ betroffen sein können
Social-Media-UserInnen	Personen, die Social Media verwenden und zu fluchtbezogenen Themen posten (beispielsweise einen Lokalausweis im Falle von konkreten Flüchtlingsbewegungen). Diese sind von Human+ passiv Betroffene, da sie durch ihre Social-Media-Aktivitäten die Grundlage für eine der Datenquellen von Human+ liefern.
NGOs und NGO-MitarbeiterInnen	Dies bezieht sich auf diejenigen NGOs und ihre MitarbeiterInnen, die beim Auftreten akuter Flüchtlingslagen humanitäre Hilfe bei der Versorgung und Unterstützung leisten.
BOS-MitarbeiterInnen	Österreichische oder deutsche hauptamtliche oder ehrenamtliche MitarbeiterInnen von Behörden oder Einsatzorganisationen mit Sicherheitsaufgaben, z.B. von Polizei, Ministerienstellen mit Sicherheitsaufgaben, Zoll, aber auch Feuerwehren und Rettungsdiensten. Hier gemeint als passiv von der Human+ Technik Betroffene, die beim Erfüllen ihres Auftrags durch die von Human+ generierten Migrationsprognosen unterstützt werden sollen.
AnwenderInnen von Human+	Ebenfalls BOS-MitarbeiterInnen, hier aber in ihrer spezifischen Rolle als aktive AnwenderInnen der Human+ Technik betrachtet.
Human Sensors – OUH	Organisierte ungebundene Helfergruppen (OUH), die als eine Untergruppe der sogenannten Human Sensors die entsprechende Meldemaske im Human+ System bedienen können, um Informationen in das System einzuspeisen. Weisen allerdings keine verlässliche Organisationsstruktur auf, und sind unter Geheimschutzaspekten als Öffentlichkeit zu rechnen (vgl. Human+ 2018a: 17), können daher keine „richtige“ Zugangsberechtigung zu Human+ bekommen.
Human Sensors – BOS	MitarbeiterInnen von BOS in ihrer Funktion als eine Untergruppe der Human Sensors, die ebenfalls die entsprechende Meldemaske im Human+ System bedienen können, um Informationen in das System einzuspeisen.

Tabelle 11: Stakeholder der Human+ Technik

Die Stakeholderanalyse beendet den ersten Schritt gemäß Human+ Bewertungsframework. In den nächsten Abschnitten werden die Hauptfrage 1 bis 3 beantwortet.

6.5 HAUPTFRAGE 1 (PRIMÄRER ZWECK) – SCHRITT 2 BIS 4

6.5.1 Unmittelbarer Zweck – Argumente pro und kontra

Hauptfrage 1: Sind die Handlungsmöglichkeiten, die durch diese Technik eröffnet oder verbessert werden, an sich ethisch legitim? (primärer Zweck)

Nach Ansicht des Verfassers der vorliegenden Arbeit muss für die Beantwortung dieser Frage zunächst zwischen einem unmittelbaren und einem mittelbaren primären Verwendungszweck differenziert werden (was sich so bei Fenner 2010 noch nicht findet). Der Grund hierfür zeigt sich im Abschnitt zur Evaluation der Pro-/Kontra-Argumente zum unmittelbaren Zweck (vgl. Abschnitt 6.5.2).

Unter **unmittelbarem Verwendungszweck** der Human+ Technik wird das verstanden, was die zu entwickelnde Technik „an sich“ leisten soll. Diese lassen sich zunächst aus dem Projektantrag erschließen:

„Der beschriebene Projektansatz zielt auf die Nutzung **sozialer Medien**, sowie deren Integration mit etablierten Methoden der **in-situ Bildverarbeitung** und der **Fernerkundung** zur Erstellung eines **integrativen Echtzeit-Lagebildes** und einer **kurzfristigen Prognose** ab“ (Human+ 2016: 1; Hervorhebungen wie im Original).

Die Erstellung eines integrativen Echtzeit-Lagebildes und die Erstellung von Migrationsprognosen sind nach Ansicht des Verfassers der vorliegenden Arbeit somit der unmittelbare Verwendungszweck der Human+ Technik. Welche Pro- und Kontra-Argumente für die Legitimität dieses unmittelbaren Zwecks lassen sich anführen?

Pro-Argumente:

- **Argument des Bedarfs:** Wie weiter oben dargestellt, ergaben Erhebungen im Arbeitspaket der Bedarfsanalyse, dass während der Bewältigung der „Flüchtlingslage 2015/2016“ zumindest episodenhaft wichtige Informationen (z.B.: Wann kommen wieviele Geflüchtete mit welchen Transportmitteln wo an?) gefehlt haben, zu deren Bereitstellung die Human+ Technik einen Beitrag leisten soll. Zudem haben die Befragung und die Fokusgruppen, die mit potentiellen AnwenderInnen nach einer Erprobung des Demonstrators stattgefunden hatten, in AP5 des Projekts ebenfalls mehrheitlich ergeben, dass im entwickelten (und noch weiterzuentwickelnden) System prinzipiell ein Mehrwert für die Arbeit von BOS gesehen wird (vgl. Human+ 2019a). Während die Human+ Technik ein TRL4 anvisierte, lässt sich daher zumindest probabilistisch annehmen, dass die Technik bei entsprechender Weiterentwicklung hierzu einen praktischen Beitrag leisten kann. Falls das zutrifft, wäre auch die andere der zwei absolut zu erfüllenden Anforderungen gemäß des DESSI-Frameworks erfüllt: nämlich dass Investitionen in Sicherheitstechniken überhaupt einen Beitrag leisten

(vgl. Abschnitt 5.9 ab Seite 79). Das andere Erfordernis, nämlich dass keine Grundrechte verletzt werden, wird im Abschnitt zur Beantwortung der Hauptfrage 3 behandelt.

Kontra-Argumente:

- **Argument der Multifunktionalität und fehlenden Zweckbindung:** Prognosen zu Flüchtlingsbewegungen, ein besserer Informationsaustausch unter (selektiven) BOS und ein gemeinsames Lagebild können multifunktional für unterschiedliche weitere Zwecke verwendet werden. Zumindest ein Teil der ermöglichten Handlungsoptionen könnte, z.B. in Abhängigkeit vom politischen Klima, somit prinzipiell auch dafür verwendet werden, etwa Grenzen rechtzeitig zwecks pauschaler Abweisung der Geflüchteten zu schließen, oder sogar Menschenrechtsverletzungen gezielter zu ermöglichen.
- **Argument der unklaren Anwendungsbedingungen:** Das Anwendungsspektrum der Human+ Technik bezieht sich auch auf das „Monitoring“ internationaler Krisenregionen. Unterbestimmt ist aber, was eine Region zu einer internationalen Krisenregion macht, und sich somit für die Erfassung durch die Human+ Technik qualifiziert. Aus Sicht der projektinternen Bedarfsträger kann dies z.B. Kriegsgebiete oder auch Gebiete mit Klimakatastrophen umfassen. Eine abschließende, transparente Klärung der Anwendungsbedingungen war während der Projektlaufzeit nicht möglich, da dies komplexe Abstimmungsprozesse zwischen ExpertInnen und zuständigen Ministerien (z.B. dem Deutschen Auswärtigen Amt) bzw. internationalen Organisationen wie UN-Institutionen zu einer noch nicht weit entwickelten Technik erfordert hätte – was dies letztendlich nicht nur zu einer Frage objektiver Kriterien, sondern auch von politischen Absichten macht.
- **Argument des unsicheren Erfolgs:** Die Kehrseite des weiter oben angeführten lediglich probabilistischen Erfolgs der Human+ (Weiter-)Entwicklungen hin zu einem voll einsatzfähigen System ist die, dass es keine Erfolgsgarantie gibt. Auch bei erfolgreichem Abschluss des Projekts und Erfüllen aller Anforderungen von TRL4 dauert die Weiterentwicklung hin zu einem einsatzfähigen System in der Regel noch weitere, mehrere Jahre, und ist von zahlreichen Unsicherheiten geprägt, z.B. von der weiteren Finanzierung, welche außerhalb der KIRAS Finanzierung sichergestellt werden muss, vom Auslaufen des Human+ Konsortialvertrags, welcher die Kooperation der Projektbeteiligten sichergestellt hat und nun durch, falls notwendig, bilaterale Kooperationen der Projektpartner ersetzt werden muss, und dergleichen.

6.5.2 Unmittelbarer Zweck – Evaluation und Entscheidung

Das Argument des Bedarfs ist ein Nützlichkeitsargument, das grundsätzlich die (Weiter-)Entwicklung der Human+ Technik rechtfertigt: Human+ kann Behörden und Organisationen mit Sicherheitsaufgaben dabei helfen, die ihnen gesetzlich auferlegten Aufgaben zu erfüllen. Dass Migrationsprognosen,

der Austausch von Informationen und ein gemeinsames Lagebild allerdings multifunktionell von diesen BOS für unterschiedliche weitere Zwecke verwendet werden können, ist kein alleiniges Human+ Charakteristikum – die meisten technischen Entwicklungen (selbst einfache Artefakte wie Messer) lassen sich multifunktional verwenden. Weiterer Klärungsbedarf zu den hierfür vorgesehenen Zwecken ist damit nötig, und wird im nächsten Abschnitt vorgenommen.

Das Argument des unsicheren Erfolgs wiederum betrifft die technische Machbarkeit, und nicht die moralische Zulässigkeit der Human+ Technik. Zwar kann eingewandt werden, dass öffentliche Forschungsgelder womöglich, statt sie in besonders unsichere Technologie-Entwicklungen zu investieren, (utilitaristisch argumentierend) in anderen Anwendungsfeldern einen größeren oder unmittelbareren Nutzen bringen können – letztendlich stellt sich diese Streitfrage aber bei praktisch allen Forschungsprojekten, insbesondere bei jeglicher Art von Forschung, die nicht direkt auf bereits marktreife Anwendungen zielt und daher mit Unsicherheiten behaftet ist. Aus pragmatischer Sicht kann zudem noch ergänzt werden, dass das Human+ Projekt zum Zeitpunkt dieser ethischen Reflexionen ja bereits bewilligt und sogar abgeschlossen wurde, sodass dieser Aspekt nur der Vollständigkeit halber angeführt wird.

6.5.3 Mittelbarer Zweck – Argumente pro und kontra

Von der Multifunktionalität von technischen Entwicklungen war bereits die Rede. Um die ethische Legitimität von multifunktionaler Technik beurteilen zu können, muss daher auch deren **mittelbarer primärer Zweck** beleuchtet werden: wofür sollen die erstellten Migrationsprognosen und das gemeinsame Lagebild (also die Ergebnisse des unmittelbaren Zwecks) in weiterer Folge eingesetzt werden? Die Schwierigkeit bei der Bestimmung des „mittelbaren primären Zwecks“ liegt, in größerem Maße als beim unmittelbaren Zweck, in der Frage, wer diesen Zweck bestimmen kann? Die Entwickler? Die Fördermittelgeber? Die späteren AnwenderInnen der Technik? Für Analysezwecke wird im Rahmen der vorliegenden Arbeit eine Festlegung vorgenommen: Der **mittelbare Zweck** kann aus dem Antrag entnommen werden, schließlich wurden die öffentlichen Fördermittel zur Finanzierung von Human+ auf Basis der dortigen Angaben bewilligt.

„[Das] Human+ Projekt soll zur **Gewährleistung von humanitärer Sicherheit und einer besseren Vorbereitung auf Migrationswellen** beitragen, unter gleichzeitiger umfassender Betrachtung von **rechtlichen, ethischen und soziologischen Rahmenbedingungen**.

[...]

Durch das so gewonnene Echtzeit-Lagebild sind die Einsatzkräfte besser auf Migrationswellen und deren Herausforderungen vorbereitet und können **umfassend informierte, passgenaue Entscheidungen** treffen“ (Human+ 2016: 1; Hervorhebungen wie im Original).

Humanitäre Sicherheit, dies als Vorausgriff auf die weiter unten noch ausführlicher erfolgende Begriffsklärung, kann in einer ersten Näherung als „Gewährleistung von elementaren Menschenrechten im Rahmen einer allgemeinen Schutzverantwortung“ bezeichnet werden (Daase 2010: 12).

Mit diesen Festlegungen kann die Hauptfrage 1 neu gestellt werden:

Ist die Verbesserung der Gewährleistung von humanitärer Sicherheit und einer besseren Vorbereitung auf „Migrationswellen“ ethisch legitim?

Dass etwa mit Migrationsprognosen auch andere Zwecke erreicht werden können, z.B. das „rechtzeitige“ Befestigen von Grenzen mit dem Ziel zwecks völliger Abschottung fällt somit in weiterer Folge in den Bereich der Hauptfrage 2 zu den durch die Technik eröffneten außerplanmäßigen Handlungen.

Festgehalten werden muss, dass die Suche nach Argumenten pro und kontra einer Verbesserung von humanitärer Sicherheit nicht gleichzusetzen ist mit der Suche nach Argumenten pro und kontra hinsichtlich der Frage, ob es moralisch geboten, erlaubt oder verboten sein soll, Maßnahmen zu treffen, dass keine irregulären MigrantInnen oder Geflüchteten das Land betreten können. Diese zweiten Situationen haben zwar gemeinsam, dass mit einer gewissen Wahrscheinlichkeit Personen ankommen werden, aber in der einen Situation wird bewusst versucht die humanitäre Sicherheit durch bessere Vorbereitung zu erhöhen, in der anderen Situation nicht.

Für die Bewertung des primären Zweckes in Hauptfrage 1 kann nach Ansicht des Verfassers der vorliegenden Arbeit von einer idealisierten Anwendungssituation ausgegangen werden (im Sinne von: die AnwenderInnen verwenden die funktionierende Technik nach bestem Wissen und Gewissen), weil Abweichungen davon in den Hauptfragen 2 und 3 betrachtet werden.

Zwei im Antrag unterbestimmte Aspekte müssten allerdings noch genauer beleuchtet werden, um den mittelbaren primären Zweck bewerten zu können: erstens was unter der Verbesserung von humanitärer Sicherheit bzw. besseren Vorbereitung auf Migrationswellen zu verstehen ist, und zweitens unter

welchen Bedingungen Human+ angewandt werden soll bzw. welche AnwenderInnen-Gruppen konkret damit einher gehen könnten.

Klärung von humanitärer Sicherheit und besserer Vorbereitung auf Migrationswellen:

„Humanitäre Sicherheit“ wird im Human+ Projektantrag nicht genau definiert. Folgende Aspekte werden aber in einem diesbezüglichen Absatz im Rahmen der Problemdefinition genannt, sodass man eine Art extensionale Definition geben kann: Aufnahme, Weiterleitung, Versorgung und Unterbringung von MigrantInnen (mit einem Verweis auf die Grenzen der etablierten Methoden des Migrationsmanagements während der „Flüchtlingskrise“, sodass es hier zu Überschneidungen zwischen beiden Konzepten kommen dürfte), aber auch Planbarkeit von (hierzu) benötigten Ressourcen wie Personal, Wasser, medizinischer, und sanitärer Artikel (Human+ 2016: 8).

Zieht man dagegen etablierte Definitionen des Begriffs heran, dann zeigt sich, dass der Begriff „**humanitäre Sicherheit**“ vergleichsweise selten (in der deutschsprachigen) Literatur genutzt wird, am prominentesten von Daase als Teil des „erweiterten Sicherheitsbegriffs“ (vgl. Daase 2010: 10): Daase verweist dabei auf einen Wandel der „Sicherheitskultur“, wonach „Sicherheit“ von militärischer Sicherheit in den letzten 50 Jahren sich graduell über die ökonomische und dann ökologische Sicherheit hin zur (nach dem Ende des Ost-West-Konflikts) humanitären Sicherheit erweitert hat. Das heißt, dass im Zuge dieser Entwicklung immer mehr in Richtung Sicherheit des Individuums geblickt wird, während davor zuerst der Staat und später noch die Gesellschaft im Vordergrund gestanden haben. „Humanitäre Sicherheit bezieht sich nicht nur auf die Sicherung von Schutzzonen und Katastrophenhilfe in Krisenregionen, sondern auch auf die Gewährleistung elementarer Menschenrechte im Rahmen einer allgemeinen Schutzverantwortung (responsibility to protect) der internationalen Gemeinschaft“ (Daase 2010: 12).⁵¹

Da „humanitäre Sicherheit“ in der deutschsprachigen Literatur kaum genutzt wird, scheint es vielversprechend, sich die Verwendung im englischsprachigen Raum anzusehen. Erschwert wird dies aber durch den Umstand, dass das deutsche Wort „Sicherheit“ im Englischen in zwei Varianten auftritt (vgl. Davey und Wootton 2016: 43): einmal als „security“ (meist gebraucht bezogen auf absichtlich, böswillig zugefügte Schäden), und einmal als „safety“ (welches sich stärker auf „zufällige“ Gefahren, z.B. Verletzungen bei Unfällen, bezieht). Gleichzeitig werden „security“ und „safety“ häufig auch im Englischen austauschbar verwendet, was eine exakte Begriffsauslegung verkompliziert.

⁵¹ Dies eröffnet nach Daase aber auch die Möglichkeit, durch die Verbindung von Menschenrechten und Sicherheit auch militärisch humanitär zu handeln oder, „humanitäre Kriege“ zu führen.

„Humanitarian security“ bezieht sich im englischsprachigen Raum in erster Linie auf die Sicherheit von MitarbeiterInnen, die im humanitären Bereich arbeiten – immer stärker verwoben mit Risikomanagement-Konzepten aus dem unternehmerischen Bereich (vgl. Hoffmann 2017: 102).

Dases Begriff von humanitärer Sicherheit durch die Betonung der Gewährleistung von Menschenrechten ist nahe an dem, was auch **humane bzw. menschliche Sicherheit** genannt wird, bzw. im englischen „human security“. Humane Sicherheit betrifft „[...] the right of people to live in freedom and dignity, free from poverty and despair and promotes the protection of their physical safety, economic and social well-being and human rights. It includes the right of all individuals, vulnerable people in particular, to live free from fear and free from want, with an equal opportunity to enjoy all their rights and fully develop their human potential“ (IOM 2019: 95). Ulbert und Werthes (2008a: 19ff) wiederum verweisen auf drei Dimensionen von menschlicher Sicherheit: a) Freiheit/Menschenrechte und Rechts-sicherheit, b) Gleichheit & soziale Gerechtigkeit (Freiheit von Not), sowie c) humanitäre Sicherheit (Freiheit von Furcht). Hier wird humanitäre Sicherheit als Unterdimension der menschlichen Sicherheit betrachtet. Menschliche Sicherheit im umfassenden Sinne inkludiert als „zentrale Strategien und Instrumente“ nicht nur (im internationalen Sinne) Maßnahmen der Konflikt- und Krisenprävention sowie der Friedenskonsolidierung, sondern auch (humanitäre) Nothilfe sowie (militärische) humanitäre Interventionen. Die AutorInnen verweisen dabei auf das „engere“ und „weitere“ Verständnis von menschlicher Sicherheit – ein engeres Verständnis grenzt es auf gewaltsame Konflikte ein, das weitere Verständnis zieht obige drei Dimensionen umfassend in Betracht.

Beiden Begriffen, humanitärer Sicherheit und humaner Sicherheit, ist gemein, dass der Bezugspunkt das menschliche Individuum ist (vgl. Ulbert und Werthes 2008: 21; Daase 2010: 10), also nicht mehr der Schutz von Staaten oder (ganzen) Gesellschaften, sondern der Schutz von einzelnen Personen, was sich sowohl auf MitarbeiterInnen im Rahmen von humanitärem Engagement wie auch von Geflüchteten beziehen kann.⁵²

Ob nun im weiteren oder im engeren Begriff, in beiden Fällen bezieht sich der „mittelbare Zweck“ von Human+ auf die Verbesserung von humanitärer Sicherheit in einer Weise, dass damit Leben geschützt werden. Bei einem erweiterten Verständnis des Begriffs sollen die Migrationsprognosen dafür eingesetzt werden, um Individuen (Geflüchtete) auf eine Weise zu versorgen, dass sie „free from fear and free from want“ sind. Die im Antrag aufgezählten Aspekte, wie Aufnahme, Weiterleitung, Versorgung und Unterbringung von MigrantInnen sowie die Verbesserung der Planbarkeit des Bedarfs an Personal, Wasser, medizinischen und sanitären Artikeln fügen sich nahtlos in dieses erweiterte Verständnis von

⁵² Der Schutz von Minderheiten durch Minderheitsrechte wird von Ulbert und Werthes (2008: 20) als Beispiel für Freiheits- und Menschenrechte genannt, welcher in ihrer Diktion somit (ebenfalls) in den Bereich der humanen Sicherheit fällt. Angesichts dessen lässt sich trotz der ausdrücklichen Formulierung der AutorInnen, dass der Bezugspunkt das Individuum sei, wohl auch für den Einbezug von Gruppen von Individuen unter den Begriff „humane Sicherheit“ argumentieren.

humanitärer Sicherheit ein. Für die nachfolgende Beurteilung wird im Sinne eines konditionalen Bezuges⁵³ dieses **erweiterte Verständnis** zu Grunde gelegt. Die alleinige Anwendung zu anderen Zwecken, z.B. des Einsatzes der Prognosen, um rechtzeitig Wegweisungen zu ermöglichen, fällt somit nicht in den Bereich der Sicherstellung von humanitärer Sicherheit, und wird im Abschnitt zu illegitimen Verwendungsweisen behandelt.

Klärung möglicher Anwendungsbedingungen und konkreter AnwenderInnen-Gruppen:

Eine weitere Facette aus dem potentiellen Anwendungskontext von Human+ bedarf an dieser Stelle noch einer Klärung, nämlich nach der für die Erreichung des primären Zweckes mit der Human+ Technik zu betrauenden **AnwenderInnen-Gruppe** und den **konkreten Anwendungsbedingungen**. Während der Projektlaufzeit von Human+ wurden diese Aspekte nur teilweise geklärt bzw. festgelegt.

Zwei grundsätzliche Möglichkeiten kristallisierten sich allerdings während verschiedener projektinterner Diskussionen heraus:

- a) Anwendung von Human+ Komponenten lediglich innerhalb von akuten Krisensituationen: Bei Katastrophen- und Krisensituationen kommt es im Bereich von BOS bzw. des staatlichen Krisen- und Katastrophenmanagements zur Einrichtung sogenannter „**Besonderer Aufbauorganisationen**“ (BAO). Einsatzorganisationen verfügen über zwei Arten von formalen Strukturen: eine bürokratische Struktur zum Aufbau und Erhalt von Fähigkeiten im sogenannten Bereitschaftszustand (d.h. außerhalb von akuten Situationen), sowie eine flexible Struktur im Einsatz (vgl. Kern 2020: 25). Die konkrete Form der Einsatzstruktur wird je spezifisch an akute Situationen angepasst (Schadenstyp, Schadensobjekt, Schadensausmaß, etc.). Für Routinelagen heißt die Einsatzstruktur „**Allgemeine Aufbauorganisation**“ (AAO). Für besondere Lagen bzw. Sonderlagen kann eine BAO eingerichtet werden. BAO inkludieren die Einrichtung sogenannter Stabsstrukturen. Stäbe sind grundsätzlich in Sachgebiete gegliedert und mit durchnummerierten S-Nummern bezeichnet. Dies trifft sowohl für Stäbe von Einsatzorganisationen auf österreichischer und deutscher Seite wie auch [Großteils] auf die Führungsgrundgebiete des Österreichischen Bundesheeres zu (vgl. Dudek 2020: 226). Relevant für den vorliegenden Kontext ist herbei besonders „S2 – Lage“, da sich dieses Sachgebiet mit der Erstellung eines umfassenden Lagebildes beschäftigt. Assistenzpersonal im Rahmen des S2 könnte somit mit der Bedienung der Human+ Technik betraut werden, die (bei einem entsprechend ambitionierten Rollout der Technik) auch im Vorfeld eine Ausbildung zum Fachanwender durchlaufen können.⁵⁴ Zu berücksichtigen ist, dass während der „Flüchtlingskrise 2015/2016“ Stäbe auf unterschiedlichen Ebenen (Bezirke, Stadt Wien, Österreich) sowie von unterschiedlichen BOS (Rotes

⁵³ Siehe Seite 27 der vorliegenden Arbeit.

⁵⁴ Während einer Evaluationsübung zu Human+ gegen Ende der Entwicklungslaufzeit des Projekts zeigte sich jedenfalls, dass eine Anwendung ohne vorherige, längere Einweisung nicht möglich ist. Die betrifft speziell die Interpretation der weniger

Kreuz, Polizei, etc.) aufgebaut wurden, sodass zu überlegen wäre, welche dieser Stäbe Zugriff auf eine (vereinheitlichte und womöglich mit demselben Datenstand abgeglichene) Human+ Technik erhalten sollen.⁵⁵ Der Vorteil der Anwendung im Rahmen von BAO liegt darin, dass die Human+ Anwendung somit zeitlich automatisch „befristet“ ist: wenn es mehr keine außergewöhnliche Situation gibt, zu deren Bewältigung die BAO einberufen wurde, wird sie aufgelöst. Dies würde sicherstellen, dass es zu keiner **ausufernden Anwendung** der Human+ Technik kommt⁵⁶, hat aber das Problem, dass eine frühprognostische Anwendung dann nicht mehr möglich ist – Human+ würde per Definition erst dann einsetzbar sein, wenn die außergewöhnliche Situation bereits in einer Weise absehbar ist, dass eine BAO errichtet wird. Human+ kann dann aber selbst keinen Beitrag zum frühzeitigen Erkennen liefern, d.h. die Vorbereitungsphase für die Lagebewältigung bzw. die Chaosphase profitiert nicht davon, womit die Phase der eigentlichen Lagebewältigung nicht verkürzt werden kann. Ein weiterer Vorteil der befristeten Nutzung im Rahmen von BAO liegt darin, dass im Falle dieser Anwendungsgruppe die Verwendung durch eine „vertrauenswürdige“ AnwenderInnen-Gruppe sichergestellt ist, nämlich durch die MitarbeiterInnen der Stäbe, welche in ihrer Arbeit ohnehin mit vertraulichen Informationen hantieren.

- b) Anwendung von Human+ Komponenten (auch) außerhalb von akuten Krisensituationen: Die Möglichkeiten einer Human+ Anwendung außerhalb von Kontexten mit ausgerufenen Stabsstrukturen (und damit außerhalb von sich abzeichnenden Katastrophen- oder Krisenzeiten) wurde bereits innerhalb des Konsortiums während der Projektlaufzeit diskutiert, und teilweise auch zum Gesprächsgegenstand im Rahmen der Interviews der Bedarfserfassung gemacht. Auch wenn hierbei keine erschöpfende Liste von AnwenderInnen erarbeitet wurde, wurde summarisch von den „einsatzrelevanten Trägern“ gesprochen (Organisationen des Bevölkerungsschutzes, mit „Flüchtlingsaufgaben“ betraute Fachorganisationen, Behörden verschiedener Ebenen und BOS). Dies schließt folglich die Human+ Anwendung durch private bzw. privatwirtschaftliche Organisationen aus. Eine potentielle AnwenderInnen-Gruppe, die vor allem im Umfeld von Einsatzorganisationen mit starkem Ehrenamtlichen-Anteil zum Tragen kommt, sind die **VOST** (Virtual Operations Support Teams). VOST sind eine mögliche Ausformung der sogenannten „digital volunteers“, und werden in erster Linie assoziiert mit dem Monitoring von sozialen Medien, der Unterstützung im Bereich

intuitiv verständlichen Ergebnisse der Algorithmen, z.B. die auf einer Karte visualisierten Schätzungen zu Ankunftsahlen auf Basis von Social-Media-Auswertungen – und die Aufklärung darüber, mit welchen Unsicherheiten diese Zahlen behaftet sind.
⁵⁵ Auf deutscher Seite gab es ebenfalls Stäbe auf unterschiedlichen Ebenen, aber mit anderen Bezeichnungen aufgrund der geringfügig anders gestalteten föderalen Struktur.

⁵⁶ So hatte ein Interview mit einem Experten vom Institut für Technikfolgenabschätzung in Wien, das der Verfasser der vorliegenden Arbeit während der Projektlaufzeit des Human+ Projekts durchgeführt hat, ergeben, dass eine zeitlich uneingeschränkte Erfassung und Analyse des gesamten Twitter-Verkehrs (worauf die Social-Media-Analyse hinauslaufe) kritisch zu sehen ist, da dieser Erkenntnisgewinn nicht in Relation zur potentiellen Privatsphäre-Beeinträchtigung stünde (vgl. Human+ 2019b). Der Experte sprach sich stattdessen für eine „sehr enge“ Definition von Anwendungskriterien der Human+ Technik aus, die auf eine „zeitnahe“ Verwendung statt eines laufenden Monitorings abziele.

Lageerfassung, der Bekämpfung von Gerüchten und der Verbreitung von offiziellen Informationen, die von regulären Katastrophenbewältigungsbehörden ausgegeben werden (vgl. Roth und Prior 2019: 53f). Roth und Prior verweisen darauf, dass VOST Großteils selbstorganisierte Gruppen wären, die aber auf offizielle Anerkennung durch den etablierten Katastrophenmanagementsektor abzielen würden. Der Begriff VOST wird seit 2011 verwendet, seit das Konzept formell im Rahmen einer Konferenz vorgestellt wurde. Organisierte Gruppen von digitalen Freiwilligen habe es aber schon zuvor gegeben. Die Autoren haben in 2018 eine Befragung durchgeführt, und dafür insgesamt 27 VOST weltweit kontaktiert. Die meisten VOST-MitarbeiterInnen in unterschiedlichen Ländern haben einen ehrenamtlichen Hintergrund im Bereich Katastrophenbewältigung, z.B. bei Freiwilligen Feuerwehren oder anderen notfallbezogenen Organisationen (vgl. Roth und Prior 2019: 56f). Während Österreich zum Zeitpunkt des Verfassens der vorliegenden Arbeit noch kein VOST aufweist, zeigt die Seite der Verbandes Europäischer VOST für Deutschland vier VOST: Baden-Württemberg, Hamburg, Mülheim an der Ruhr, Rheinland-Pfalz (vgl. VOST Europe 2021). Auch wenn es keine einheitlichen Standards für die Rekrutierung von Freiwilligen für VOST gibt, werden Mitglieder zumindest von manchen VOST (z.B. im Fall der VOST Baden-Württemberg) auch einer Sicherheitsüberprüfung unterzogen, unterliegen Verschwiegenheitspflichten, und manchmal ist in jedem Fall auch eine aktive Mitgliedschaft in Bevölkerungsschutzorganisationen notwendig (von denen die Mitglieder empfohlen werden), da im Einzelfall VOST-MitarbeiterInnen mit vertraulichen Informationen konfrontiert sein können (vgl. Feuerwehr Kernen 2018). In jedem Fall würde eine Anwendung außerhalb von BAO die langfristige, ereignisunabhängige Human+ Verwendung ermöglichen, was einerseits das Risiko einer „ausufernden Technikanwendung“ mit sich bringt, andererseits aber auch dauerhaftes Monitoring ermöglicht, um Früherkennung zu betreiben.

Mit diesen Klärungen lassen sich nun folgende Argumente für und gegen den mittelbaren Human+ Zweck formulieren:

Pro-Argumente:

- **Argument der Erfüllung von Grundrechten:** Die Verbesserung von humanitärer Sicherheit dient gemäß obigen Definitionen entweder unmittelbar dem Schutz von Leben oder in der erweiterten Fassung der Verbesserung der Qualität dieses Lebens durch Freiheit von Angst und Not. Besondere Bedeutung erhält dieses Argument, wenn man sich die Art der Bedürfnisse vor Augen führt, die im Rahmen der akuten Betreuung von Geflüchteten gestillt werden (Obdach, Wasser, Nahrung, medizinische und hygienische Versorgung, in weiterer Folge aber auch Strom für das Laden des Smartphones). Diese können, unter Bezug auf häufig in ökologischen Kontexten geführte Debatten, auch als „**grundlegende Bedürfnisse**“ [basic needs] bezeichnet werden: während sich diese je nach Organismus unterscheiden können, beziehen sie sich im Falle von Menschen auf diejenigen Dinge,

die rationale und aufgeklärte Menschen als essentielle Teile ihrer Existenz als Personen erfordern (vgl. Taylor 2011: 272f). Das umfasst z.B. Substistenz (im Sinne von Lebensnotwendigkeiten) wie auch Sicherheit, Autonomie und Freiheit. Erst durch das Stillen dieser grundlegenden Bedürfnisse werden Menschen in die Lage versetzt, nach anderen Zielen überhaupt streben zu können (und sei es weil sie, auf einer basalen Ebene, ansonsten ihren Organismus nicht lebensfähig erhalten können). Solche grundlegenden, bzw. **vitalen Bedürfnisse**, die Voraussetzung dafür sind, überhaupt nicht-grundlegende Bedürfnisse verfolgen zu können, weil erst durch sie ein Mindestmaß an Handlungsfähigkeit [agency] erzielt wird (vgl. Alvarez 2009: 491ff) sind zudem vergleichsweise unabhängig vom fraglichen kulturellen Kontext. Sie können somit mit einiger Berechtigung universell angenommen werden, und sei es aus dem Grund, dass für die Feststellung, dass dies nicht der Fall sei, die Person, die diese Urteile fällt, ihrerseits ihre grundlegenden Bedürfnisse so gestillt haben muss, dass sie dieses Urteil fällen kann (vgl. Alvarez 2009: 488). Der mittelbare primäre Zweck von Human+ gestaltet sich somit als Beitrag, um die Grundrechte von Geflüchteten besser wahren zu können (z.B. das Recht auf Leben [vgl. EU 2012: Artikel 2] durch basale Versorgung mit Wasser und Nahrung, oder das Recht auf Unversehrtheit durch Versorgung mit notwendigen medizinischen oder hygienischen Artikeln [vgl. EU 2012: Artikel 3]).

Kontra-Argumente:

- **Argument des Pulleffekts:** Mitunter wird an Maßnahmen, die unmittelbaren humanen bzw. humanitären Zwecken an Geflüchteten dienen, bemängelt, dass diese zwar in unmittelbarer Hinsicht das Leid von MigrantInnen und Geflüchteten mildert, gesamtheitlich und auf längere Frist aber zu größerem Leid führt, weil die kurzfristige humanitäre Hilfe als Pull-Faktor dienen kann und weitere Migration auslöst (bzw. bestehende Migration in andere Bahnen lenkt). So wird etwa bei den Rettungen im Mittelmeer, bei welchen es um das Retten von unmittelbar in Seenot geratenen MigrantInnen geht, bemerkt, dass zwar das Retten von (potentiell) vom Ertrinken bedrohten Menschen wichtig sei, dass aber die Tatsache von solchen Rettungen dazu führt, dass sich immer mehr MigrantInnen auf potentiell immer seeuntüchtiger werdenden Schiffen auf die gefährvolle Reise über das Mittelmeer machen, um in Europa zu landen. So lautet etwa eine Stellungnahme des österreichischen Außenministeriums zum österreichischen Veto gegenüber einem Neustart der EU-Mission Sophia⁵⁷: „Österreich befürwortet eine aktive Rolle der EU zur Unterstützung der Bemühungen in Libyen, ist aber bekanntlich gegen eine Wiederaufnahme von Maßnahmen, die neue

⁵⁷ Diese dient an sich der Überwachung des Waffenembargos in Libyen, gemäß DerStandard (2020a) wurde aber befürchtet, dass mit der dazu notwendigen Aussendung von Überwachungsschiffen auch vermehrt ein Zustrom von Geflüchteten zu erwarten sei.

Pull-Faktoren für Migrationsbewegung über das Mittelmeer schaffen“ (DerStandard 2020a). Sophia wird hier als Pull-Faktor identifiziert. Im Sinne eines Analogieschlusses könnte für die Verbesserung der humanitären Versorgung daher auch argumentiert werden, dass diese Verbesserung als Pull-Faktor für weitere Geflüchtete dienen könnte.

- **Argument des „function creep“:** Der Terminus **“function creep”**⁵⁸ lässt sich folgendermaßen definieren: „[...] function creep of *X* happens when *X*'s function, i.e. what *X* is meant to perform, expands or shifts slowly, gradually, and/or imperceptibly“ (Koops 2021: 37). Der Begriff wird häufig uneinheitlich gebraucht und meistens nur lose definiert – so ist nicht in jeder Verwendung in der Literatur das Element einer langsamen, kaum wahrnehmbaren Wandlung vorhanden (vgl. Koops 2021: 37ff). In jedem Fall geht es aber darum, dass sich etwas über seinen ursprünglichen Zweck in einer Weise hinausbewegt, welche durch ihre EntwicklerInnen, BenutzerInnen oder aber die Öffentlichkeit (vermeintlich) nicht vorhersehbar gewesen wäre: ein System, eine Technologie, eine Datenbank, ein Projekt, ein Gesetz, häufig mit Bezug auf die Verarbeitung von persönlichen Daten. Oft wird der Begriff „function creep“ gebraucht, um auf eine drohende Gefahr oder ein Risiko hinzuweisen. Eine erweiterte Version dieser Definition, die auf datenverarbeitende Systeme zugeschnitten ist, ist folgende: „Function creep comes down to an expansion or shift in the proper activity of a data-processing system in relation to its purpose or use, which is considered contestable because people insufficiently recognise that it involves a qualitative change that raises public-policy concerns“ (Koops 2021: 53). Zentral in dieser zweiten Definition ist, dass die Erweiterung deshalb problematisch sei, weil sie schwer wahrnehmbar vor sich gehe, und daher etwa keine öffentlichen Debatten über die Wünschbarkeit dieser Änderung möglich seien (womit etwa nicht grundsätzlich innovative Änderungen oder Adaptionen gemeint sind, welche nach Ansicht des Autors ein natürliches Phänomen in sozialen Systemen wären). Gründe dafür, weshalb eine Funktionserweiterung (von ihren Gegnern) abgelehnt werden kann, können u.a. folgende sein (vgl. Koops 2021: 53): das System wird ineffektiv oder nicht mehr handhabbar (z.B. durch Hinzufügen immer weiterer Features), es entsteht ein Legitimationsdefizit (z.B. durch „Mission Creep“, den Missbrauch von Macht bzw. Handlungsoptionen, oder durch sogenannte Sekundärnutzungen), oder aber es fehlen angemessene Checks & Balances (weil das System in andere Bereiche mit anderen Normen expandiert).

Kritiker könnten gegenüber der Human+ Technik nun einwenden, dass auf Grund der vielseitigen Verwendbarkeit der generierten Migrationsprognosen in Kombination mit der vergleichsweise un-

⁵⁸ An dieser Stelle wird „function creep“ als Überbegriff verwendet. Koops (2021) verweist auf zahlreiche, damit verwandte Konzepte, die in verschiedenen Disziplinen benutzt werden, z.B. purpose creep (im Falle der Datenverwendung), mission creep, surveillance creep.

terbestimmten Definition von „humanitärer Sicherheit“ und „besseren Vorbereitung auf Migrationswellen“ einerseits schon eine Mehrdeutigkeit im Projekt-Antrag angelegt ist: wird etwa die „bessere Vorbereitung auf Migrationswellen“ aus der Perspektive eines (wie oben angeführt) eher auf volkswirtschaftliche Nützlichkeit der Ankommenden achtenden Migrationsmanagements gesehen, so ergeben sich andere Perspektiven, als wenn diese Vorbereitung eher aus Sicht der humanen Sicherheit betrachtet wird. Zudem bedeuten die Festlegungen im Antrag nicht, dass diese festgelegten Zwecke auch für etwaige Weiterentwicklungen der Technik durch die einzelnen Partnerinstitutionen des Human+ Projekts (deren geistiges Eigentum die Entwicklungen bleiben) gelten müssen, und somit nicht etwa auch in andere Richtungen fortgeführt werden können.

6.5.4 Mittelbarer Zweck – Evaluation und Entscheidung

Das „Argument der Erfüllung der Grundrechte“ spricht zunächst für die ethische Zulässigkeit des mittelbaren primären Zwecks von Human+. Problematisch ist die Unterbestimmtheit der Begriffe „humanitäre Sicherheit“ und „bessere Vorbereitung auf Migrationswellen“. Diese wurden oben unter Rekurs auf die Literatur zwar weiter beleuchtet, im Endeffekt aber erforderte es eine Setzung der Bedeutung, wobei grob abweichende Bedeutung in den Bereich einer „nicht-intendierten Anwendung“ verschoben und im Abschnitt zu Hauptfrage 2 betrachtet werden müssen. Gemäß der vorgenommenen Setzung bedeutet „humanitäre Sicherheit“ in besonderem Maße Versorgung in akuten Flüchtlingslagen, speziell was die Deckung von grundlegenden Bedürfnissen betrifft. Dieses Argument spricht somit für die Legitimität des mittelbaren primären Zwecks.

Dagegen spricht das Argument des Pulleffekts, wonach das Wissen um gute akute Versorgung mittelfristig motivierend wirken dann, dass weitere MigrantInnen sich auf eine gefährliche Reise machen. Hierzu ist zunächst zu sagen, dass es unwahrscheinlich scheint, dass eine gute Akutversorgung vor Ort eine große Rolle in der persönlichen Migrations- oder Fluchtentscheidung spielt. Das Argument des Pulleffekts wurde formuliert in Form eines Analogieschlusses von der Seerettung im Mittelmeer auf Inlandssituationen. Tatsächlich deuten neuere Forschung daraufhin, dass kein Zusammenhang zwischen der Präsenz von (Seerettung betreibenden) NGOs und MigrantInnen, die von der Libyschen Küste aufbrechen, festgestellt werden kann (vgl. Cusumano et al 2019: 7). Stattdessen wurde eine starke Korrelation festgestellt zwischen MigrantInnen, die aufbrechen, und den jeweils vorherrschenden Wetterbedingungen bzw. der politischen Situation vor Ort. Und wenn schon das Vorliegen der Seenotrettung keinen Zusammenhang mit Aufbrüchen hat (bei welchen es besonders direkt um Leben und Tod geht), dann ist zu erwarten, dass erst recht eine bessere Not-Versorgung in Österreich keinen Einfluss darauf haben wird. Zudem wird in der Literatur häufig die allzu schematische Einteilung in Pull- und Pushfaktoren kritisiert, zugunsten von ausgefeilteren Konzepten wie Migrationsnetzwerken oder

-Infrastrukturen. Auch wenn hierzu weitere Forschung nötig ist, lässt sich das Argument des Pulleffekts auf sachlicher Ebene dadurch in Zweifel ziehen.

Das Argument des „function creep“ stellt die Befürchtung in den Vordergrund, dass trotz Bestimmung des mittelbaren primären Zweckes die Anwendung der Human+ Technik auf andere Weise als wie festgelegt vor sich gehen könnte – was einerseits mögliche Weiterentwicklungen über TRL4 hinaus betrifft, andererseits auch die Anwendungskontexte selbst. Dieses Risiko ist nach Ansicht des Verfassers der vorliegenden Arbeit durchaus ernst zu nehmen, und unterstreicht die Wichtigkeit der Hauptfrage 2 zu den Möglichkeiten, moralisch illegitime Anwendungen zu unterbinden. An dieser Stelle kann hierzu schon vorausgeschickt werden, dass klare Methoden, solche Zweckentfremdungen zu unterbinden, nicht möglich sind. Es bleibt somit ein Spannungsverhältnis zwischen der Legitimität des mittelbaren primären Zweckes, und dem Risiko, dass die Technik anderweitig angewandt wird.

Weiter vorne wurde das Metaprinzip des Negativen Utilitarismus formuliert (vgl. Seite 39), welches besagt, dass in Abwägungsfragen eine Verminderung von Leid und negativen Konsequenzen gegenüber der Förderung von Glück bzw. positiven Konsequenzen Vorrang haben sollte, und die Variante mit dem geringsten Ausmaß an Leid gewählt werden sollte. Welche Implikationen ergeben sich für die Beurteilung des mittelbaren primären Zwecks? Dieser besteht in der **Gewährleistung von humanitärer Sicherheit und einer besseren Vorbereitung auf Migrationswellen**. Dies bedeutet im Wesentlichen die Unterstützung von Akutmaßnahmen, die sich auf die Erfüllung grundlegender Bedürfnisse beziehen. Der Asymmetrie von Leid und Glück folgend (vgl. S. 37), handelt es sich beim Bereitstellen von Nahrung, Schlafmöglichkeiten und ähnlichen Versorgungsangeboten nicht um Maßnahmen, die das Glück fördern sollen, sondern um solche, die typische Quellen von Leid bekämpfen, z.B. Hunger, fehlendes Obdach, Mangel an Kleidung und medizinischer Versorgung.⁵⁹ Da das Verfolgen des mittelbaren primären Zwecks folglich ein Beitrag zur Reduzierung von Leid ist, verglichen mit der Alternative, dies nicht zu tun, spricht vor dem Hintergrund von Hauptfrage 1 das Metaprinzip des Negativen Utilitarismus für die Verfolgung des mittelbaren primären Zwecks.

Wird Human+ dagegen zu anderen Zwecken verwendet (z.B. zur Schließung von Grenzen), dann würde ein völlig anderer primärer Zweck verfolgt werden, und die im Rahmen der vorliegenden Arbeit getroffenen ethischen Beurteilungen müssten grundlegend überdacht werden. Solche vom primären Zweck abweichende Maßnahmen werden in der vorliegenden Arbeit als durch Human+ ermöglichte, außerplanmäßige Handlungen in den Analysen zur Hauptfrage 2 betrachtet. An dieser Stelle sei auch bereits erwähnt, dass in den Analysen zur Hauptfrage 3 (Folgen und Nebenwirkungen der planmäßigen

⁵⁹ Würde man die genannten Maßnahmen zur Verbesserung der humanitären Sicherheit (Bereitstellung von Nahrung, Obdach, etc.) dagegen als glücksfördernd interpretieren, dann wäre implizit die Asymmetrie von Leid und Glück in Abrede gestellt, und leidvermindernde und glücksfördernde Maßnahmen hätten dadurch denselben Bewertungsmaßstab – was diejenigen Kritikpunkte am Glücksverständnis zuließe, die weiter oben bereits vorgebracht wurden.

Verwendungsweise von Human+) eine Reihe von ethischen Konfliktpunkten aufgeworfen wird, z.B. das zweifelhafte Einverständnis von Social-Media-UserInnen in die Verwendung ihrer Posts durch die Human+ Technik. Während manche dieser Folgen ethisch unerwünscht scheinen, etwa weil sie aus abstrakter Sicht zum Teil auch „Leid“ erhöhen, oder vor dem Hintergrund der verschiedenen Prinzipien mittlerer Reichweite, müssen sie für eine Letztbeurteilung gegenüber den intendierten Folgen der (Leid-verniedrigenden) Maßnahmen, die sich aus dem Anstreben des mittelbaren primären Zwecks von Human+ ergeben, abgewogen werden (vgl. hierzu Kapitel 6.7 ab Seite 132).

6.5.5 Zwischenfazit

In diesem ersten Bewertungsabschnitt ging es um die Beantwortung der Hauptfrage 1:

Sind die Handlungsmöglichkeiten, die durch diese Technik eröffnet oder verbessert werden, an sich ethisch legitim? (primärer Zweck)

Diese Frage wird eingedenk obiger Einschränkungen hiermit bejaht (speziell in Form von BAO, da hier tendenziell die Gefahr einer ausufernden Technikanwendung niedriger ist). Mögliche Missbrauchsmöglichkeiten und negative Folgen bei der Verfolgung des primären Zweckes werden in den Hauptfragen 2 und 3 beantwortet.

6.6 HAUPTFRAGE 2 (ERMÖGLICHTE AUßERPLANMÄßIGE ANWENDUNGEN) – SCHRITT 2 BIS 4

6.6.1 Klärung der Möglichkeiten außerplanmäßiger Handlungen

In diesem Abschnitt geht es um die Beantwortung der Hauptfrage 2:

Hauptfrage 2: Werden ausreichend Maßnahmen ergriffen, um alle moralisch bedenklichen außerplanmäßigen Handlungsweisen auszuschließen, die sich damit abseits vom primären Zweck durchführen lassen?

Um dies zu leisten, müssen zuerst zwei vorgelagerte Fragen beantwortet werden, nämlich:

Unterfrage 2a: Welche außerplanmäßigen Handlungsweisen abseits vom primären Zweck lassen sich durchführen?

sowie

Unterfrage 2b: Welche davon sind moralisch bedenklich?

Wie bei den meisten Techniken, lassen sich auch bei Human+ die Entwicklungen bzw. ermöglichten Handlungen für andere als den primären Zweck verwenden. Um daher zu den außerplanmäßigen Handlungsweisen zu gelangen, müssen mögliche Abweichungen vom primären Zweck betrachtet werden. Folgende Überlegungen bieten hierfür nach Ansicht des Verfassers der vorliegenden Arbeit einen heuristischen Wert:

- Die Human+ Komponenten werden korrekt genutzt, um den unmittelbaren Zweck zu erreichen (Migrationsprognose und gemeinsames Echtzeitlagebild), aber diese Outputs werden nicht für die Verbesserung der humanitären Sicherheit verwendet, sondern für **andere mittelbare** Zwecke.
- Die Human+ Komponenten werden genutzt, um **andere unmittelbare** Zwecke als wie die Migrationsprognose und ein verbessertes Echtzeitlagebild zu erreichen, d.h. es geht erst gar nicht um die Erstellung von Migrationsprognosen und eines gemeinsamen Echtzeitlagebildes, sondern Human+ Komponenten werden (adaptiert) für die Erreichung anderer Informationsoutputs genutzt.

Da wie oben schon angeführt zahlreiche Anwendungsbedingungen nicht final geklärt sind (z.B.: Wer genau soll die Technik anwenden? Wann soll sie eingesetzt werden: laufend oder anlassbezogen?), können diese Aspekte nur vorläufig unter den außerplanmäßigen Handlungsweisen beleuchtet werden, obwohl auch hier rote Linien, die nicht überschritten werden sollten, skizziert werden können. Zudem gibt es mögliche außerplanmäßige Handlungsweisen, die sofort eröffnet werden, solche, die mit geringem (Weiterentwicklungs-)Aufwand eröffnet werden, und drittens solche, die erst mit größerem Weiterentwicklungsaufwand möglich werden. Weiters können auch durch Fehleranfälligkeit oder **Manipulierbarkeit** eröffnete Handlungsspielräume hier inkludiert werden.

Nachfolgende Tabelle gibt einen Überblick über naheliegende außerplanmäßige Handlungsweisen, und dient somit der Beantwortung von Unterfrage 2a. Zudem werden auch „moralisch bedenkliche“ Ausformungen oder Komponenten dieser Handlungsweisen aufgezeigt, und somit die Beantwortung von Unterfrage 2b durchgeführt.

Betroffene Human+ Komponente	außerplanmäßige Handlungsweise
Migrationsprognose	
Migrationsprognose gesamt	<p>Ergreifen von Maßnahmen auf Basis von durch die Human+ Technik bereitgestellten Migrationsprognosen, die aber nicht der Verbesserung der humanitären Versorgung von Schutzsuchenden dienen oder von der im Abschnitt „6.5.3 Mittelbarer Zweck – Argumente pro und kontra“ (vgl. S. 113ff) gemachten Klärung von „humanitärer Sicherheit“ abweichen [anderer mittelbarer Zweck]</p> <p>moralisch bedenklich: ja, wenn Zwecke angestrebt werden, die der Verbesserung der humanitären Versorgung zuwiderlaufen, z.B. frühzeitiges Schließen von Grenzen mit der Intention, damit eine Versorgung hinfällig zu machen, bis hin zu im schlechtesten Fall einer militärischen Verwendung von Prognosen um mit Waffengewalt Migrationsbewegungen zu unterbinden (Dual-Use-Problematik)</p>
	<p>Weitergabe von aufbereiteten Informationen zu Migrationsprognosen nicht an die dafür vorgesehenen und verantwortlichen BOS (bzw. deren dafür qualifizierte Unterstützungsgruppen wie ggf. VOST) [anderer mittelbarer Zweck]</p> <p>moralisch bedenklich: ja, besonders wenn die Weitergabe in unqualifizierte Hände erfolgt (z.B. gewaltbereite Rechtsradikale, parteipolitische AkteurInnen, usw.)⁶⁰</p>
	<p>Ermöglichung der Bedienung durch nicht-vorgesehene BedienerInnen [anderer unmittelbarer Zweck]</p> <p>moralisch bedenklich: ja, wenn diese nicht-vorgesehenen BedienerInnen zudem grundsätzlich abweichende Zwecke verfolgen, z.B. kommerzielle oder politische Zwecke</p>
speziell die Social-Media-Auswertung der Migrationsprognose	<p>Weitergabe von durch den Social-Media-Crawler gesammelten Social-Media-Daten, die persönliche Bezüge und Identifikation von Personen möglich machen, was besonders dann gravierend ist, wenn der Social-Media-Crawler nicht undifferenziert, sondern anhand einer Liste von Schlagwörtern sammelt (z.B. „refugees welcome“), wodurch die Identifizierung von Personen die zu diesen bestimmten Themen posten möglich wird [Verwendung des Crawlers für einen anderen unmittelbaren Zweck]</p> <p>moralisch bedenklich: möglicherweise, hängt von den Zwecken der Weitergabe ab, aber in jedem Fall würde dies eine Änderung des in Human+ angedachten Modus und somit einen function creep bedeuten</p>
	<p>Gezielte Identifizierung von Einzelpersonen, die in den analysierten Social Media (z.B. auf Twitter) solche Posts absetzen, die einen inhaltlichen Bezug zu Flucht und Migration haben, z.B. AktivistInnen oder Schutzsuchende selbst⁶¹ [anderer unmittelbarer Zweck]. Geringer Adaptionaufwand, da die Analysemöglichkeiten von Social-Media-Posts bereits für die Identifikation von Posts mit Migrationsbezug ausgelegt wurden.</p> <p>moralisch bedenklich: ja, da die Identifikation von Einzelpersonen kein Teil des ursprünglichen unmittelbaren Zwecks ist und insbesondere da die systematische Identifikation mit großem Missbrauchspotential einhergeht</p>
	<p>Ermöglichung von nicht in Human+ vorgesehenen „semantischen Analysen“ der angehäuften Big Data, um Informationen zu generieren, die „so“ nicht in den Daten enthalten sind, aber aus der Informationsgesamtheit abgeleitet werden können und zu anderen Zwecken genutzt werden können (z.B. Inferenzschlüsse auf die sexuelle Orientierung) [anderer unmittelbarer Zweck]. Größerer Adaptionaufwand, da die Entwicklung der Human+ Komponenten nicht auf die Anreicherung mit anderen Daten ausgelegt ist.</p> <p>moralisch bedenklich: eher ja, entspricht nicht dem ursprünglichen unmittelbaren Zweck, und die Adaption der Human+ Algorithmen zur Generierung weiterer persönlicher Daten birgt großes Missbrauchspotential</p>
In-Situ-Daten (UAV, Sensortürme, Kamerasysteme)	<p>Ausspähen der lokalen Bevölkerung durch etwaige installierte Sensortürme oder Drohnenaufklärung [anderer unmittelbarer Zweck]</p> <p>moralisch bedenklich: ja, weil die Anwendung der Human+ Technik für die Bevölkerungsüberwachung einen völligen Kontextwechsel und function creep bedeutet, und eine solche Anwendung erst einer speziellen Legitimation und Klärung der dann neuen Anwendungsbedingungen bedarf</p>
	<p>Manipulation der in das Human+ System einfließenden In-Situ-Informationen, um strategische Zwecke zu erreichen (z.B. Über- oder Untertreibung von Migrationsbewegungen) [Manipulation]</p> <p>moralisch bedenklich: ja, weil gegen Auflagen der Datenintegrität verstoßen wird. Hierbei handelt es sich eher um eine Form von Cyberkriminalität.</p>
Social-Media- und In-Situ-Daten in Kombination	<p>Wenngleich sehr unwahrscheinlich und praktisch kaum möglich: Bei entsprechend genauer Preisgabe des aktuellen Aufenthaltsorts einer Social-Media-UserIn über die Social-Media-Aktivitäten und gleichzeitigem Vorhandensein von In-Situ-Sensoren am Aufenthaltsort wäre es zumindest theoretisch denkbar, bei rechtzeitiger Sichtung der Social-Media-Daten diese Daten mit In-Situ-Daten anzureichern, um Bildinformationen über den Autor zu erhalten. [anderer mittelbarer Zweck]</p> <p>moralisch bedenklich: ja, da die Identifikation von Einzelpersonen kein Teil des ursprünglichen unmittelbaren Zwecks ist und die systematische Identifikation mit großem Missbrauchspotential einhergeht</p>

Human Sensors Daten	Manipulation der Human-Sensors-Informationen um strategische Zwecke zu erreichen (z.B. Über- oder Untertreibung von Migrationsbewegungen) [Manipulation] moralisch bedenklich: ja, da bewusst gegen Auflagen der Integrität verstoßen wird
Satellitenfernerkundungsdaten	Weitergabe von Satellitenfernerkundungs-Auswertungen an EmpfängerInnen abseits der bei der Bewältigung von Flüchtlingslagen für die Sicherstellung humanitärer Sicherheit notwendigen BOS [anderer mittelbarer Zweck] moralisch bedenklich: hängt von den EmpfängerInnen ab, birgt aber jedenfalls Risiken und ist jedenfalls dann moralisch bedenklich, wenn die EmpfängerInnen dabei Vertraulichkeitsaspekte verletzen und dadurch Personen in Flüchtlingslagern und dergleichen zusätzlichen Risiken ausgesetzt werden
gemeinsamer Echtzeitlageüberblick	
gemeinsamer Echtzeitlageüberblick gesamt	Weitergabe von Informationen aus dem Echtzeitlagebild an ungeeignete EmpfängerInnen (z.B. Medien, gewaltbereite Gruppen, aber auch parteipolitische Akteure, etc.) [anderer mittelbarer Zweck] moralisch bedenklich: ja, da diese Informationen zweckentfremdet werden können und dann Schadenspotentiale ermöglichen (z.B. können feindselig gesinnte Akteure gezielt dort Sabotage betreiben, wo der Echtzeitlageüberblick wenig BOS-Ressourcen ausweist), sowie eine solche Weitergabe das Vertrauen der beteiligten BOS in das Human+ System schädigt und damit der gemeinsame Echtzeitlageüberblick nicht mehr bespielt wird
	bewusstes Einspielen von Falschinformationen in das Echtzeitlagesystem [Manipulation] moralisch bedenklich: ja, weil gegen Auflagen der Datenintegrität verstoßen wird
	Strategische Manipulation von Informationen des Echtzeitlagesystems (z.B. durch missbräuchliche Verwendung des Sichtungstools) [Manipulation] moralisch bedenklich: ja, weil gegen Auflagen der Datenintegrität verstoßen wird
	Identifikation einzelner Flüchtlinge (was die Gestaltung des Human+ Systems auf eine Weise voraussetzt, dass solche Informationen im Echtzeitlagebild angezeigt werden, die eine tatsächliche Identifikation auch ermöglichen) [anderer unmittelbarer Zweck] moralisch bedenklich: ja, insofern die Technologie nur über Gruppen Auskunft geben soll und bei individuellen Informationen der Schutz persönlicher Informationen weniger gegeben erscheint; wenn es um individuelle Personen geht, sollten die etablierten Informations- und Kommunikationswege der beteiligten BOS eingehalten werden)

Tabelle 12: Überblick über außerplanmäßige Handlungen

6.6.2 Maßnahmen zur Verhinderung außerplanmäßiger Handlungen

Für die Beantwortung von Hauptfrage 2, ob alle Maßnahmen ergriffen werden, um obige angeführte moralisch bedenkliche außerplanmäßige Handlungen zu verhindern, ist zunächst vorzuschicken, dass prinzipiell nie „alle“ Maßnahmen ergriffen werden können. Bei der Human+ Technik handelt es sich nicht um eine „abgegrenzte Sachtechnik“, für welche etwa gezielt technische Vorkehrungen getroffen werden können, sondern bezogen auf den Einsatz um ein „sozio-technisches System“, welches im Falle einer hypothetischen Anwendung in soziale Systemzusammenhänge eingebunden wäre (Ropohl 2016: 12). Da es als solches einer gesellschaftlichen Prägung und Gestaltung sowohl fähig wie auch bedürftig ist (vgl. Simonis 2013a: 12), dürften nicht allein technische, sondern müssen nach Ansicht des Verfassers der vorliegenden Arbeit auch soziale Vorkehrungen in den Blick genommen werden. Grundsätzlich lassen sich also zwei Schienen von „Gegenmaßnahmen“ zu moralisch bedenklichen, au-

⁶⁰ Bei diesem wie bei ähnlichen Punkten ist auch jeweils eine „interne“ Verfehlung denkbar, d.h. dass an sich berechnete Personen missbräuchliche Handlungen setzen, oder aber ein Zugriff von außen erfolgt, und etwa Manipulationen des Systems durch Einspielen von Falschinformationen durch unberechtigte Dritte erfolgen.

⁶¹ Beispiele hierfür finden sich etwa in der Türkei, wo gezielt Twitter-UserInnen identifiziert wurden, die Aktivismus im Rahmen von Protesten gezeigt hatten (vgl. WinFuture 2013).

ßerplanmäßigen Handlungen unterscheiden: **technische Maßnahmen** und **organisatorische Maßnahmen**. Da die Human+ Technik im Rahmen des Projekts auf TRL4 entwickelt wird, werden sowohl bereits angedachte wie auch für eine Weiterentwicklung auf ein anwendungsnahes TRL naheliegende technische Maßnahmen betrachtet, bzw. Empfehlungen für den organisationalen Einsatz gegeben.

Technische Maßnahmen

- Trennung von „Rohdaten“ und „Informationsprodukt“: Wie in Tabelle 12 dargestellt, beziehen sich zahlreiche moralisch bedenkliche außerplanmäßige Handlungen (z.B. Weitergabe von gesammelten Social-Media-Daten, bzw. Identifizierung von Personen die zu Migrationsthemen in Social Media aktiv sind) darauf, dass auf die gecrawlten Social-Media-Daten zugegriffen wird. Technische Vorkehrungen hierfür, wie sie auch in den Konzepten zur Systemarchitektur von Human+ vorgesehen sind, zielen auf eine Trennung von Rohdaten und dem erst aus ihnen destillierten Informationsprodukt ab. Während die Rohdaten (z.B. gesammelte Tweets) angereichert sind mit Informationen zum Verfasser und weiteren Metainformationen, enthält das daraus erstellte Informationsprodukt nur noch aggregierte Informationen, zum Beispiel Zahlenangaben auf einer Geoinformationskarte. Eine Zwischenrolle nimmt das in der Systemarchitektur vorgesehene sogenannte „**Sichtungstool**“ ein: Im Human+ Demonstrationssystem war dies ein Werkzeug, mit dem die Informationsinputs (z.B. gesammelte Social-Media-Daten, Meldungen von Human Sensors, etc.) vom sogenannten „**Sichter**“, nach Relevanz und Vertrauenswürdigkeit eingeschätzt und so für die weitere Auswertung bzw. die Weiterreichung an den Einsatzstab, der das Human+ System bedient, als ausreichend geeignet bzw. nicht-geeignet eingestuft werden konnten. Während Einzelpersonen (z.B. eben die genannten SichterInnen) somit je nach konkreter Ausgestaltung des Systems noch auf Rohdaten zugreifen können, bleibt die Mehrheit der Personen, die die Human+ Technik verwenden, auf das bereits prozessierte Informationsprodukt verwiesen. Während der Projektlaufzeit wurde bereits die Notwendigkeit diskutiert, ob SichterInnen überhaupt Rohdaten wie „Echt-Tweets“ (mit Namen, ggf. Bild des Verfassers des Tweets, dem Postingtext selbst, geographische Position) in vollem Umfang zur Beurteilung des Tweets angezeigt bekommen müssen, da dies eben potentiell die Identifizierbarkeit von zu diesem Thema postenden Twitter-UserInnen erlaubt. Aus Sicht der am Human+ Projekt beteiligten Einsatzkräfte war dies aber eine wichtige Voraussetzung für die Beurteilung von Relevanz und Plausibilität der auf diesen gesammelten Daten aufbauenden Prognosen. Da zu den in Human+ entwickelten Prognosealgorithmen noch keine Erfahrungswerte vorliegen, konnte während der Projektlaufzeit darüber kein abschließendes Urteil gefällt werden. Falls sich der Output der Prognosen entsprechend bewährt, könnte eine Einblendung von Originaltexten zwecks Sichtung dann womöglich entfallen, auch wenn sich dies negativ auf die Nachvollziehbarkeit der Informationen auswirkt.

- Beschränkte Speicherung von Rohdaten: Während im Human+ Forschungsprojekt die gesammelten Rohdaten (da sie auch für andere Forschungsprojekte verwendet werden können) vorgehalten wurden, ergaben Diskussionen mit den im Projekt beteiligten BOS, dass eine dauerhafte Speicherung nicht notwendig wäre, d.h. nach Extraktion der Prognosen bzw. Bewältigung der Lage können diese gelöscht werden. Eine automatisierte Speicherung im Rahmen von Privacy-by-Design-and-by-Default wäre somit prinzipiell möglich.
- Beschränkung von Informationen im gemeinsamen Echtzeitlagebild: Während zur Unterstützung der Aufgabenbewältigung für die BOS eine Vielzahl an Informationen im durch Human+ erzeugten Echtzeitlagebild möglich sind, wurde schon während der Projektlaufzeit beschlossen, keine individuellen Daten zu Geflüchteten zu eruieren und zu erfassen, sondern nur aggregierte Daten, die sich z.B. auf Gruppen von Geflüchteten beziehen, über das Lagebild auszutauschen. Dies verhindert, dass Einzelpersonen über Human+ identifiziert oder potentiell sensible Daten über das System angezeigt werden.
- Rechte- und Rollenmanagement: Im Pflichtenheft zum Human+ Demonstratorsystem (vgl. Human+ 2018d: 74f) wurde ein (TRL4 angemessenes) rudimentäres Rechte- und Rollensystem vorgesehen, welches das Anlegen von unterschiedlichen System-User-Typen erlaubt, welche dann mit für diese AnwenderInnen-Rolle maßgeschneiderten Systembefugnissen ausgestattet werden können. Diese Funktionalitäten erlauben es etwa, die Rolle des oben genannten „Sichters“ anzulegen, und so die Trennung von Rohdaten und Informationsprodukt für den überwiegenden Teil der AnwenderInnen des Systems sicherzustellen. Außerdem wird damit der allgemeine Systemzugriff auf Personen mit angelegten Userprofilen eingeschränkt. Zudem wurde die Berücksichtigung einer sogenannten „Behördenkennung“⁶² für datenschutzrechtliche Aspekte vorgesehen, sodass Zugriffsmöglichkeiten im Falle eines interorganisationalen Human+ Betriebes auch über verschiedene BOS hinweg differenziert gestaltet werden können. Für ein praxisbetriebstaugliches, höheres TRL können diese Komponenten weiter ausgebaut werden.
- Nachvollziehbarkeit der Systemverwendung: Um das Einspielen von Falschinformationen in unterschiedliche Datenquellen (z.B. gefälschte Human-Sensors-Meldungen) oder die missbräuchliche Verwendung der Sichtungstools zu erschweren oder zumindest nachvollziehbar zu machen, wurde während des Projekts auch die Möglichkeit diskutiert, eine Art Log-File zu führen, welches unterschiedliche UserInnen-Aktivitäten im System verzeichnet und protokolliert. Zumindest bei Weiterentwicklung in Richtung Anwendungsreife kann dies für weitere Transparenz sorgen, und außerdem in der Aufarbeitung des Einsatzes im Nachhinein unterstützen.

⁶² Eine 13 stellige, von der Firma CrisCom erstellte, proprietäre Behördenkennung, welche eine BOS eindeutig nach Staat, Bundesland, Organisation etc. identifiziert.

- Systementwicklung gemäß den „guten Standards“ der IT-Sicherheit: Als letzter Punkt ist hier die Einhaltung von guten Standards der IT-Sicherheit (z.B. Data-Protection-by-Design-and-by-Default) bei der (Weiter-)Entwicklung der Komponenten zu nennen, um unberechtigte Zugriffe durch Ausnutzen von Systemschwächen oder Programmierungslücken zu vermeiden.

Während obige technische Maßnahmen das Human+ System als „technisches“ System in seinem Missbrauchspotential begrenzen, müssen organisationale Maßnahmen dazu beitragen, es auch als „sozio-technisches“ System ethisch verträglich zu gestalten. Diese Maßnahmen können als „soziale Fest-schreibungen“ (Fenner 2010: 230) betrachtet werden, die die Anwendung auf bestimmte, ethisch legitime Anwendungsformen „fixieren“.

Organisationale Maßnahmen:

- Beschränkter AnwenderInnenkreis: Da sich gezeigt hat, dass eröffnete, zweckentfremdete Anwendungsmöglichkeiten nur begrenzt durch technische Maßnahmen verhinderbar sind, stellen sich besondere Anforderungen an die Personen, die das Human+ System oder seine Komponenten verwenden. Eine Möglichkeit, der Erfüllung dieser Anforderungen Genüge zu tun, besteht in der Beschränkung des AnwenderInnenkreises auf Personen, die im Rahmen von BOS oder in speziellen Konstellationen (wie VOST, oft mit gesonderter Sicherheitsüberprüfung) tätig sind. So sieht etwa das in Abstimmung mit den BedarfsträgerInnen (d.h. den involvierten BOS) des Projekts entwickelte Pflichtenheft folgende AnwenderInnen vor: „Die Zielgruppen sind die Bevölkerungsschutz- und Katastrophenschutz-Einheiten und Behörden auf Kreisebene, Bezirksebene und Landesebene, dazu die einsatzdurchführenden BOS wie Feuerwehr, Polizei, Rettungs- und Sanitätsdienst, THW und organisierte ungebundene Helfergruppen (OUH) [...]“ (Human+ 2018d: 20). Während eine solche BenutzerInnen-Vorauswahl mögliche Verfehlungen nicht im Vorhinein völlig ausschließt, arbeiten MitarbeiterInnen in BOS bereits jetzt in Einsätzen mit potentiell vertraulichen Informationen und sind daher speziellen Überprüfungen und Dienstvorschriften unterworfen. Gleichwohl gibt es Systemkomponenten, z.B. das Sichtungstool, für dessen Bedienung noch verschärfte Anforderungen formuliert werden können, z.B. in höherem Maße politische Neutralität, da über das Sichtungstool entschieden wird, welche Informationen Bestandteile des gemeinsamen Lagebildes von Human+ werden. Von der Verwendung durch reine Privatpersonen oder Privatorganisationen, die nicht wie BOS einem entsprechenden Reglement unterworfen sind, wird daher abgeraten. Durch das schon oben erwähnte Rechte- und Rollensystem kann hier auch in Abstufungen vorgegangen werden, sodass z.B. sogenannte organisierte ungebundene Helfergruppen in Form von Human-Sensors zwar Informationen in das System einspeisen können (welche dann noch von SichterInnen beurteilt werden), aber nicht selbst direkt Informationsprodukte abrufen können.

- Code-of-Conduct (COC) für AnwenderInnen: Neben der Auswahl entsprechenden Personals wäre die Formulierung eines COC eine Möglichkeit, die Verwendung der Human+ Techniken in gewünschte Bahnen zu lenken. Ein solcher COC sollte die AnwenderInnen über die „legitimen“ Anwendungszwecke der Technik informieren, und Beispiele für klar davon abweichende Anwendungen aufzeigen. Ein solcher COC kann auch Kriterien bzw. Auflistungen enthalten, unter welchen Bedingungen an welche Institutionen oder Personen Informationen, die über das Human+ System generiert wurden, weitergegeben werden dürfen.

Da das Human+ Forschungsprojekt ein relativ niedriges TRL anvisierte und somit nahe an einem Grundlagenforschungsprojekt war, können obige Punkte nur skizzenhaft aufzeigen, welche Gegenmaßnahmen zur Verhinderung außerplanmäßiger Anwendungen möglich sind – zumal die Weiterentwicklung auf ein höheres TRL nicht im Projektverbund stattfindet, sondern ggf. durch diejenigen Projektpartner, die für die einzelnen Komponenten im Rahmen des Human+ Forschungsprojekts verantwortlich waren. Für die Beantwortung der Hauptfrage 2 bedeutet dies, dass diese Frage nicht klar bejaht werden kann. Es kann nicht garantiert werden, dass „alle“ Maßnahmen zur Verhinderung von Missbrauch getroffen werden, speziell angesichts der Tatsache, dass auch der spätere Einsatzkontext (welche BOS verwenden welche Human+ Komponenten) nicht genau festgelegt ist. Es konnte aber aufgezeigt werden, dass eine Kombination aus technischen und organisationalen Maßnahmen dabei helfen kann, das System ethisch verträglicher zu gestalten und die in obiger Tabelle angeführten außerplanmäßigen Handlungsweise zu erschweren oder sogar zu verhindern.

6.7 HAUPTFRAGE 3 (FOLGEN UND NEBENFOLGEN DER PLANMÄßIGEN ANWENDUNG) – SCHRITT 2 BIS 4

6.7.1 Argumente pro und kontra

Hauptfrage 3 lautet: *„Gefährden die Folgen und Nebenwirkungen der planmäßigen Technikenutzung die Lebensbedingungen oder die Lebensqualität gegenwärtiger oder zukünftiger Menschen?“*

Um die Gefährdung für Lebensbedingungen und die Lebensqualität strukturiert zu beleuchten, wird an dieser Stelle eine Ethical Matrix erstellt, die zeigt, ob vor dem Hintergrund der oben aufgestellten Prinzipien mittlerer Reichweite für unterschiedliche Stakeholdergruppen ethische Konflikte bzw. potentielle Beeinträchtigungen durch die Human+ Komponenten auftreten. Die im Hintergrund stehende Überlegung ist somit die, dass im Falle von absehbaren moralischen Konflikten auch in einem abstrakten Sinne Gefährdungen vorliegen können (v.a. bei Konflikten vor dem Hintergrund des Prinzips der

Nichtschädigung). Um zudem ein möglichst abgerundetes Bild für etwaige Gewichtung-Angelegenheiten zu erhalten, werden nicht nur negative, sondern auch (weitere) positive Auswirkungen in der Ethical Matrix verzeichnet. Um zu veranschaulichen, ob es sich um die Berücksichtigung eines Prinzips [grün] oder den Verstoß gegen ein Prinzip [rot] handelt, werden die entsprechenden Passagen zusätzlich eingefärbt. Ein Hinweis: Argumente, die bereits bei der Behandlung von möglichen Missbrauchsmöglichkeiten (Hauptfrage 2) erläutert wurden, werden hier nicht noch einmal angeführt, da es in diesem Abschnitt um die Folgen und Nebenwirkungen der **planmäßigen** Techniknutzung geht.

Einige der angeführten Argumente betreffen zudem mehrere Stakeholder-Gruppen in ähnlicher Weise simultan, sodass einem methodischen Ratschlag von Bleisch und Huppenbauer (2014: 51) gefolgt wird und das Argument übergreifend für die betroffenen Stakeholder-Gruppen gemeinsam formuliert wird. Dementsprechend scheint es in den passenden Zellen der Ethical Matrix auch mehrfach auf.

Stakeholder \ Prinzip	Respekt vor der Autonomie (Prinzip der Selbstbestimmung)	Prinzip der Nichtschädigung	Prinzip der Fürsorge	Prinzip(ien) der Gerechtigkeit
Geflüchtete („Flüchtlinge“)		Kontra-Argument 1: potentielle Verunsicherung bei der Smartphone- und Social-Media-Verwendung von Geflüchteten durch eventuelle Befürchtung einer weiteren Maßnahme der „digitalen Überwachung“	Pro-Argument 1: höhere Chance rechtzeitiger und angemessener Versorgung	
Zivilbevölkerung			Pro-Argument 2: Bessere Informationen zu Flüchtlingsbewegungen und Ankunftszeiten, dadurch bessere Ressourceneffizienz bei BOS	Pro-Argument 3: Zivilbevölkerung finanziert im Wesentlichen über Steuern auch die technische Ausstattung von BOS (und damit auch Human+); Rekompensation durch die Arbeit der Behörden
In-Situ-Betroffene	Kontra-Argument 2: Privatheitsbedrohung durch verwendete Technik Kontra-Argument 3: kein „informed consent“ oder „opt-out“ möglich, daher potentielle Verletzung der „Verhaltensautonomie“ der Betroffenen	Kontra-Argument 4: potentieller Schaden besteht in „Normalisierung“ des Verhaltens bei erkennbaren Überwachungstechnologien		
Social-Media-UserInnen	Kontra-Argument 3: Privatheitsbedrohung durch verwendete Technik Kontra-Argument 5: kein „informed consent“ für die Verwendung der erzeugten Social-Media-Daten, kein „opt-out“ möglich, weil Möglichkeit weder gewährt wird noch die Human+ Technik überhaupt in einer Weise bekannt werden dürfte, dass „opt-out“ realistisch wäre	Kontra-Argument 6: Schaden im Sinne des Kontrollverlusts über die Datenverwendung		Kontra-Argument 7: keine „Kompensation“ für das Erzeugen von Daten für die Social-Media-basierten Komponenten
NGOs und NGO-MitarbeiterInnen		Pro-Argument 4: NGO-MitarbeiterInnen und NGOS profitieren in der Zusammenarbeit mit BOS ebenfalls von genaueren Informationen		
BOS-MitarbeiterInnen	Pro-Argument 5: Bedarfsträger-Interviews in AP2.AT-DE zeigten den erwarteten Mehrwert und den Wunsch nach einer Technik, die durch Migrationsprognosen und einen besseren Informationsaustausch bei der Bewältigung von Lagen unterstützt		Pro-Argument 2: Bessere Informationen zu Flüchtlingsbewegungen und Ankunftszeiten, dadurch bessere Ressourceneffizienz bei BOS	
AnwenderInnen von Human+	Pro-Argument 5: Bedarfsträger-Interviews in AP2.AT-DE zeigten den erwarteten Mehrwert und den Wunsch nach einer Technik, die durch Migrationsprognosen und einen besseren Informationsaustausch bei der Bewältigung von Lagen unterstützt Kontra-Argument 8: Gefahr von „Sachzwängen“ durch verordnete Anwendung der Human+ Technik	Kontra-Argument 9: bei wenig intuitiver Bedienung oder hoher Lernkurve zumindest (zunächst) ein Mehraufwand beim Einlernen oder Anwenden des Systems [Kompensation: einfache Anwendbarkeit anstreben] Kontra-Argument 10: Einsatz des Systems muss auf der richtigen Ebene erfolgen, da ansonsten „Ressourcenabzüge“ zu erwarten sind		
Human Sensors – OUH	Kontra-Argument 8: Gefahr von „Sachzwängen“ durch verordnete Anwendung der Human+ Technik positiv gewendet steht aber ein zusätzliches Hilfsmittel zur Verfügung			
Human Sensors – BOS	Kontra-Argument 8: Gefahr von „Sachzwängen“ durch verordnete Anwendung der Human+ Technik positiv gewendet steht aber ein zusätzliches Hilfsmittel zur Verfügung			

Tabelle 13: Ethical Matrix zu Folgen und Nebenfolgen der planmäßigen Technikverwendung

Kontra-Argumente:

Kontra-Argument 1: Argument der Verunsicherung: Verunsicherung bei Geflüchteten durch eine zusätzliche Analysetechnik

Das Smartphone hat für zahlreiche Geflüchtete einen besonderen Stellenwert, wie Forschungen zeigen. So dient es etwa als wichtiges Werkzeug, um sich während der irregulären Migration von Menschenschmugglern (bzw. der Notwendigkeit, sich auf deren Informationen zu verlassen) unabhängiger zu machen (vgl. Dekker et al 2018: 7). Zudem besteht auf Seiten der Geflüchteten ein Bewusstsein dafür, dass Behörden sich durchaus Methoden der „digital surveillance“ bedienen können, um den Grenzschutz durchzusetzen. Human+ ist zwar nicht primär als Überwachungstechnik ausgelegt (da etwa keine Smartphones lokalisiert werden und kein individuelles Tracking erfolgt), dennoch können die Komponenten, die auf eine Analyse des Social-Media-Verkehrs hinauslaufen (bei welcher etwa nach Möglichkeit auch die Georeferenzierung der Posts ausgelesen wird) als eine solche Technik (miss-)verstanden werden. Das heißt, Human+ könnte bei einem entsprechenden Publikwerden potentiell verunsichernd auf Geflüchtete wirken, was in jedem Fall eine nicht-intendierte Nebenfolge der planmäßigen Human+ Anwendung wäre.

Kontra-Argument 2: Privatheitsbedrohung durch verwendete Technik

Der In-Situ-Teil und die Social-Media-Auswertungen im Rahmen von Human+ können als potentielle Bedrohung für die Privatsphäre der durch diese Technik betroffenen Personen betrachtet werden. Während In-Situ-Aufnahmen (z.B. durch Kameras oder UAV) durch die Aufnahmen selbst bei ausreichender Auflösung etwa Gesichter, Nummernschilder, und dergleichen erkennen lassen, besteht die Privatheitsbedrohung bei der Auswertung von Social-Media-Daten (z.B. der im Projekt vornehmlich angepeilten Auswertung von Tweets) darin, dass persönliche Daten wie der Benutzername des Profils, das Benutzerbild, die geographische Position des Tweets, und natürlich der Tweetinhalt selbst womöglich Rückschlüsse auf den/die VerfasserIn des Tweets zulassen.

Kontra-Argument 3: Argument der Verletzung der Verhaltensökonomie von In-Situ-Betroffenen

Ein Teil der Human+ Komponenten bezieht sich für die Anreicherung des gemeinsamen Echtzeitlagebildes für BOS auch auf die Auswertung von sogenannten In-Situ-Informationsquellen, also z.B. Videoüberwachungskameras⁶³ auf öffentlichen und für die Einsatzlage relevanten Plätzen. Dies wäre beispielsweise während der „Flüchtlingslage 2015/2016“ der Westbahnhof gewesen (wo schon eine Infrastruktur aus Videoüberwachungskameras vorhanden ist). Eine andere beispielhafte Möglichkeit wäre für staatsgrenznahen Bereiche auch Kamerateürme, oder auch die Aufnahme von Videos aus der Luft durch UAV („Flugdrohnen“), die durch den Einsatz von Algorithmen hinsichtlich der Zahl und Bewegungsrichtung der aufgenommenen Personen ausgewertet werden können, um zu beantworten, wieviele Menschen sich auf der anderen Seite der Grenze (falls die Übergänge geschlossen sind) stauen, und wieviele Menschen in einer gewissen Zeit einen bestimmten Grenzübergang übertreten haben (falls die Grenzen geöffnet sind).⁶⁴ Während es sich auf den ersten Blick um heterogene Datentypen handelt, ist ihnen allen gemein, dass es sich dabei um Informationsquellen handelt, die (anders als etwa Social-Media-Auswertungen) zumindest potentiell die lokal ansässige Bevölkerung der damit abgedeckten Gebiete betreffen kann (die etwa in grenznahen Wohnbereichen auch auf UAV-Aufnahmen abgebildet sein können). In Human+ wurden in erster Linie Algorithmen für die Auswertung dieser Datentypen erforscht (z.B. für die Analyse von Videoaufnahmen von UAV oder festinstallierten Kameras, um die Anzahl von Personen, die ein bestimmtes Gebiet passiert haben, abzuschätzen, d.h. für Bewegungsstromanalysen). Für dieses Argument wird angenommen, dass bei einer Weiterentwicklung des Systems tatsächlich auf solche Bilddaten zugegriffen werden kann oder entsprechende UAV-Einheiten eingesetzt werden. Grundsätzlich können solche Aufnahmen entweder auf **transparente bzw. offene Weise** erfolgen (z.B. indem UAV-Einsätze angekündigt werden, oder im Falle von Videoüberwachung entsprechende Schilder und gut sichtbare Kameras aufgestellt werden), oder aber auf **verdeckte Weise** (bei Fehlen dieser Maßnahmen, sodass der Umstand, in einem Gebiet gefilmt zu werden, den Betroffenen nicht bekannt wird). Das vorliegende Argument bezieht sich auf den Fall einer verdeckten Datenaufnahme, und besagt, dass Betroffene der In-Situ-Aufnahme durch deren verdeckte Natur keine Möglichkeit haben, der Betroffenheit durch diese Informationsquellen (informiert) zuzustimmen

⁶³ Zu diesen Kamerasystemen müssten zunächst einmal die notwendigen Schnittstellen geschaffen werden. Im Hintergrund der Ausführungen zu den Kameras stehen somit eher grundsätzliche Überlegungen, die eine Weiterentwicklung des Projekts betreffen. Während des Human+ Projekts selbst wurden keine Schnittstellen geschaffen, sondern es wurde an Algorithmen geforscht, die die Auswertung von Kamera-Streams ermöglichen würden, wenn solche Schnittstellen verfügbar wären.

⁶⁴ In einer Human+ Besprechung, in der auch die BedarfsträgerInnen des Projekts vertreten waren, wurde anekdotisch auf die Bedeutung von Einschätzungen von „erhöhten Positionen“ während der Situation 2015/2016 an den österreichischen Grenzübergängen verwiesen: dort habe es einen Rückstau von Menschen auf der ungarischen Seite der Grenze gegeben, und von österreichischer Seite musste das Grenzpersonal auf die Dächer von LKWs und dergleichen klettern, um einen erhöhten Blick „über“ die Grenze hinweg zu bekommen und um abschätzen zu können, wie weit der Rückstau von Personen, die über die Grenzübergänge wollten, ins ungarische Hinterland hineinreicht.

oder sich dagegen auszusprechen, bzw. ihr Verhalten an den Umstand des Gefilmt-Werdens anzupassen. Dies bedeutet, dass zwar keine „Verunsicherungen“ durch die Datenaufnahmen auftreten (wie bei einer offenen Datenerhebung), aber „[d]ie Unwissenheit über stattfindende Überwachungsmaßnahmen kommt somit einer Täuschung gleich. Diese verletzt die Verhaltensautonomie der Betroffenen, da sie keine Möglichkeit haben, selbstbestimmt auf die Überwachungssituation zu reagieren“ (Krüger 2018: 21). Während man beim Wissen, gefilmt zu werden, sich prinzipiell dann im eigenen Verhalten danach ausrichten kann, fehlt diese Möglichkeiten bei verdeckten Methoden. Hierbei handelt es sich somit um eine Beeinträchtigung des Prinzips des Respekts vor der Autonomie.

Kontra-Argument 4: Argument der Normalisierung des Verhaltens

Dieses Argument bezieht sich ebenfalls auf die Gruppe der In-Situ-Betroffenen, allerdings nicht mehr vor dem Hintergrund des Prinzips des Respekts vor der Autonomie, sondern hier wird eine Verletzung des Prinzips der Nichtschädigung angenommen. Wenn nämlich statt einer verdeckten Datenerhebung durch die In-Situ-Datenquellen eine offene Datenerhebung vorgenommen wird, stellt sich die Problematik einer sogenannten „**Normalisierung**“ des Verhaltens. Normalisierung meint, dass bereits das bloße Vorliegen von erkennbaren Überwachungstechnologien (wie beispielsweise Kameras) dazu führen kann, dass sich Menschen in ihrem Verhalten danach ausrichten, was sie für das gesellschaftlich erwünschte Verhalten erachten (vgl. Koch 2014: 169). Bei Videoüberwachung „[...] besteht etwa die Möglichkeit, dass nicht nur kriminelles Verhalten oder das Verhalten, welches als Indiz für kriminelles Verhalten gelesen wird, unterbunden wird, sondern generell jedes sozial auffällige Verhalten“ (Koch 2014: 169).

Werden Menschen beobachtet oder fühlen sich zumindest beobachtet, so verstärkt dies den Anpassungsdruck (vgl. Koch 2014: 175). Der Umstand der Beobachtung selbst kann die Selbstaufmerksamkeit steigern und die Aufmerksamkeit darauf richten, ob den Normalitätsstandards genügt wird oder nicht. Gleichzeitig können wachgerufene Verhaltenserwartungen diffus sein, sodass verschiedenste Normalitätserwartungen abgerufen werden können, die auch von den UrheberInnen dieser Überwachungs- bzw. Beobachtungsmaßnahmen nicht erwünscht sein müssen.

Zudem wird Videoüberwachung als häufiges Instrument von solchen Regierungen eingesetzt, deren Legitimität von der eigenen Bevölkerung nicht breit akzeptiert wird und somit nicht auf die Loyalität der BürgerInnen setzen können (vgl. von Silva-Tarouca Larsen 2011: 87). Sichtbare Videoüberwachung wird dort eingesetzt, um durch die Angst, „erwischt zu werden“ die Verhaltensübereinstimmung der BürgerInnen mit unterdrückenden Gesetzen zu erzwingen. Einschüchterung durch Beobachtung des Verhaltens und Zwang (da Ungehorsam entdeckt und sanktioniert wird) sind so Begleiterscheinungen

von Videoüberwachung in bestimmten politischen Settings. Während dies für österreichische Verhältnisse zwar nicht zutrifft, kann sichtbare Videoüberwachung womöglich solche Assoziationen bei denjenigen Personen auslösen, die aus einem unterdrückerischen Regime nach Österreich eingereist sind, sodass Normalisierung womöglich auch in Österreich bei diesen Personen einschüchternd wirken kann.

Während Normalisierung gesamtgesellschaftlich wünschenswerte Auswirkungen haben kann (z.B. Normalisierung dahingehend, dass kein kriminelles Verhalten an den Tag gelegt wird), stellt sich bei der ethischen Beurteilung von Normalisierung die Frage, welche Art von Verhalten normalisiert wird, z.B. nur kriminelles⁶⁵ oder bereits lediglich auffälliges Verhalten? Normalisierung, so das Argument, schränkt Selbstbestimmungsmöglichkeiten und freie Persönlichkeitsentfaltung von Personen ein, ohne dass diese durch ihr Handeln die Sicherheit von anderen Personen gefährden, und die Verwendung von Kameras kann zu dieser Normalisierung führen.

Kontra-Argument 5: Argument der zweifelhaften Einwilligung von Social-Media-UserInnen:

Ein Teil der Human+ Technik zielt auf die Auswertung von Social-Media-Posts ab. Diese werden sowohl hinsichtlich quantitativer Kriterien (z.B. „Anomalien“ hinsichtlich der Postinghäufigkeit je nach geographischer Region) als auch qualitativer Kriterien analysiert (z.B.: Werden Migrations- oder Flucht-bezogene Termini, z.B. „Flüchtlinge“ verwendet?). Im Human+ Forschungsprojekt wurde als Referenz-Plattform für die Forschungsarbeiten Twitter gewählt. Mittels eines Crawlers, welcher über ein API („application programming interface“) eine offizielle Schnittstelle zu Twitter verwendet, kann ein bestimmter Prozentsatz von (als öffentlich deklarierten) Tweets in einer Datenbank für die weitere Verwendung abgespeichert werden. Zumindest auf der Ebene dieser Rohdaten werden u.a. die Tweet-ID, der Profilname derjenigen Person, die den Tweet absetzt, etwaige Geo-Daten, Datum und Uhrzeit des Posts sowie Inhalt und etwaige beigefügte Bilder abgespeichert (vgl. Human+ 2018d: 89). Twitter-UserInnen haben nicht ausdrücklich einer Datensammlung durch die Human+ Technik zugestimmt, sehr wohl aber setzt die Nutzung von Twitter die Zustimmung zu den AGB voraus, in denen es u.a. heißt:

„Durch Übermittlung, Veröffentlichung oder Anzeigen von Inhalten auf oder über die Dienste gewähren Sie uns eine weltweite, nicht ausschließliche, unentgeltliche Lizenz (mit dem Recht zur Unterlizenzierung), diese Inhalte in sämtlichen Medien und über sämtliche Verbreitungswege (die gegenwärtig bekannt sind oder in Zukunft bekannt

⁶⁵ Selbst bei der Normalisierung von kriminellem Verhalten, worauf die Human+ Technik in keiner Weise abzielt, stellen sich Probleme der Vermischung von Moralität, Normativität und Kriminalität, d.h. dass das Verhalten als wünschenswert betrachtet wird, von dem angenommen wird, dass es „normal“ ist – während aber das, was als normal angenommen wird, auch stark davon abhängt, was als wünschenswert gilt (vgl. Koch 2014: 169).

sein werden) zu verwenden, zu vervielfältigen, zu reproduzieren, zu verarbeiten, anzupassen, abzuändern, zu veröffentlichen, zu übertragen, anzuzeigen und zu verbreiten. Mit dieser Lizenz erteilen Sie uns die Erlaubnis, Ihre Inhalte weltweit verfügbar zu machen und dies auch Dritten zu ermöglichen.“ (Twitter AGB 2019).

Während Twitter-UserInnen somit keiner spezifischen Nutzung durch Human+ zugestimmt haben, haben sie einer allgemeinen Nutzungserlaubnis in den AGB zugestimmt. Heißt das, dass der Datenverwendung durch Human+ implizit zugestimmt wurde? Zunächst ist vorzuschicken, dass zufolge einer Studie des Deutschen Instituts für Vertrauen und Sicherheit im Internet unter knapp mehr als 1.000 deutschen Internet-NutzerInnen ab 14 Jahren ein großer Teil der Befragten die AGB von KommunikationsdienstleisterInnen nicht liest (60% der Befragten; 20% haben dagegen geantwortet, AGBs genau bzw. detailliert zu lesen; der Rest wies sich keiner Kategorie zu; vgl. DIVSI 2015: 32). Vor allem junge und gebildete UserInnen würden AGBs nicht oder nur flüchtig lesen. Demgegenüber geht die genauere Lektüre von AGBs häufiger mit einem „ungutem Gefühl“ bei der Zustimmung oder sogar der Ablehnung der Nutzung eines Dienstes einher. Gleichzeitig existieren bei vielen NutzerInnen zumindest relativ konkrete Vorstellungen davon, was (grundsätzlich) Inhalte von AGBs sind.

Für den Kontext der vorliegenden Arbeit bedeutet dies, dass speziell die Berücksichtigung des Prinzips des Respekts vor der Autonomie im Falle von Social-Media-UserInnen fraglich ist. Während der „Idealfall“ dieser Berücksichtigung die Übernahme des aus dem biomedizinischen Bereich bekannten Konzepts der „informierten Zustimmung“ auch auf den IKT-Bereich wäre, sind im Falle von Human+ diese Idealbedingungen nicht erfüllt: den Twitter-UserInnen fehlt nicht nur die Information über die konkreten Forschungsprojekte und kommerziellen Verwendungen, die auf ihre Posts durch die AGB-Zustimmung zugreifen, sondern es ist zu erwarten, dass ein bedeutender Teil der UserInnen auch die AGBs (die oft in komplexer Sprache und sehr ausschweifend formuliert sind) nicht gelesen hat, sondern ohne Lektüre der AGBs diesen zugestimmt hat. Auch fehlen weitere Charakteristika, die für eine informierte Zustimmung nötig sind, z.B. Kontrolle, ob die AGBs auch wirklich verstanden wurden. Auf Grund der fehlenden Kenntnis der weiteren, genauen Datenverwendung, ist auch ein „Ausstieg“ aus dem Forschungsprozess (oder der späteren Echtverwendung) bezüglich Human+ nicht möglich – lediglich die komplette Einstellung der Twitternutzung würde dem gleichkommen. Da Human+ außerdem auf die Auswertung von Massendaten in dieser Hinsicht angewiesen ist, ist das individuelle Einholen einer Nutzungserlaubnis auch bei entsprechendem „Goodwill“, mehr als das rechtliche Mindestmaß zu erfüllen, forschungspraktisch nicht möglich. Das vorliegende Argument besagt somit, dass die Nutzer-Einwilligung in die Auswertung der Daten durch Human+ auch bei Berufung auf deren Zustimmung durch AGBs zweifelhaft bleibt.

Kontra-Argument 6: Argument des Schadens durch Kontrollverlust bei Social-Media-UserInnen

Während also die Zustimmung von Social-Media-UserInnen hinsichtlich ihrer Datenverwendung gemäß des obigen Arguments zumindest angezweifelt werden kann, stellt sich die Frage, wie weiter zu verfahren ist, falls man die Zustimmung tatsächlich als unzureichend erachtet. Erleiden die Social-Media-UserInnen einen Schaden durch die Datensammlung und Verwendung, der sie nicht im ethischen Sinne ausreichend zugestimmt haben? Und wenn ja, auf welche Weise? Grundsätzlich lassen sich zum Thema Schutz der Privatheit [privacy protection] zwei Ansätze unterscheiden. Der erste ist die sogenannte **Schadens-basierte Theorie** [harm-based theory], welche besagt „[...] that so long as the data can be protected from attack by hackers or others wishing to ‚do‘ something harmful once gaining access, the privacy of the subjects can be maintained“ (Zimmer 2010: 321). Solange also die Daten nicht missbräuchlich verwendet werden, dass den Opfern ein tatsächlicher, konkreter Schaden widerfährt, könne auch nicht von Schaden gesprochen werden. Hierbei handelt es sich folglich um eine konsequentialistische Theorie. Anders gesagt: Während zwar prinzipiell nie ein hundertprozentiger Schutz von (ohne schädigende Absicht) gesammelten Daten a priori garantiert werden kann, gilt die Privatsphäre solange als unverletzt, bis a posteriori der Schutz durchbrochen wird, die gesammelten Daten von Personen durch die AngreiferInnen gestohlen werden und mit ihnen auch tatsächlich Schaden angerichtet wird. Kritisieren lässt sich an dieser Theorie nicht nur, dass nie völliger Schutz garantiert werden kann, sondern dass im Falle von AngreiferInnen, die widerrechtlich auf die Daten zugreifen indem sie Schutzmaßnahmen durchbrechen, dann aber die Daten nicht weiterverwenden um konkrete Schäden anzurichten, die Privatheit nach wie vor als geschützt gelten würde. Die zweite Theorie ist die **Würde-basierte Theorie** [dignity-based theory], derzufolge „[...] one does not need to be a victim of hacking, or have a tangible harm take place, in order for there to be concerns over the privacy of one’s personal information. Rather, merely having one’s personal information stripped from the intended sphere of the social networking profile, and amassed into a database for external review becomes an affront to the subjects’ human dignity and their ability to control the flow of their personal information“ (Zimmer 2010: 321). Dieser Theorie zufolge erleiden die Betroffenen bereits dadurch einen Schaden ihrer Würde und ihrer Datenautonomie, dass ihre persönlichen Informationen überhaupt aus der Sphäre des sozialen Netzwerkes abgegriffen und extern gespeichert werden. Zimmer zufolge wird über die angemessene Theorie in Fachkreisen nach wie vor gestritten. Für die Human+ Technik sind die Postings (im Falle des Demonstrators: v.a. die Tweets) der Social-Media-UserInnen nicht aufgrund der darin möglicherweise enthaltenen persönlichen Daten zu den UrheberInnen (etwa deren Präferenzen bezüglich verschiedener Sachgebiete) interessant, sondern aufgrund ihrer Relevanz und ihres Informationsgehalts für die Prognose von Migrationsbewegungen. Dennoch stellen persönliche Infor-

mationen gemäß Definition der DSGVO⁶⁶ einen Bestandteil der Daten dar, da sowohl der Twitter-Nutzer-Name, der Originaltext (der auch weitere persönliche Daten enthalten kann) oder auch die Geoposition in den Metadaten enthalten sein kann. Während für den Human+ Demonstrator bzw. die Entwicklungen während der Projektlaufzeit vor allem Twitter relevant war, trifft diese grundlegende Problematik auch auf andere Plattformen zu. Facebook etwa ist nach Auskunft des TechnikerInnenteams von Human+ restriktiver beim Datenzugang als wie Twitter, und ermöglicht es seinen NutzerInnen, verschiedene Sichtbarkeitsgrade von Posts und Profilen zu konfigurieren, sodass die Problematik bei anderen Plattformen bei Beachten der konfigurierten Sichtbarkeit entweder entschärft würde, oder aber bei der Nutzung von Möglichkeiten, sich über die von den Profil-EigentümerInnen eingestellten Sichtbarkeitsgrade hinwegzusetzen sogar verstärkt wird, da hier noch explizite Erwartungen der NutzerInnen an die Privatheit von als „privat“ konfigurierten Postings verletzt werden. Demgegenüber kann bei Twitter argumentiert werden, dass öffentliche Sichtbarkeit für Tweets charakteristisch, ja ein Wesensmerkmal ist. Proponenten der Würde-basierten Theorie des Schadens können jedenfalls dennoch argumentieren, dass den Social-Media-UserInnen durch das Crawlen ihrer Daten ein Schaden entsteht, da die Daten aus der ursprünglich von den UserInnen beabsichtigten Sphäre entfernt und neu zusammengestellt werden, und sie die Kontrolle über ihre Daten verlieren.

Kontra-Argument 7: Keine Entschädigung von Social-Media-UserInnen

Die Social-Media-Komponente von Human+ involviert das (anlassweise oder kontinuierliche) Crawlen von Posts auf Social Media (z.B. auf Twitter, worauf der Fokus während des TRL4-Projekts gelegt wurde). Diese Posts werden von Privatpersonen aus unterschiedlichen persönlichen Gründen verfasst. Die Bereitstellung einer Datenbasis für Human+ kann dabei als persönlicher Motivationszweck angesichts der öffentlichen Unbekanntheit des Projekts ausgeschlossen werden. Dies bedeutet in der Konsequenz, dass die Human+ Technik als eine mögliche Datenquelle Tweets heranzieht, mit (siehe oben) fraglicher Einwilligung der NutzerInnen, die zu völlig anderen Zwecken verfasst wurden, diese Tweets

⁶⁶ **Personenbezogene Daten** laut DSGVO sind „[...] alle Informationen, die sich auf eine identifizierte oder identifizierbare natürliche Person (im Folgenden ‚betroffene Person‘) beziehen; als identifizierbar wird eine natürliche Person angesehen, die direkt oder indirekt, insbesondere mittels Zuordnung zu einer Kennung wie einem Namen, zu einer Kennnummer, zu Standortdaten, zu einer Online-Kennung oder zu einem oder mehreren besonderen Merkmalen, die Ausdruck der physischen, physiologischen, genetischen, psychischen, wirtschaftlichen, kulturellen oder sozialen Identität dieser natürlichen Person sind, identifiziert werden kann“ (EU 2016: Artikel 4, Ziffer 1). Außerdem unterscheidet die DSGVO eine Kategorie von besonders **sensiblen Daten**, nämlich solche, „[...] aus denen die rassische und ethnische Herkunft, politische Meinungen, religiöse oder weltanschauliche Überzeugungen oder die Gewerkschaftszugehörigkeit hervorgehen, sowie die Verarbeitung von genetischen Daten, biometrischen Daten zur eindeutigen Identifizierung einer natürlichen Person, Gesundheitsdaten oder Daten zum Sexualleben oder der sexuellen Orientierung einer natürlichen Person“ (EU 2016: Artikel 9, Ziffer 1). Die Verarbeitung dieser Daten ist grundsätzlich untersagt, außer bestimmte Ausnahmebestimmungen (z.B. ausdrückliche Einwilligung) treffen zu.

unentgeltlich nutzt, und den UrheberInnen der Tweets auch keinerlei sonstige Entschädigung zukommen lässt. Vor dem Hintergrund des Prinzips der Gerechtigkeit scheint somit fehlende Reziprozität vorzuliegen.

Kontra-Argument 8: Argument der Sachzwänge

Unter „Sachzwang“ wird eine Vielzahl unterschiedlicher Phänomene bezeichnet. Im politischen Bereich etwa ist damit oft gemeint, dass der Entscheidungsspielraum eingeschränkt ist durch gesetzliche, wirtschaftliche oder gesellschaftliche Faktoren, die sich (angeblich) kurzfristig nicht ändern lassen (vgl. Innerhofer 2018: 15f). Als Sachzwänge werden auch in einem alltagssprachlichen Sinne Einschränkungen des Handlungsspielraums von Personen bezeichnet, um bei ihnen ein gewünschtes Verhalten zu fördern (ebd.: V). Eine stärker wissenschaftliche Definition lautet: „Ein Sachzwang ist eine pädagogisch motivierte, technische Intervention, die zu einer Reduktion des Handlungsspielraums führt“ (ebd.: 139). Pädagogisch meint in diesem Sinne: nicht naturhaft gegeben. Während **harte Sachzwänge** alternative Pläne zur Realisierung von Zielen vollständig aus dem Handlungsspielraum entfernen, sind **weiche Sachzwänge** solche im übertragenen Sinne, die symbolhaft deutlich machen, dass bestimmte Lösungswege nicht mehr gewählt werden sollten (vgl. ebd.: 141).

Für den vorliegenden Kontext besagt das Argument der Sachzwänge, dass die Bereitstellung einer Technik, die eigentlich laut Antrag zur Unterstützung von BOS-FachanwenderInnen gedacht ist und die menschliche fachliche Kompetenz ergänzen soll, als „weicher Sachzwang“ wirkt. Letztendlich, so der Gedanke, handelt es sich dann nicht um ein optionales Werkzeug, das bei Bedarf eingesetzt werden kann, sondern dessen Einsatz den verantwortlichen Personen qua seiner Verfügbarkeit „nahegelegt“ wird. Im Hintergrund dieser Problematik steht auch jeweils das Erfordernis, sich im Rahmen von BOS-bezogenen Tätigkeiten (spätestens in der Aufarbeitung von Einsätzen) für die jeweiligen Entscheidungen rechtfertigen zu müssen. Der Einsatz von Unterstützungssoftware wie Human+ kann im Zweifel wie die „sicherere“ Alternative gegenüber einem Verzicht auf eine solche Verwendung wahrgenommen werden. Verstärkt wird dies durch den Umstand, dass bei Human+ im Bereich der Migrationsprognose induktive, datenbasierende Algorithmen verwendet werden, denen im Rahmen von Sicherheitstechnologien oft eine „Aura der Objektivität“ (vgl. Baur-Ahrens und Krüger 2015: 32) gegenüber einer vermeintlichen menschlichen Subjektivität anhaftet, was es schwieriger macht, gegen die maschinell erzeugten Prognosen auf Basis eigener Expertise zu argumentieren. Sachzwänge scheinen in jedem Fall eine Beschränkung der Autonomie der Betroffenen darzustellen, und somit eine Verletzung des Prinzips des Respekts vor der Autonomie zu implizieren.

Kontra-Argument 9: Argument der steilen Lernkurve

Im Rahmen der Evaluation des Demonstrators im THW Ausbildungszentrum in Neuhausen auf den Fil dern am 27.9.2019 und 28.9.2019 hat sich gezeigt, dass zumindest das dort vorgestellte Demonstra torsystem trotz mehrstündiger Vorstellungsvorträge nicht intuitiv genug anwendbar war, um auch nur in einer simulierten Stabsituation „ad hoc“ eingesetzt zu werden (vgl. Human+ 2019a: 17). Die Einar beitung der AnwenderInnen in ein System wie Human+ erfordert offenbar mehr Zeit und eine anwen dungsnahere Einschulung als lediglich in Form von Vorträgen. Denn: Zahlreiche der Funktionen, die der versammelte Übungsstab für die Abarbeitung eines Stabsrahmenübungsrehbuchs hätte verwen den können, wurden nicht genutzt, oder Informationen in der Geoinformationssystem-Lagekarte des Demonstrators wurden falsch interpretiert. Dies ist besonders dann gravierend, wenn Outputs des Hu man+ Systems, welche lediglich als Hinweise für das mögliche (bestenfalls: sehr wahrscheinliche) Vor liegen von Migrationsbewegungen interpretiert werden sollten, stattdessen als „Fakten“ über Migra tionsbewegungen missinterpretiert werden, d.h. die Begrenzungen des Systems und seine Natur als unsicherheitsbehaftetes Prognosesystem nicht ausreichend verstanden werden. Eine Ursache für die ses Missverständnis kann etwa die im letzten Argument schon erwähnte Aura der Objektivität von Ergebnissen, die durch Algorithmen zustande kamen, sein, die vergessen lässt, dass Prognosen prinzi piell stets einen gewissen Grad an Unsicherheit haben. Dies wird dann zu einem ethisch relevanten Faktor, wenn Sachzwänge die Anwendung der Human+ Technik bei ungenügender Ausbildung der Be dienerInnen für den Umgang mit der Technik forcieren, Begrenzungen der Ergebnisse und wie verant wortungsvoll damit umgegangen werden kann⁶⁷, unterbeleuchtet bleiben, und so insgesamt durch vermeidbare Falschentscheidungen eher Ressourcen aus dem Krisenstab abgezogen als bereitgestellt werden.

Kontra-Argument 10: Einsatz des Systems auf der richtigen Anwendungsebene nötig

Gemäß Diskussionen mit den als Projektpartnern involvierten BOS muss der Einsatz des Human+ Sys tems auf der „richtigen Anwendungsebene“ erfolgen, da sonst unverhältnismäßiger „Ressourcenab zug“ zu erwarten ist. Die „richtige“ Anwendungsebene hängt allerdings stark von der letztendlich bei einem marktreifen (d.h. weit über TRL4 hinaus weiterentwickelten) System tatsächlich realisierten Funktionalitäten ab, und weiters von der Benutzerfreundlichkeit, dem notwendigen Schulungsauf wand, und den schon weiter oben diskutierten Festlegungen hinsichtlich der konkreten Anwendungs bedingungen (vgl. S. 117). So würde es etwa keinen Sinn machen, jeden Einsatzstab auf Gemeinde-

⁶⁷ Verantwortungsvoller Umgang könnte etwa darin bestehen, Prognosen des Human+ Systems als Auslöser für zusätzliche Verifizierungsmaßnahmen zu verwenden, anstatt umfassende Entscheidungen alleine auf den Prognosen basierend zu tref fen.

oder womöglich auch nur Bezirksebene mit einem System auszustatten, welches hauptsächlich nur auf einer gröber granularen Ebene (z.B. Bundesländer oder Staaten) Migrationsprognosen tätigen kann. Sinnvoller wäre da der Einsatz auf einer höheren Ebene unter Zuhilfenahme von FachanwenderInnen, die dann mit nachgelagerten Stäben und Stellen die Informationen austauschen. Da diese Aspekte allerdings nicht vorab festgelegt sind, besteht zumindest das Risiko eines Einsatzes auf einer falschen Anwendungsebene, z.B. durch missgeleitetes Engagement, das durch undifferenzierte Technikbegeisterung von BOS-MitarbeiterInnen „auf der falschen Ebene“ ausgelöst wird.

Pro-Argumente:

Pro-Argument 1: Verbesserte Chance auf rechtzeitige und angemessene Versorgung

Dieses Argument besagt, dass (angenommen die Technik funktioniert wie beabsichtigt und wird korrekt angewendet) die Migrationsprognosen und das gemeinsame Lagebild dazu beitragen, dass Geflüchtete rechtzeitig angemessene Versorgung erhalten. Es ist somit eine grobe Zusammenfassung und Neuformulierung derjenigen Aspekte, wie sie im Abschnitt zur Hauptfrage 1 in der Rubrik der Legitimität des mittelbaren Verwendungszwecks diskutiert wurden, und wird daher an dieser Stelle nur kurz erwähnt.

Pro-Argument 2: Bessere Informationen zu Flüchtlingsbewegungen und Ankunftszeiten, dadurch bessere Ressourceneffizienz bei BOS

Falls Human+ dazu führt, dass sich BOS rechtzeitiger und angemessener auf Migrationsbewegungen einstellen können, dann ist dies auch ein Beitrag zu einer besseren Ressourceneffizienz. Davon profitieren einerseits die involvierten BOS selbst, weil sie die zur Verfügung stehenden Mittel (an Personal, an aufzubauender Infrastruktur, an Versorgungsgütern, etc.) zielgerichteter verwenden können. Andererseits hat davon auch die allgemeine Zivilgesellschaft indirekt einen positiven Effekt, da eine geringere Belastung von BOS bedeutet, dass die BOS auch mehr Ressourcen für die Bewältigung von auch in migrationsbezogenen Krisenzeiten ja weiterhin anfallenden anderen Ereignissen (z.B. Naturereignissen) haben. Hierbei handelt es sich also um ein Nützlichkeitsargument vor dem Hintergrund des Prinzips der Fürsorge.

Pro-Argument 3: Gerechter Beitrag zu einer arbeitsteiligen Gesellschaft

Vor dem Hintergrund des Prinzips der Gerechtigkeit, welches u.a. als Forderung nach einer fairen Verteilung von Lasten und Pflichten gelesen werden kann, zeigt sich die Entwicklung, Verwendung und

Anschaffung einer Technik wie Human+ (wie übrigens auch jedes anderen technischen Hilfsmittels) als Beitrag zur Unterstützung von BOS: Da Human+ öffentliche Forschungsgelder durch das „Bundesministerium für Verkehr, Innovation und Technologie“ (BMVIT) über das KIRAS Forschungsprogramm bzw. Fördermittel von deutscher Seite durch das Bundesministerium für Bildung und Forschung (BMBF) im Rahmen des Programms „Forschung für die zivile Sicherheit“ erhielt (vgl. DHPol 2021), und sich solche öffentliche Forschungsgelder letztendlich durch Steuergelder speisen, kann in extenso die Forschung an Human+ aus einer Gerechtigkeitsperspektive auch als Beitrag der Gesamtgesellschaft an einer besseren Ausstattung für BOS gelesen werden.

Pro-Argument 4: Indirekter Gewinn durch Zusammenarbeit mit BOS

Bei der Bewältigung der „Flüchtlingslage von 2015/2016“ waren neben BOS im klassischen Sinn auch NGO sowie „grassroots“-ähnliche Bewegungen wie Train of Hope involviert. Als ein Erfordernis für die Human+ Verwendung wurde im Zuge der Anforderungserhebung immer wieder von Seiten der befragten BOS-MitarbeiterInnen betont, dass die Weitergabe von durch Human+ generierte Informationen heikel sein kann, und auf BOS beschränkt sein sollte. Dies sollte verhindern, dass etwa Migrationsbewegungen von anderen politischen Akteuren (z.B. rechtsradikalen Gruppierungen, bzw. Zitat aus einer Human+ Besprechung: „dem Naziverein von nebenan“) instrumentalisiert werden können. Dennoch scheint es nicht ausgeschlossen, dass selektive Informationen auch mit den bei der Bewältigung mit-involvierten nicht-staatlichen Akteuren geteilt werden können, wodurch letztendlich nicht nur BOS, sondern auch beteiligte NGOs durch die Human+ Technik profitieren können – zumal wenn diese NGOs selbst einen Beitrag für die humanitäre Sicherheit und Versorgung leisten. Diese Weitergabe kann dann allerdings nicht über einen „Anschluss“ an das Human+ System, sondern müsste wie erwähnt als selektive Weitergabe von Einzelinformationen aus dem Human+ System erfolgen.

Pro-Argument 5: Argument der Akzeptanz: Im Rahmen des Human+ Projekts wurde eine Erprobung des entwickelten Demonstrators durch die Anwendung innerhalb eines zweitägigen Übungsszenarios vorgenommen (vgl. Human+ 2019a). Dabei wurden im Vorfeld aufbereitete Daten und ein „Einsatzdrehbuch“ verwendet, sodass es sich hierbei nicht um einen Echteinsatz handelt, sondern um einen Testeinsatz, der dazu dienen sollte, weiteres Feedback von den potentiellen Human+ AnwenderInnen durch Simulierung eines Einsatzkontextes zu erhalten. Während und nach dieser Erprobung wurden u.a. durch sozialwissenschaftliche Methoden⁶⁸ auch die Einstellungen der involvierten Personen zur

⁶⁸ Zum Beispiel Fokusgruppen, Einzelinterviews, und quantitative Kurzbefragungen vor und nach dem Erproben des Systems.

Human+ Technik hinsichtlich verschiedener Akzeptanz-Dimensionen (Nutzungsintention, Benutzerfreundlichkeit, erwarteter Mehrwert, usw.) erhoben. Insgesamt zeigten die Untersuchungen, dass eine grundsätzliche Akzeptanz und Nutzungsintention gegenüber den gezeigten Komponenten vorhanden war. So zeigte etwa die quantitative Befragung unter den n=21 Testpersonen aus dem BOS Kreis, die in die Systemanwendung involviert waren, dass 16 Personen die Aussage „Ich würde für ähnliche Anwendungsfälle in Zukunft gerne eine solche Software verwenden“ auf einer 5-stelligen Skala von „1 – trifft voll zu“ bis „5 – trifft gar nicht zu“ mit der Note 1 bzw. 2 bewerteten (4 Personen verteilten die Note 3 und 4, bzw. 1 Person enthielt sich einer Angabe; vgl. Human+ 2019a: 37). Auch wenn im Detail zahlreiche Verbesserungsvorschläge zu Tage gefördert wurden, und die einzelnen Komponenten von Human+ (z.B. der erwartete Nutzen der Social-Media-Auswertungen, des gemeinsamen Lagebildes oder auch der Meldemasken) unterschiedlich bewertet wurden, kommt den Ergebnissen doch der starke Hinweischarakter zu, dass ein entsprechend optimiertes System auf große Akzeptanz bei den potentiellen NutzerInnen stoßen würde.

Pro-Argument 6: Argument des Bedarfs

Dieses Argument wurde bereits weiter oben bei der Behandlung des primären Zwecks vorgebracht, und wird hier nur der Vollständigkeit halber noch einmal genannt (vgl. S. 111 der vorliegenden Arbeit).

6.7.2 Evaluation und Entscheidung

In diesem Abschnitt werden zunächst die Kontra-Argumente bewertet, um zu entscheiden, ob diese so schwerwiegend sind, dass sie prinzipiell nicht mehr durch Pro-Argumente aufgewogen werden können. Falls die Bewertung zum Schluss kommt, dass eine Abwägung von Kontra- und Pro-Argumenten zulässig ist (und vorausgeschickt werden kann hier, dass dies der Fall sein wird), dann werden die Pro-Argumente einer Evaluation unterzogen, um am Ende die Abwägung vorzunehmen und die Entscheidung zur Hauptfrage 3 zu treffen.

Eine Reihe der Kontra-Argumente verweist auf grundsätzliche Verletzungen der Prinzipien mittlerer Reichweite, welche allerdings zu großen Teilen vor dem Hintergrund der Hauptfrage 3 (Gefährdung der Lebensbedingungen und der Lebensqualität aktueller und zukünftiger Generationen durch die Folgen und Nebenfolgen der Anwendung der Technik) lediglich geringe praktische Auswirkungen haben dürften, oder für die zumindest entsprechende niedrigschwellige Gegenmaßnahmen vorgeschlagen werden können.

„**Kontra-Argument 1: Argument der Verunsicherung: Verunsicherung bei Geflüchteten durch eine zusätzliche Analysetechnik**“ kann dadurch abgeschwächt werden, dass es wenig wahrscheinlich ist,

dass Human+ als Fachtechnik für einen sehr eingeschränkten NutzerInnenkreis große öffentliche Publizität erreicht und somit tatsächlich verunsichernd wirken könnte. Eine Gegenmaßnahme wäre, das bestehende Risiko dadurch auszuschalten, dass von Anfang an transparent mit dieser Technik umgegangen wird, und dabei aber möglichen Missverständnissen proaktiv zu begegnen. Da Human+ keine Überwachungstechnik für die Verfolgung einzelner Personen (oder auch nur bestimmter Gruppen von Geflüchteten) ist, und auch keine speziellen Funktionen hierfür unterstützt, können schon vorab Informationen dazu veröffentlicht werden, die darlegen, was genau der Zweck der Technik ist, mit einer ausreichend genauen Beschreibung der Funktionsweise, und einer Erläuterung des Unterschieds zu tatsächlichen Surveillance-Lösungen (z.B. Tracking von Smartphones oder gezieltes Tracking individueller Social-Media-Profile).

Dem **„Kontra-Argument 2: Privatheitsbedrohung durch verwendete Technik“** kann zumindest teilweise technisch begegnet werden, indem bei den Social-Media-Auswertungen möglichst auf die Anzeige und die (langfristige) Speicherung der zugrunde liegenden Rohdaten (d.h. der Social-Media-Posts, wie z.B. Tweets) verzichtet wird, und stattdessen nur daraus erstellte Informationsprodukte angezeigt werden. Dem steht aber entgegen, dass zumindest der/die SichterInnen (vgl. S. 129) des Human+ Systems aus fachlichen Gründen Zugriff auf die Ursprungsdaten haben sollten, und somit die Speicherung der Rohdaten in einer Datenbank notwendig wird, wenngleich auf diese dann nur ein nochmal eingeschränkter AdressatInnenkreis Zugriff hätte. Bei den In-Situ-Auswertungen wurde für die Analyse von Bewegungsströmen durch Kameras während des Human+ Projekts konzeptionell die Möglichkeit angestrebt, die Auswertung „live“ ohne Zwischenspeicherung der Kamera-Aufnahmen zu bewerkstelligen, sodass letztendlich nur verschiedene Auswertungsergebnisse gespeichert würden, z.B. die Anzahl der Personen, die in einem bestimmten Zeitraum einen abgesteckten Bereich in die eine oder andere Richtung verlassen haben. Prinzipiell wäre eine „on-board-Auswertung“ auch bei UAV-Aufnahmen möglich – dies war jedoch nicht Teil des Human+ Projekts. Festzuhalten bleibt, dass auch bei Ergreifen von Gegenmaßnahmen potentiell die Privatheit von Betroffenen durch die Technik bedroht bleibt. Auch können durch COC bzw. durch die Dienstvorschriften der AnwenderInnen genaue Grenzen für den Umgang mit den erhobenen Daten gesetzt werden, um die Privatheitsbedrohung zu reduzieren und stärker in den Bereich von unsachgemäßen Anwendungen zu verlagern.

Das **„Kontra-Argument 3: Argument der Verletzung der Verhaltensökonomie von In-Situ-Betroffenen“** kann nur entkräftet werden, wenn statt einer verdeckten eine öffentliche Datenaufnahme gewählt wird, womit das **„Kontra-Argument 4: Argument der Normalisierung des Verhaltens“** schlagend wird. Hierzu ist zu sagen, dass in Kontra-Argument 4 sehr grundsätzlich von einer Normalisierung gesprochen wurde, als möglicher nicht-intendierter Begleiteffekt der Human+ In-Situ-Datenerfassung,

welche sich bei jeder Form von offen erfolgender Kamera- oder Videoüberwachung stellt. Eine Quantifizierung der tatsächlich erfolgenden Normalisierung scheint angesichts dessen, dass die Datenaufnahme keine gezielte Reaktion auf bestimmte Arten von kriminellem oder auch nur abweichendem Verhalten ist, schwierig, dürfte aber in etwa dem Ausmaß entsprechen, wie sie durch jede Überwachungskamera in U-Bahnen oder Geschäften erfolgt. Damit ist dann auch schwer zu sagen, worin denn „eigentlich“ der Schaden besteht, falls das Verhalten in Richtung erwarteter Normvorstellung adaptiert würde (d.h.: in welche Richtung das Verhalten denn tatsächlich angepasst wird), sodass es sich hierbei zwar um einen grundsätzlichen Punkt handelt, aber die konkreten Auswirkungen sich in Grenzen halten dürften (ausgenommen die mögliche Verunsicherung oder das Auslösen von Gefühlen des Verfolgt-Werdens bei Personen die aus repressiven Regimes kommen, wie weiter oben bereits beschrieben). Eine Möglichkeit, zumindest zusätzliche Normalisierung zu vermeiden, bestünde in der ausschließlichen Nutzung von bereits bestehenden Kamera-Systemen (in welche Richtung zahlreiche Diskussionen während des Human+ Projekts auch abzielten). Während Normalisierung bei öffentlicher Datenerhebung somit nicht ausgeschlossen werden kann, übersteigt sie nicht das Ausmaß der im Falle von U-Bahnen und öffentlichen Plätzen bereits stattfindenden.

„Kontra-Argument 5: Argument der zweifelhaften Einwilligung von Social-Media-UserInnen“ sowie das **„Kontra-Argument 6: Argument des Schadens durch Kontrollverlust bei Social-Media-UserInnen“** bleibt nach Ansicht des Verfassers der vorliegenden Arbeit grundsätzlich bestehen, sodass die Beeinträchtigung gegen das Prinzip des Respekts vor der Autonomie abgewogen werden muss gegenüber positiven Auswirkungen des Projekts in anderen Bereichen. Allenfalls kann darauf geachtet werden, für die Social-Media-Auswertungen hauptsächlich solche Plattformen zu wählen, bei denen argumentiert werden kann, dass die Erwartungshaltung der Social-Media-UserInnen hinsichtlich der „Wahrscheinlichkeit“ einer öffentlichen Datenverwendung und Einsehbarkeit ihrer Postings besonders hoch ist (wie etwa bei Twitter). Dies führt aber nicht zu einer völligen Entkräftigung dieser Argumente, da die Postings dennoch in einen anderen Kontext überführt werden, was die Erwartungshaltung zumindest mancher UserInnen verletzen kann.

„Kontra-Argument 7: Keine Entschädigung von Social-Media-UserInnen“ kann insofern relativiert werden, dass zwar den Betroffenen in ihrer Rolle als Social-Media-UserInnen keine Entschädigung für die Nutzung ihrer Posts zukommt, aber sie können in Form der Teilhaberschaft an anderen Stakeholdergruppen (z.B. Zivilgesellschaft, oder wenn sie selbst Geflüchtete sind) sehr wohl auch positive Folgen durch das Projekt erfahren. Zudem kann argumentiert werden, dass die Nutzung von Twitter selbst für die NutzerInnen nur deswegen unentgeltlich möglich ist, weil sie den Nutzungsbedingungen zustimmen, welche ausdrücklich eine Weiterverwendung der Daten durch Dritte erlaubt, wodurch sich

Twitter (teil)finanziert. Davon abgesehen lässt sich dieser Aspekt nicht durch organisatorische oder technische Maßnahmen sinnvoll behandeln, und muss daher Teil des Abwägungsprozesses werden.

Die Argumente „**Kontra-Argument 8: Argument der Sachzwänge**“, „**Kontra-Argument 9: Argument der steilen Lernkurve**“ sowie „**Kontra-Argument 10: Einsatz des Systems auf der richtigen Anwendungsebene nötig**“ beziehen sich entweder auf die AnwenderInnen des Human+ Systems oder im Fall von Argument 8 auch auf Human Sensors (OUH und BOS). Nach Ansicht des Verfassers der vorliegenden Arbeit können diese Argumente durch entsprechende organisatorische Rahmenbedingungen bei der späteren Anwendung gemildert oder sogar ausgeschaltet werden: „Weiche“ (bzw. inoffizielle) Sachzwänge können durch klare Anwendungsvorgaben so geregelt werden, dass die Optionalität des Human+ Einsatzes an Kriterien geknüpft und damit abgesichert ist, auf die sich AnwenderInnen dann beziehen können, um ihre Entscheidung zum Verzicht oder Einsatz von Human+ besser rechtfertigen zu können. Damit dies nicht trotzdem als von außen übergestülpter „harter Sachzwang“ wahrgenommen wird, etwa durch zu harte Anwendungskriterien, wäre es sinnvoll, diese Kriterien und Richtlinien in enger Kooperation mit den potentiellen AnwenderInnen zu erarbeiten. Diese Anwendungsvorgaben können in ähnlicher Weise auch für Human Sensors erarbeitet werden. Klare, nach rationalen Kriterien erarbeitete Anwendungsrichtlinien können zudem dem Einsatz des Systems auf einer „falschen Ebene“ vorbeugen, um etwa zu verhindern, dass technikbegeisterte Einzelpersonen Human+ qua seiner Eigenschaft als neuer Technik einsetzen wollen, ohne auf die tatsächliche Angemessenheit zu achten (Kontra-Argument 10). Bestandteil dieser Anwendungsregularien müssen dann (neben der schon erwähnten Klärung der AnwenderInnen-Gruppe und der Anwendungsumstände im Detail) auch die notwendigen Schulungen sein, da sich gezeigt hat, dass ein System wie Human+ nicht für eine intuitive Verwendung gestaltet werden kann. Diese Schulungen müssen dann auch ausreichend auf die Begrenzungen des Systems hinweisen, um einer Scheinobjektivität durch Algorithmen vorzubeugen.

Zwischenzeitlich kann festgehalten werden, dass es für eine Reihe der behandelten Kontra-Argumente technische oder organisatorische Gegenmaßnahmen gibt, die die dahinterstehenden Problematiken entweder abschwächen oder sogar ausschalten können. Einige Schwierigkeiten bleiben aber bestehen, weil sie sozusagen in der Human+ Technik grundgelegt sind (z.B. „**Kontra-Argument 2: Privatheitsbedrohung durch verwendete Technik**“, „**Kontra-Argument 4: Argument der Normalisierung des Verhaltens**“, „**Kontra-Argument 5: Argument der zweifelhaften Einwilligung von Social-Media-UserInnen**“, „**Kontra-Argument 6: Argument des Schadens durch Kontrollverlust bei Social-Media-UserInnen**“, sowie „**Kontra-Argument 7: Keine Entschädigung von Social-Media-UserInnen**“). Rückbezogen auf Hauptfrage 3 lässt sich sagen, dass diese Aspekte zwar nicht die Lebensgrundlage aktueller oder zukünftig lebender Menschen bedrohen, Bedrohungen der Privatsphäre können aber durchaus als Be-

einträchtigung der Lebensqualität gesehen werden – v.a. wenn neben dem Schaden durch Kontrollverlust auch noch ein tatsächlicher Schaden eintritt, etwa das System unsachgemäß verwendet wird. Weiters lässt sich hinterfragen, ob bei einer nichterfolgten „Entschädigung“ für das Verwenden von Social-Media-Posts tatsächlich eine Beeinträchtigung der Lebensqualität vorliegt, da die Posts völlig unabhängig von Human+ von den Social-Media-UserInnen erstellt wurden, ohne eine etwaige Entschädigung im Vorfeld überhaupt erwartet zu haben. Dennoch kann aus Gerechtigkeitsperspektive von mangelnder Reziprozität gesprochen werden.

Zusammenfassend scheinen die nicht auszuräumenden ethischen Konfliktpunkte solche darzustellen, die nicht solch grundlegender (etwa weil gegen Grundrechte gravierend verstoßender) Natur sind, dass sie nicht zumindest prinzipiell gegenüber den zu erwartenden Vorteilen aufgewogen werden könnten.⁶⁹

Das macht es notwendig, nun die Pro-Argumente einer Bewertung zu unterziehen, um sie am Ende gegen die Kontra-Argumente abwägen zu können.

„Pro-Argument 1: Verbesserte Chance auf rechtzeitige und angemessene Versorgung“ sowie **„Pro-Argument 2: Bessere Informationen zu Flüchtlingsbewegungen und Ankunftszeiten, dadurch bessere Ressourceneffizienz bei BOS“** beziehen sich auf den Haupt-Daseinszweck von Human+. Wenn die Prognosen und das gemeinsame Lagebild nicht dazu führen, die Chance auf rechtzeitige und angemessene Versorgung zu verbessern, fehlt die erste Hauptvoraussetzung laut DESSI-Framework, die Investitionen in Sicherheitstechniken rechtfertigen, nämlich eine tatsächliche Sicherheitsverbesserung. Während der oben schon angesprochene Demonstrator-Test in Neuhausen grundsätzlich gezeigt hat, dass zahlreiche der entwickelten Algorithmen für die Prognose Potential haben, war man aufgrund des TRL4 von einer tatsächlichen Anwendbarkeit naturgemäß noch entfernt. Zudem ist zu erwähnen, dass ein hypothetisches Human+ Echtsystem ja nicht nur auf technischer Ebene funktionieren muss, sondern auch noch fachgerecht innerhalb der richtigen Umstände zu nutzen ist, also als sozio-technisches System korrekt angewandt werden muss, was entsprechende Festlegungen, die schon mehrfach angesprochen wurden, erfordert. Pro-Argument 1 und 2 lassen sich somit nur probabilistisch bewerten und akzeptieren, ohne allerdings genau quantifiziert werden zu können. Von den organisatorischen Rahmenbedingungen der Zusammenarbeit von BOS und NGOs hängt es dann auch ab, ob **„Pro-Argument 4: Indirekter Gewinn durch Zusammenarbeit mit BOS“** in der Realität zutrifft.

Dass der Versuch einer Weiterentwicklung in Richtung höheres TRL erfolgen soll, wird durch **„Pro-Argument 6: Argument des Bedarfs“** zumindest gestützt, denn wenigstens den Konzepten hinter dem

⁶⁹ Vorausgesetzt, dass das zweite essentielle Kriterium laut DESSI-Framework, nämlich die Achtung der Grundrechte, in diesem Fall auf Privatheit, durch eine eingeschränkte Speicherung von Social-Media-Rohdaten und eine Beschränkung auf die „SichterInnen“ geschützt wird.

Human+ Projekt wird von befragten BOS-MitarbeiterInnen ein Mehrwert zugeschrieben. Und die Tatsache, dass der Versuch unternommen wird, BOS durch neuartige Projektentwicklungen zusätzliche Hilfsmittel an die Hand zu geben (**Pro-Argument 3: Gerechter Beitrag zu einer arbeitsteiligen Gesellschaft**) hängt streng genommen nicht vom Erfolg des Unterfangens ab, sondern bei einer pflichtethischen Leseart des Prinzips der Gerechtigkeit bereits von der dahinterliegenden Einstellung.

Dem „**Pro-Argument 5: Argument der Akzeptanz**“ dagegen muss aus ethischer Sicht mit Vorbehalten begegnet werden. Weil es sich um einen im Rahmen der Technikfolgenabschätzung sehr zentralen Unterschied handelt, soll an dieser Stelle die Differenz zwischen den beiden Teilen des Begriffspaares **Akzeptanz-Akzeptabilität** im Detail erläutert werden. **Akzeptanz** kann als „[...] die Übernahme von bestimmten Propositionen, Sachverhalten, politischen Agenden etc. in eigene Handlungen verstanden werden [...]“, und bezieht sich somit auf „[...] kognitive, emotionale und behaviorale Elemente des positiven Umgangs mit Elementen der Welt, dergestalt, dass diese Elemente handlungsrelevant sind und nicht zur Abstoßung, zur Missachtung, zur Auflehnung oder dergleichen führen“ (Wolkenstein 2014: 226). Das heißt, dass man das, was man „akzeptiert“, nicht unbedingt mögen oder gutheißen muss, lediglich dass sie nicht in Widerstände übersetzt werden. Während Akzeptanz recht einheitlich gefasst wird, finden sich in der Literatur unterschiedliche Auffassungen von Akzeptabilität. Als **Akzeptabilität** etwa kann entweder verstanden werden, wenn eine Sache „[...] fähig und würdig ist, Akzeptanz zu gewinnen [...]“ (Wolkenstein 2014: 226f), etwa Akzeptanz verdient und nicht moralisch oder rechtlich verboten ist, oder aber als Synonym für Richtigkeit bzw. konstruktivistisch als Eigenschaft des „von-allem-akzeptiert-Werdens“. In seinen weiteren Ausführungen führt Wolkenstein zwar die Überlegung ein, dass Akzeptabilität im politischen Sinne auch Akzeptanz erfordert und damit die Begriffe verschränkt sind. Im Allgemeinen bezeichnet aber Akzeptabilität die normative Ebene als Übereinstimmung mit moralischen, rechtlichen oder kulturellen Wertmaßstäben, und Akzeptanz eine Art individuelle Einstellung, die über sozialwissenschaftliche Befragungen erfasst und aggregiert werden kann. Das bedeutet auch, dass man durch Werbung und „akzeptanzschaffende Maßnahmen“ oder aber durch entsprechendes Aufheizen des politischen Klimas für die passende „Stimmung“ bei einer betroffenen Gruppe sorgen kann, dass Maßnahmen oder Techniken faktisch akzeptiert werden, die bei der Betrachtung anhand von moralischen oder rechtlichen Kriterien nicht legitim erscheinen (und umgekehrt). Eine etwas andere Definition liegt wiederum der TA-Richtung der „Rationalen Technikbeurteilung“ zugrunde: **Akzeptabilität** wird hier zwar ebenfalls als normativer Begriff verstanden, allerdings lege dieser die „Akzeptanz von risikobehafteten Optionen mittels rationaler Kriterien des Handelns unter Risikobedingungen“ fest (Gethmann und Sander 1999: 146). Bezogen auf das Risikoverhalten (und die Nutzung oder Ablehnung von bestimmten Instanzen von Technik kann als Risikoverhalten interpretiert werden, was besonders am Beispiel des Sicherheitsgurts verständlich wird) ist etwa das

Verhalten akzeptabel, das „[...] ein kognitiv und operativ perfekter Entscheider angesichts mehrerer Handlungsalternativen zeigen würde“ (Gethmann und Sander 1999: 146). Diese Definition erlaube es „**hypothetische Imperative**“ zu formulieren, sodass Personen, die bestimmte Risiken akzeptieren, auch vergleichbare Risiken akzeptieren sollten, z.B. wer das Risiko beim Bergsteigen akzeptiert, sollte auch das Fliegen mit Linienmaschinen riskieren. Akzeptabilität zielt hier somit auf die Binnenrationalität von Individuen ab, was Gethmann und Sander zum **Prinzip der pragmatischen Konsistenz** verdichten: „Hat jemand durch die Wahl einer Lebensform den Grad eines Risikos akzeptiert, so darf dieser auch für eine zur Debatte stehende Handlung unterstellt werden“ (Gethmann und Sander 1999: 146).

Nach Grunwald können weder Akzeptanz noch Akzeptabilität für sich genommen ausreichend für die konstruktive Behandlung von **Technikkonflikten** beitragen, weil beide für sich genommen auf Schwierigkeiten stoßen (vgl. Grunwald 2010: 242). So sagt faktische Akzeptanz nichts über die ethische Rechtfertigbarkeit aus, ist stets auch immer nur eine gegenwärtige empirische Momentsituation und kann sich für die Zukunft ändern (Extrapolationsproblem, Stabilitätsproblem). Zudem würden gemäß Arrow-Theorem der Entscheidungstheorie individuelle Präferenzen nicht widerspruchsfrei zu einer wohl definierten Gesamtnutzenfunktion aggregiert werden können (Aggregationsproblem), sodass Technikkonflikte gar nicht gelöst werden können, indem alle Präferenzen gleichermaßen befriedigt werden. Akzeptabilität (im Sinne von Gethmann und Sander) wiederum erfordere für das Vergleichen von Risiken mit anderen Risiken eine Art einheitlicher Skala (welche es nicht gäbe), wodurch die Frage, welche Risiken als gleichwertig und vergleichbar betrachtet werden, selbst zum Gegenstand von Kontroversen wird. Risiken sind immer Risiken für „jemanden“, sodass es keine einheitlichen, sondern unterschiedliche Perspektiven geben dürfte. Und zuletzt erfordere jenes Konzept von Akzeptabilität zu starke Rationalitätsanforderungen – in demokratischen Systemen würde niemand auf (rationale) Konsistenz geprüft werden bzw. dies Voraussetzung für gesellschaftliche Teilhabe sein.⁷⁰

Grunwald (2010: 244f) ergänzt, dass statt allgemeiner Akzeptanz gegenüber einer Technik eher Situationen unterschiedlicher Wahlmöglichkeiten gegenüber der Exposition (d.h. dem Ausgesetzt-sein) einer Technik unterschieden werden müssten: Zumutungen mit individuell kontrollierbaren Ausweichmöglichkeiten, mit einfachen Ausweichmöglichkeiten, mit beschwerlichen Ausweichmöglichkeiten, und ohne Ausweichmöglichkeiten. Die Grad einer etwaigen Technikkonfliktsituation hänge stark von der Möglichkeit der Einflussnahme auf die **Exposition** ab.⁷¹

⁷⁰ Grunwald schließt daraus, dass Technikkonflikte nur prozedural gelöst werden können, d.h. in geregelten Verfahren (vgl. Grunwald 2010: 243).

⁷¹ Eine wichtige Rolle spielt hier das Thema Information: so können gentechnisch veränderte Lebensmittel von Konsumenten durch entsprechendes Kaufverhalten vermieden werden, vorausgesetzt es gibt entsprechende Kennzeichnungen. Gerade wenn es um eine Technik ohne Ausweichmöglichkeit geht, müsse nach Grunwald nicht vom individuellen, sondern vom gesellschaftlichen Standpunkt aus eruiert werden (etwa in Form von Standards, Risikogrenzen, und dergleichen). Bei Fragen von

Angesichts dieser unterschiedlichen Bedeutungsdimensionen ist zunächst festzuhalten, dass Human+ (anders als z.B. Sicherheitsgurte) nicht die Art von Risiko-Bezug („Handeln unter Risikobedingungen“) aufweist, wie sie der **Akzeptabilitäts**-Definition von Gethmann und Sander zugrunde liegt. Da es sich bei Human+ um eine Technik für Fachanwender handelt, ist auch der Aspekt der demokratischen Legitimation bzw. damit verbundener Verfahren zumindest unter Vorbehalt zu betrachten, anders als etwa bei den klassischen **Technikkonflikte** auslösenden Themenbereichen wie etwa Atomkraft (auch in Form von Kraftwerken), Gentechnik oder Nanotechnik, die eine andere Art von gesellschaftsweiter Betroffenheit bzw. Exposition implizieren als wie Human+. Lediglich für die Betroffenheiten durch die Social-Media-Auswertungen und die In-Situ-Aufnahmen kann argumentiert werden, dass hiervon auch die allgemeine Bevölkerung nachteilig betroffen sein könnte – mit geringen Möglichkeiten, dieser Exposition zu entgehen, wenn man entweder Twitter nutzen möchte oder sich in einem bestimmten Gebiet aufhalten möchte. Nach Ansicht des Verfassers der vorliegenden Arbeit kann im Falle von Human+ dennoch nicht von **Technikkonflikten** gesprochen werden, weil es gar keine größeren oder auch nur kleineren demographischen Gruppen gibt, die Parteien in einem solchen Konflikt sein könnten. Akzeptabilität wird folglich für Human+ festgelegt als abhängig von der umfassenden Betrachtung ethischer Gesichtspunkte vor dem Hintergrund unterschiedlicher Prinzipien mittlerer Reichweite, d.h. als ethische Zulässigkeit. Wenn dadurch allerdings Akzeptanz und (ethische) Akzeptabilität konzeptionell unterschiedliche Bedeutungen gewinnen und daher in der Analyse voneinander unterschieden werden, relativiert dies die Bedeutung der weiter oben angeführten „Zustimmung“ potentieller AnwenderInnen zu Human+, auf die im „**Pro-Argument 5: Argument der Akzeptanz**“ verwiesen wird. Zwar mag es ethisch nicht irrelevant sein, wenn sich potentielle AnwenderInnen-Gruppen für oder gegen die Nutzung einer bestimmten Technik aussprechen (etwa vor dem Hintergrund von Sachzwang-Abwägungen), aber die Tatsache alleine, dass die Technik akzeptiert wird, kann in der Bewertung der Technik lediglich eine untergeordnete Rolle spielen.

Wie lassen sich nun die Kontra-Argumente gegen die Pro-Argumente abwägen? Da die Bewertung mehrerer Argumente konditionaler Natur ist, sollten für die Abwägungen zwei Szenarien unterschieden werden: **Szenario A**, in welchem die dargestellten „Gegenmaßnahmen“ zur Abfederung zahlreicher vorgebrachter Gegenargumente umgesetzt wurden, sowie **Szenario B**, in welchem sie nicht umgesetzt wurden, sodass diese beiden Szenarien die Pole eines Spektrums bilden. In beiden Fällen wird davon ausgegangen, dass der nur probabilistisch zu bewertende Nutzen von Human+ tatsächlich ein-

Technikkonflikten müsse daher ein solcherart kollektives Problem auf der Ebene demokratischer Meinungsbildung und Entscheidungsfindung gelöst werden, auf eine Weise, die Luhmann als Herstellung von Legitimität durch Verfahren charakterisiert hat, und mit welchen die Gesellschaft die Akzeptanz der Zumutbarkeit vor dem Hintergrund öffentlichen Interesses festlegt.

tritt, da ansonsten gemäß DESSI-Hauptkriterium 1 (Sicherheitsgewinn) von der Technik generell abzuraten wäre. Zu fordern, dass der erwartete Sicherheitsgewinn garantiert, und nicht nur mit einer bestimmten Wahrscheinlichkeit, eintreten müsse, scheint dagegen als zu starkes Kriterium, schließlich würde sich dieses Argument dann auch gegen sämtliche andere Arten von sicherheitsbezogener Grundlagenforschung wenden, deren Verwertungs- und Erfolgspotential im Vorfeld nicht genau quantifizierbar ist.

Szenario A) In diesem Szenario bedroht die Sammlung und Auswertung von Social-Media-Daten, obwohl für Human+ nicht die persönlichen Daten der VerfasserInnen der Social-Media-Posts relevant sind, zumindest grundsätzlich deren Privatsphäre, da die Daten aus dem ursprünglichen Verwendungskontext entnommen wurden und möglicherweise private Daten den „SichterInnen“ vorgelegt werden. Der Verwendung ihrer Social-Media-Posts zum ausdrücklichen Zwecke von Human+ haben die Social-Media-UserInnen nicht zugestimmt, lediglich in Form einer Art Generalvollmacht, die in den Nutzungsbedingungen der Plattform enthalten sind. Sie erhalten dafür keine Entschädigung, und können auch nichts dagegen unternehmen, dass ihnen die Kontrolle dieser „gecrawlten“ Daten verloren geht. Außerdem kommen auf öffentliche Weise entweder Videokameras oder UAV-Luftbildaufnahmen zum Einsatz, die zumindest prinzipiell Verhaltensnormalisierungen der Personen in ihrem Einflussbereich auslösen können, auch wenn deren Umfang zumindest im Falle von Videokameras nicht das Ausmaß typischer Überwachungskameras wie etwa in U-Bahn und Supermarkt überschreitet⁷². Dem steht auf der anderen Seite die Chance gegenüber, dass durch den (korrekten) Einsatz der Technik BOS ankommende Geflüchtete rechtzeitiger und angemessener akutversorgen können und dadurch allgemein die Ressourcen von BOS effizienter genutzt werden. Durch (selektiven) Informationsaustausch können auch in der Versorgung tätige NGOs unterstützt werden. Die Weiterentwicklung bzw. der Einsatz der Human+ Technik kommt dem von den BOS artikulierten Bedarf an einer solchen Technik entgegen, und stellt einen kleinen Beitrag in der Verteilung von Lasten (Entwicklungskosten), Pflichten (Tätigkeiten der BOS sind gesetzlich vorgeschrieben) und Vorteilen (Arbeitserleichterung bei BOS bzw. Genuss von BOS-Leistungen bei der Zivilgesellschaft) dar. Zudem besteht die Aussicht, dass die Human+ AnwenderInnen die Technik auch akzeptieren.

Exakte Quantifizierungen in diesem Abwägungsprozess sind nach Ansicht des Verfassers der vorliegenden Arbeit nicht möglich (etwa im Sinne eines „utilitaristischen“ Kalküls, wie es sich etwa bei Bentham angelegt findet). Eine Verbesserung der Sicherheit und Versorgung von Geflüchteten durch bessere

⁷² Somit könnte ein in der Diktion von Gethmann und Sander's „hypothetischer Imperativ“ skizziert werden: wer die Videoüberwachung in Supermärkten und U-Bahnen hinsichtlich der Verhaltensnormalisierung akzeptiert, müsse sie auch im Falle von Human+ Anwendungen akzeptieren.

Migrationsprognosen berührt deren vitalste Interessen, während eine (un)effiziente Ressourcennutzung von BOS aufgrund deren oft Leib und Leben betreffende Tätigkeitsfelder ebenfalls Auswirkungen auf vitalste Interessen anderer Personen haben können. Ein Grund hierfür wäre etwa, weil Kräfte im Falle einer Flüchtlingslage auf eine Weise gebunden sind, dass für die Bewältigung anderer Krisen und Katastrophen weniger Ressourcen zur Verfügung stehen, was bis zu einer Bedrohung des Lebens der von diesen anderen Ereignissen betroffenen Personen führen kann. Demgegenüber scheinen die Nachteile des Systems (Verhaltensnormalisierung, fehlende Entschädigung der UrheberInnen bei Social-Media-Analysen, im Personenkreis „eingeschränkte“ Privacy-Verletzungen bei Sichtung von Originalposts durch SichterInnen ohne damit einhergehenden Missbrauch im engeren Sinne) zwar unangenehm und aus ethischer Sicht prima facie abzulehnen, aber letztendlich auf einer weniger vitalen Ebene zu liegen und daher im Abwägungsprozess kompensiert werden zu können. Zumindest in Szenario A kann daher argumentiert werden, dass der Nutzen der Human+ Technik (zumindest probabilistisch) die Nachteile überwiegt, welche zudem nicht essentielle Lebensbedingungen bedrohen, sondern „lediglich“ Teilaspekte der Lebensqualität. Dies ist zwar bedauerlich, aber im Falle der Güterabwägung nach Ansicht des Verfassers der vorliegenden Arbeit zulässig.

Ähnlich lässt sich auch unter Bezug auf das Metaprinzip des Negativen Utilitarismus argumentieren: Die Human+ Technik leistet einerseits einen Beitrag zur Reduzierung von Leid (vgl. für die Argumentation S.123). Andererseits führen die erwähnten prima facie Nachteile der Technik unbeabsichtigt ebenfalls zu Leid, welches aber auf einer weniger vitalen Ebene liegt als das Leid, das durch die Technik reduziert wird. Wenn angenommen wird, dass durch das Ermöglichen einer besseren Vorbereitung von BOS auf ankommende Geflüchtete die humanitäre Sicherheit (besser) gewährleistet werden kann und so im Falle eines Human+ Einsatzes weniger Leid (insgesamt) vorhanden ist als wie im Falle eines Nicht-Einsatzes, dann müsste dem Metaprinzip des Negativen Utilitarismus folgend der Einsatz von Human+ gegenüber einem Nichteinsatz bevorzugt werden. Hier wird aber auch sichtbar, dass viele der eben genannten Annahmen auf Wahrscheinlichkeitsschlüssen basieren, z.B. dass die Human+ Technik auf eine Weise funktionieren und korrekt eingesetzt wird, dass erstens zutreffende Prognosen für Migrationsbewegungen gewonnen werden, und dass diese zweitens auch tatsächlich für die Verbesserung der humanitären Sicherheit verwendet werden. Betrachtet man die Einsatzlage von 2015/2016 als Referenz, dann lässt sich argumentieren, dass zumindest in einigen Phasen, z.B. in der Frühphase bzw. der „Chaosphase“, in welcher die BOS häufig mit fehlenden Informationen zu kämpfen hatten, eine einsatzreife Human+ Technik einen wertvollen Beitrag geleistet hätte (vgl. Abschnitt 6.4.1.3 ab S. 103). Aus ethischer Sicht wäre daher im Falle einer Weiterentwicklung und eines Einsatzes der Human+ Technik in der Zukunft eine Neubewertung der Technik wünschenswert, wenn sich diese zum jetzigen

Zeitpunkt lediglich probabilistisch zu bewertenden Vorannahmen dann auch empirisch bewerten lassen.

Szenario B) In diesem Szenario kommt gegenüber den Nachteilen in Szenario A hinzu, dass

- es Verunsicherungen (bei Geflüchteten) durch Human+ gibt (was sich etwa in Form von noch weiter reduzierter Smartphone-Nutzung während der Flucht auswirken könnte – auf den wichtigen Stellenwert des Smartphones als Werkzeug, das den Geflüchteten auf der Reise zusätzliche Autonomie verleiht, wurde oben bereits hingewiesen),
- eine größere Privatheitsbedrohung für Social-Media-UserInnen durch einen erweiterten Kreis an Personen, die systematisch zusammengestellte Rohdaten⁷³ einsehen können (z.B. alle Human+ BedienerInnen), was die Einfachheit von missbräuchlichen Verwendungen (Hauptfrage 2) erhöht,
- Videoaufnahmen bzw. UAV-Aufnahmen auf versteckte Weise erfolgen,
- und die Anwendung der Human+ Technik durch unklare Rahmenumstände und schlechte Schulungslagen für die BedienerInnen mit Sachzwängen, Bedienungsschwierigkeiten und Ressourcenverluste durch Einsatz auf der „falschen“ Ebene einhergeht.

Auch in Szenario B wird davon ausgegangen, dass die Human+ Technik ordnungsgemäß angewandt wird, also nicht auch noch zusätzlicher Missbrauch des Systems, wie er in der Hauptfrage 2 behandelt wurde, schlagend wird, auch wenn diese Missbrauchsmöglichkeiten durch obige Faktoren leichter möglich werden.

Diese Faktoren erhöhen insgesamt nicht nur die Nachteile für diejenigen, die eigentlich von Human+ profitieren sollten (Geflüchtete im Sinne der Akutversorgung, bzw. BOS), sondern auch für manche der Stakeholdergruppen, die keine direkten Vorteile durch Human+ erfahren. Hierunter fallen beispielsweise Social-Media-UserInnen, welche als „unentgeltliche“ InformationslieferantInnen möglichst vor allen Nachteilen durch das System bestmöglich geschützt werden müssen. Das DESSI-Grundkriterium 2, wonach eine Sicherheitstechnik Grundrechte (z.B. das Recht auf Privatheit) wahren muss, scheint hier bereits wesentlich gefährdeter zu sein als wie in Szenario A. Eine weitere solche Gruppe ist die der In-Situ-Betroffenen: Durch den verdeckten Einsatz der In-Situ-Datenquellen (z.B. verdeckte Videosysteme an Hotspots) wird zwar keine Normalisierung des Verhaltens bewirkt, aber wie oben dargestellt stärker das Prinzip des Respekts vor der Autonomie verletzt.

⁷³ Auch in Szenario B sei angenommen, dass die Sammlung der Social-Media-Rohdaten durch einen Crawler in Übereinstimmung mit den AGBs der jeweiligen Social-Media-Plattform erfolgt, und daher die offiziell zur Verfügung gestellte Programmierschnittstelle bzw. API genutzt wird.

Während zwar auch in Szenario B die Beförderung vitaler Interessen von Geflüchteten durch einfachere Arbeitsumstände für BOS im Vordergrund steht, relativiert sich der Nutzen für BOS durch nachteilige Anwendungsumstände und schwächt damit das probabilistische „**Pro-Argument 1: Verbesserte Chance auf rechtzeitige und angemessene Versorgung**“.

Die Güterabwägung fällt in Szenario B somit insgesamt nachteiliger aus, und aus ethischer Sicht scheint damit die Zulässigkeit von Human+ deutlich fraglicher. Auch dem Metaprinzip des Negativen Utilitarismus folgend ist das Ausmaß an Leid für die von negativen Folgen des Human+ Einsatzes Betroffenen in Szenario A geringer als in Szenario B, sodass diesem Metaprinzip folgend Maßnahmen gesetzt werden sollten, dass im Falle einer Weiterentwicklung und spätere Anwendung der Human+ Technik (eher) der Stand von Szenario A als wie Szenario B erreicht wird.

6.8 EMPFEHLUNGEN (SCHRITT 5 IMPLEMENTIERUNG)

Im Ablaufschema von Bleisch und Huppenbauer (2014: 15) inkludiert die Behandlung von Implementierungsfragen (Schritt 5) eine Abschätzung der Möglichkeiten der Implementierung, Vorschläge zu Maßnahmen, wie die Implementierung der moralischen Lösung durchgeführt oder zumindest befördert werden kann (z.B. durch Anreizsysteme oder Übergangsfristen), sowie das Ausloten von anderen Perspektiven (z.B. politischen und wirtschaftlichen Perspektiven).

Im Rahmen der vorliegenden Arbeit wird für Schritt 5 nur ein reduzierter Themenumfang verfolgt, nämlich noch einmal die Empfehlungen aufzulisten, die in den vorherigen Abschnitten zur Behandlung der bei Human+ auftretenden moralischen Konflikte und Probleme bereits vorgeschlagen wurden. Wie sich gezeigt hat, hängt die Beurteilung der ethischen Zulässigkeit maßgeblich von einer Reihe von Faktoren ab, die erstens in der Technik selbst begründet sind und die in der Weiterentwicklung von TRL4 auf ein tatsächliches Anwendungsniveau noch in die eine oder andere Richtung fortgeführt werden können, oder aber von organisationalen Faktoren, die selbst bei einem marktreifen System unterschiedlich ausgestaltet werden können. Das Ausloten anderer Perspektiven entfällt aus Machbarkeitsgründen – weder kann im Rahmen der vorliegenden Arbeit umfassend das Meinungsbild von politischen Entscheidungsträgern bzw. Entscheidungsträgern in BOS, die auf der strategischen Ebene Entscheidungen fällen, eingeholt werden, noch können wirtschaftliche Ausblicke gegeben werden, da Human+ als TRL4 Forschungsprojekt von einem inhomogenen Forschungskonsortium durchgeführt wurde, und sich wirtschaftliche Verwertungsaspekte je nach fraglicher Komponente des Human+ Systems und der noch zu erfolgenden Weiterentwicklung schwer abschätzen lassen.

Aus **organisatorischer Sicht** lassen sich für einen späteren Human+ Echteinsatz bzw. die dafür notwendige Weiterentwicklung folgende konzeptionelle Klärungen als dringend notwendig empfehlen:

- Klärung und Festlegung der Anwendungsbedingungen und der AnwenderInnengruppe: Soll die Human+ Technik nur anlassweise verwendet werden, und kann dadurch ggf. die Anwendung BAO-Settings gekoppelt werden? Oder soll Human+ für dauerhaftes Monitoring in langfristig bestehende Strukturen (z.B. Ministerien oder andere BOS) eingegliedert werden? Wer genau soll zur AnwenderInnengruppe gehören? Welchen Kriterien müssen AnwenderInnen genügen (bzw. wie soll der AnwenderInnenkreis beschränkt werden)? Während in den obigen ethischen Analysen davon ausgegangen wurde, dass der AnwenderInnen-Kreis einen BOS-Konnex aufweist, so könnte bei genaueren Festlegungen auch stärker die Spezifika der einzelnen BOS bzw. das ihnen zugrunde liegende „Ethos“ miteinbezogen werden, was Human+ als sozio-technisches System umfassender beleuchten lässt.
- Klärung und Festlegung der Anwendungszwecke: Während allgemein als legitimer Human+ Verwendungszweck eine Verbesserung der humanitären Sicherheit in Flüchtlingslagen genannt werden kann, zeigte die Analyse zum Begriff „humanitäre Sicherheit“, dass der Begriff in einem weiten und einem engen Verständnis aufgefasst werden kann. Zudem ist bisher unterbestimmt, an welchen Kriterien sich etwa die Anwendung im Falle eines dauerhaften Monitorings ausrichten soll. Falls bestimmte Aspekte der Human+ Technik aus ethischer Perspektive nicht unbedenklich sind (z.B. das Crawlen von Social-Media-Daten vor dem Hintergrund des Respekts vor der Autonomie), dann sollte die Anwendung nicht willkürlich erfolgen, sondern sich an transparenten Kriterien orientieren, sodass die Notwendigkeit einer Anwendung gegenüber den Verletzungen der unterschiedlichen Prinzipien abgewogen werden kann. Außerdem dient dies dazu, weichen Sachzwängen vorzubeugen, und ist auch eine Voraussetzung für einen gegenüber der Öffentlichkeit transparenten Umgang mit der Human+ Technik.
- Formulierung eines COC für AnwenderInnen: Sobald die Anwendungsbedingungen, die AnwenderInnengruppe und die Anwendungszwecke für einen potentiellen Echteinsatz der Human+ Technik festgelegt sind, (dann erst) macht die Formulierung eines COC für die AnwenderInnen Sinn, um ihnen ein zusätzliches Hilfsmittel an die Hand zu geben, um legitime von missbräuchlichen Verwendungsweisen zu unterscheiden, und sich auch gegen etwaigen Druck von anderen BOS oder eigenen Vorgesetzten hinsichtlich einer illegitimen Verwendung entschieden wehren zu können.
- Anwendungsorientiertes Training für die AnwenderInnengruppe: Dieser Punkt zielt darauf ab, nicht auf eine intuitive oder „schnell zu erlernende“ Anwendbarkeit des Human+ Systems zu vertrauen, sondern die intendierten NutzerInnen zu Human+ FachanwenderInnen auszubilden, damit die Technik zu einem gut gehandhabten Hilfsmittel wird, und nicht unnötig die in BOS meist gut eingespielten Arbeitsabläufe durch ungeübten Umgang mit dem System behindert werden. Dies

trägt auch dazu bei, die Begrenzungen der Technik bzw. der erstellten Prognosen besser zu vermitteln und so erst eine aufgeklärte Human+ Anwendung zu ermöglichen.

Während die Umsetzung dieser vier Vorschläge aus ethischer Sicht zwar wünschenswert wäre, um weitere ethische Beurteilungen zu ermöglichen, function creep zu erkennen und womöglich zu verhindern, und Missbrauch einfacher durch organisationale Maßnahmen eindämmen zu können, ist aus praktischer Perspektive dieses Unterfangen zumindest fraglich. Da die Human+ Komponenten auf TRL 4 sind und die Weiterentwicklung über das Projekt hinaus in der Verantwortung der jeweiligen Projektpartner liegt, wird sich naturgemäß die Weiterentwicklung (außer bei erneuter Beteiligung öffentlicher Fördermittel) weniger an ethischen Kriterien als wie an ökonomischer Verwertbarkeit ausrichten. An dieser Stelle muss auch empfohlen werden, im Falle einer Weiterentwicklung und Anwendung der Human+ Technik die in der vorliegenden Arbeit nur probabilistischen Annahmen (z.B. zum Nutzen der durch die Human+ Technik erzeugten Prognosen für BOS) empirisch zu untersuchen.

In **technischer Hinsicht** können noch einmal folgende, v.a. in Abschnitt "6.6.2 Maßnahmen zur Verhinderung außerplanmäßiger Handlungen" (S. 128; für Details siehe weiter oben) schon erarbeitete Aspekte angeführt werden:

- Trennung von Rohdaten und aufbereitetem Informationsprodukt, mit nochmals eingeschränktem Zugriff auf die zu Grunde liegenden Rohdaten, falls diese etwa aus Privatheitsperspektive nicht unbedenklich sind.
- Beschränkte Speicherung von Rohdaten
- Beschränkung der Informationen im gemeinsamen Echtzeitlagebild
- Etablierung eines Rechte- und Rollenmanagements
- Nachvollziehbarkeit der Systemverwendung sicherstellen
- Systementwicklung gemäß den „guten Standards“ der IT-Sicherheit
- Verwendung von bereits bestehenden Video-Systemen für die In-Situ-bezogenen Datenquellen, anstatt zusätzliche Systeme hochzuziehen, um der Normalisierungsproblematik nicht zusätzliche Brisanz zu geben
- Auswahl solcher Social Media für die Social-Media-Datensammlung und Analyse, bei denen in besonderer Weise auf deren öffentlichen Charakter schon durch die UrheberInnen Wert gelegt wird (z.B. bei Twitter, aber auch im Falle öffentlicher Gruppen in andere Social-Media, z.B. auf Facebook)

7 SCHLUSS

Die vorliegende Masterarbeit im Rahmen des „Interdisziplinären Masterstudiums Ethik für Schule und Beruf“ verfolgte ein doppeltes Ziel: einerseits die vertiefte ethische Beurteilung des KIRAS Sicherheitsforschungsjprojekts Human+, mit welchem der Verfasser bereits in beruflichem Rahmen zu tun hatte. Andererseits die Erarbeitung eines Bewertungsframeworks hierfür, welches exemplarisch auf das genannte Human+ Projekt Anwendung fand.

Zunächst zum Bewertungsframework, welches methodisch die Grundlage für die inhaltliche Arbeit bildet: Eine Anforderung an das erarbeitete Framework war, dass es auch für die weitere berufliche Praxis im Rahmen von Sicherheitsforschungsjprojekten Anwendung finden kann. Zentral für die Erarbeitung des Bewertungsframeworks waren zunächst die sogenannten drei „Hauptfragen“ (vgl. Fenner 2010: 228), welche sich auf die Legitimität des primären Einsatzzweckes, das Ausmaß des Ergreifens von Gegenmaßnahmen für moralisch bedenkliche außerplanmäßigen Handlungsweisen, und die Folgen und Nebenwirkungen der planmäßigen Anwendung beziehen. Diese Hauptfragen wurden erweitert und vertieft, z.B. durch die Unterscheidung von unmittelbarem und mittelbarem primären Zweck, sowie durch die Unterfragen zur Hauptfrage 2. Zudem wurde durch Rückgriff auf verschiedene AutorInnen das Metaprinzip des Negativen Utilitarismus formuliert.

Diese drei Hauptfragen wurden eingebettet in das grundsätzliche Ablaufschema von Bleisch und Huppenbauer (2014), welches verschiedene Ablaufphasen vorschlägt (1. Analyse des Ist-Zustandes, 2. Benennung der moralischen Frage, 3. Analyse der Argumente, 4. Evaluation und Entscheidung, 5. Implementierung). Während die Phase Analyse des Ist-Zustandes sowie Implementierung die äußere Klammer bildeten, wurden für die Beantwortung von Hauptfrage 1 und 3 jeweils die Schritte 2 bis 4 des Ablaufschemas in einer Schleife durchlaufen, wobei die Hauptfragen als moralische Frage für Schritt 2 dienten. Hauptfrage 2 wurde davon abweichend in einem abgekürzten Prozess behandelt.

In den jeweiligen Schleifen für die Hauptfragen 1 bis 3 kamen sodann unterschiedliche Elemente der in Kapitel 5 der vorliegenden Arbeit vorgestellten Ansätzen und Frameworks der Technikethik zum Einsatz. Eine besondere Rolle spielten hier einerseits prima facie Prinzipien mittlerer Reichweite, wie sie dem Principlism von Beauchamp und Childress (2009) entnommen wurden, aber mit einer neuen Lesart versehen, wie sie aus IKT-bezogenen Zuspitzungen (z.B. dem Menlo-Report [vgl. Menlo 2013] oder dem Ethical Impact Assessment [vgl. Wright 2011]) des Principlism ersichtlich wurden. Strukturiert wurden die möglichen Auswirkungen zu den Folgen und Nebenfolgen von Human+ in der Beantwortung der Hauptfrage 3 zudem durch die Erstellung einer Ethical Matrix (vgl. Mepham et al 2006), welche Stakeholder und Prinzipien noch einmal explizit zusammenbrachte. Außerdem wurde an un-

terschiedlichen Stellen auf das Metaprinzip des Negativen Utilitarismus rekurriert, um in Abwägungssituationen zu entscheiden, in denen sich die Alternativen besonders hinsichtlich ihrer negativen Auswirkungen bzw. dem (allerdings nur probabilistisch abzuschätzenden) Ausmaß an verursachtem Leid deutlich unterscheiden.

Da bei Human+ während der Laufzeit des Projekts zudem die tatsächlichen Schritte der Projektpartner nach Abschluss des Projekts in Richtung einer späteren (etwaigen) Weiterentwicklung vom TRL4 zu einem tatsächlich anwendungsreifen Produkt mit höherem TRL nicht vollständig abschätzbar waren, und auch die späteren Einsatzbedingungen nur im Sinne von konditionalen Bezügen (vgl. Grunwald 2010: 151) im Sinne von „Was wäre wenn?“ analysiert werden konnten, wurden für die Beantwortung der Hauptfrage 3 zwei Szenarien A und B vorgestellt, die sich hinsichtlich der ethischen Zulässigkeit des Projekts strukturell breit unterscheiden und so die Pole eines Spektrums bilden. Szenario A umreißt dabei eine Entwicklung, die die ethisch relevanten Weiterentwicklungs- und Anwendungsempfehlungen berücksichtigt, und Szenario B skizziert die Auswirkungen, wenn diese Beachtung unterlassen wird.

Insgesamt wurde im Rahmen der vorliegenden Arbeit somit ein Ablaufschema für die ethische Analyse konstruiert, welches zwar vornehmlich mit IKT-Projekten im Blick erstellt wurde, aber durch die Verwendung der sehr allgemein formulierten Prinzipien mittlerer Reichweite und die nicht speziell auf IKT-ausgerichteten drei Hauptfragen auch für andere Technikkontexte genutzt werden kann.

Nach dieser methodischen Zusammenfassung, kann nun noch einmal eine Synopse über die inhaltlichen Ergebnisse, d.h. die ethische Beurteilung im engeren Sinne, gegeben werden. Auch hier werden die Ergebnisse entlang der drei Hauptfragen strukturiert.

Hauptfrage 1: Sind die Handlungsmöglichkeiten, die durch diese Technik eröffnet oder verbessert werden, an sich ethisch legitim? (primärer Zweck)

Der sogenannte **unmittelbare** primäre Zweck wurde in der Erstellung von Migrationsprognosen und einem gemeinsamen Echtzeit-Lagebild ermittelt. Aus ethischer Sicht wurde festgestellt, dass zwar ein Nützlichkeitsargument für die Zulässigkeit dieser Zwecke spricht, die bereitgestellten Informationsprodukte aber letztendlich multifunktional verwendbar sind, und die späteren Anwendungsbedingungen nicht ausreichend festgelegt sind. Daher wurde zusätzlich zum unmittelbaren primären Zweck auch der **mittelbare** primäre Zweck betrachtet, welcher in der Verbesserung von humanitärer Sicherheit und einer besseren Vorbereitung auf Migrationswellen zu sehen ist. Um den mittelbaren primären Zweck zu analysieren, wurde zunächst der Begriff „humanitäre Sicherheit“ nachgeschärft und für den Kontext der vorliegenden Analyse ein erweiterter Begriff von humanitärer Sicherheit angenommen. Anschließend wurden zwei mögliche Varianten der Anwendung von Human+ unterschieden: Anwendung innerhalb von akuten Krisen- und Katastrophenbedingungen (was dann die Einflechtung in BAO

ermöglicht, und auf eine zeitlich eingeschränkte Anwendung hinausläuft), oder aber auch außerhalb von Krisen- und Katastrophensbedingungen (was „ausufernde“ Technikanwendung implizieren kann, aber auch langfristige Monitoring-Aktivitäten zur Früherkennung ermöglicht). Der Einsatz von Human+ für die Verbesserung der humanitären Versorgung und der besseren Vorbereitung wurde als ethisch legitim beurteilt, weil es um einen Beitrag für die Erfüllung von grundlegenden Bedürfnissen von Geflüchteten geht, wenngleich das Argument des function creep auf die Gefahr hinweist, dass diese Technik auch anderweitig genutzt werden könnte, womit sich allerdings Hauptfrage 2 beschäftigt.

Hauptfrage 2: Werden ausreichend Maßnahmen ergriffen, um alle moralisch bedenklichen außerplanmäßigen Handlungsweisen auszuschließen, die sich damit abseits vom primären Zweck durchführen lassen?

Unterfrage 2a: Welche außerplanmäßigen Handlungsweisen abseits vom primären Zweck lassen sich durchführen?

Unterfrage 2b: Welche davon sind moralisch bedenklich?

Im Abschnitt zu Hauptfrage 2 wurde eine Reihe von möglichen außerplanmäßigen Handlungsweisen, die durch Human+ eröffnet werden, eruiert und hinsichtlich ihrer moralischen Bedenklichkeit bewertet. Daraufhin wurde eine Reihe von technischen und organisationalen Gegenmaßnahmen skizziert. Aufgrund des niedrigen TRL4, bei welcher die konkreten Funktionen eines möglichen Endprodukts noch nicht vollumfänglich realisiert sind und v.a. die konkreten Anwendungsbedingungen nur umrissen werden können, kann Hauptfrage 2 nur tentativ in dem Sinne bejaht werden, dass zumindest eine Reihe von Maßnahmen skizziert werden kann, die außerplanmäßige Handlungsweisen erschweren, wovon ein Teil dieser möglichen Maßnahmen bereits während der Projektlaufzeit in der Systemarchitektur von Human+ grundgelegt wurde.

Hauptfrage 3: Gefährden die Folgen und Nebenwirkungen der planmäßigen Techniknutzung die Lebensbedingungen oder die Lebensqualität gegenwärtiger oder zukünftiger Menschen?

Für die Beantwortung dieser Frage wurde eine Ethical Matrix erstellt, welche die möglichen Folgen und Nebenwirkungen der Human+ Anwendung auf die in „Tabelle 11: Stakeholder“ (S. 110) identifizierten Betroffenen vor dem Hintergrund der unter „6.3 Prinzipien mittlerer Reichweite des Human+ Bewertungsframeworks“ (vgl. S. 95ff) dargestellten Prinzipien strukturiert anführt und beleuchtet. Diese Betroffenheiten wurden in Form von „Argumenten“ schriftlich dargelegt, welche sowohl negative (Verletzungen der Prinzipien) als auch positive (Berücksichtigung der Prinzipien) umfassen. Zahlreiche Pro- und Kontra-Argumenten wurden angeführt und miteinander abgewogen, und in Form einer Gegenüberstellung von zwei Szenarien ethisch beurteilt (Szenario A, in welchem die möglichen Maßnahmen zur Entkräftung der Gegenargumente ergriffen wurden, und Szenario B, in welchem diese fehlen).

Während am Ende des Abwägungsprozesses in Szenario A die Hauptfrage 3 so beantwortet wurde, dass die Gefährdungen für Lebensbedingungen und Lebensqualität durch Human+ geringfügig genug sind, dass die bestehenden ethischen Konfliktpunkte durch die Vorteile überwogen werden können, wurde im Falle von Szenario B das Fazit kritischer gesehen, sodass von einer Human+ Verwendung bzw. Weiterentwicklung aus ethischer Sicht abgeraten würde.

Die ethische Zulässigkeit von Human+ hängt somit am Ende von der ernsthaften Realisierung und Befolgung einer Reihe von Weiterentwicklungs- und Anwendungsempfehlungen ab, wie sie in Kapitel „6.8 Empfehlungen (Schritt 5 Implementierung)“ (S. 157ff) bereits zusammengefasst wurden.

Abschließend sei auf eine Reihe von methodischen und inhaltlichen Beschränkungen der vorliegenden Arbeit verwiesen. Zunächst zu den **methodischen Beschränkungen**: Da eine Anforderung an das Framework die Anwendbarkeit auf andere Sicherheitsforschungsprojekte war, mit denen der Verfasser beruflich zu tun hat, wurde besonderer Wert darauf gelegt, dass die ethische Analyse in klaren Schritten erfolgt, und ein nachvollziehbarer „Plan“ hinter den Ausführungen erkennbar wird. Die Kehrseite davon ist, dass das entworfene Framework generisch und schematisch wirkt. Generischer Schematismus ist somit zugleich Stärke wie auch Schwäche der vorliegenden Arbeit.

Zweitens finden prima facie Prinzipien mittlerer Reichweite Anwendung, wie sie im Principlism von Beauchamp und Childress (2009) grundgelegt sind. Neben allen Vorteilen und inhaltlichen Stärken handelt sich die vorliegende Arbeit damit auch den Nachteil ein, dass grundlegende Kritik am Principlism auch auf sie zutrifft, nämlich die Begründung dieser prima facie Prinzipien in der „gemeinsamen Moral“. Wer das Vorliegen einer solchen gemeinsamen Moral (welche sich nach Beauchamp und Childress letztendlich in Form von relativ wenigen, sehr abstrakten Prinzipien niederschlägt) anzweifelt, der wird auch die Arbeit mit den Prinzipien mittlerer Reichweite wenig überzeugend finden – oder aber eine andere Begründung für diese fordern.

Ganz grundsätzlich ist aber zu sagen, dass ethische Werkzeuge voraussetzen, dass die AnwenderInnen dieser Werkzeuge Teil der gleichen moralischen Gemeinschaft sind (vgl. Moula und Sandin 2015: 271). Nach Ansicht des Verfassers der vorliegenden Arbeit kann diese Aussage ausgeweitet werden auch auf die intendierten RezipientInnen der ethischen Analysen, da ansonsten die solcherart ermittelten Befunde zur ethischen Akzeptabilität keine argumentative Wirkung und damit auch keine Akzeptanz und Chance auf Berücksichtigung durch die verantwortlichen Entscheider finden können.

Während für die vorliegende Arbeit der Weg über Prinzipien mittlerer Reichweite gewählt wurde, wäre es alternativ auch möglich gewesen, „Wertesysteme“ hierfür heranzuziehen, und diese stattdessen etwa in die Ethical Matrix einzuflechten. So stellt etwa der Verein Deutscher Ingenieure in der VDI

Richtlinie 3780 ein Werteoktagon⁷⁴ zur Verfügung, welches auch häufige Instrumental- und Konkurrenzbeziehungen zwischen den Werten illustriert (vgl. Ropohl 2016: 224). Eine mögliche Kritik an solchen Wertedarstellungen ist allerdings, dass streng genommen solche Wertesysteme nicht universell, sondern für jede objektive Sachlage gesondert zu konstruieren seien (unter Einbezug aller von einer Technik betroffenen sozialen Akteure und Gruppen), und zudem dass beteiligte Entscheidungssubjekte (z.B. Individuen, Korporationen, Gesellschaften) die Beziehungen zwischen den Werten anders interpretieren⁷⁵ und auch die Wertpräferenzen anders setzen können (vgl. Ropohl 2016: 225). Da das VDI Werteoktagon durch empirische Werteforschung ermittelt wurde, kann daran kritisiert werden, dass es sich nur um Momentaufnahmen handelt, die normativ schwer zu begründen sind, will man sich nicht in einen Sein-Sollen-Fehlschluss begeben. Grunwald verweist bezüglich des VDI Werteoktagons ebenfalls auf die Kritik aus dem Bereich der philosophischen Ethik, wonach der Bezug auf die vom VDI definierten Werte die Gefahr eines naturalistischen Fehlschlusses aufweisen, weil sie das faktisch Akzeptierte als nicht-hinterfragte normative Basis für Zukünftiges ausgeben (vgl. Grunwald 2010: 111). Zudem bestehe die Gefahr von Wertkonflikten, für deren Lösung die angegebenen Werte selbst nicht ausreichen.

Dieser Art von Relativismus entgeht der Ansatz einer „gemeinsamen Moral“ (Beauchamp und Childress 2009) bzw. eines „ethischen Minimalkonsenses“ (Ropohl 2016: 314), zu welchem laut Ropohl auch der Negative Utilitarismus gezählt werden kann.

In jedem Fall bestimmt die Auswahl der Prinzipien oder Werte, was bei der ethischen Analyse in den Blick genommen wird. Ein Anspruch auf Vollständigkeit kann daher auch im Falle umfangreicher Analysen nicht erhoben werden, was gleichermaßen auf Arbeiten im Rahmen der Technikethik wie auch der TA zutrifft. Selektivität ist somit eine unvermeidbare methodische Beschränkung.

Angesichts der Erkenntnisziele des skizzierten Vorgehens wurde außerdem dediziert von einem „moral presentism“ ausgegangen (Boenink et al 2010: 2): Das (langfristig) dynamische Verhältnis von Moral und Technikentwicklung wird daher weitgehend (bewusst) ausgeblendet, weil es um Weiterentwicklungsempfehlungen für die nächsten paar Jahre geht. Dementsprechend wenig Bezugspunkte hat es zu den „techno-ethischen Zukunftsszenarien“ gegeben.

Als letztes sei noch darauf verwiesen, dass der Verfasser der vorliegenden Arbeit zwar nach allen Möglichkeiten versucht hat, den universelle Standpunkt der Moral (vgl. Fußnote 37 auf Seite 89 der vorlie-

⁷⁴ Die Dimensionen hierbei sind Gesellschaftsqualität, Umweltqualität, Gesundheit, Sicherheit, Funktionsfähigkeit, Wirtschaftlichkeit (einzelwirtschaftlich), Wohlstand (gesamtwirtschaftlich) und Persönlichkeitsentfaltung (vgl. Ropohl 2016: 224).

⁷⁵ Auch wie die jeweiligen Wertdimensionen zueinander gewichtet werden sollen, ist umstritten und erfordert im Zuge der TA spezielle konsensbildende Verfahren (vgl. Simonis 2013a: 14).

genden Arbeit) in der Abwägung der Argumente gegeneinander einzunehmen, dass aber nichtsdestotrotz andere Personen (mit anderen Voraussetzungen und Ansichten) die Gewichtung anders vornehmen könnten. Umso wichtiger war es, die Sammlung und Analyse von möglichen Pro- und Kontra-Argumenten von ihrer Evaluation zu trennen, um es den LeserInnen der vorliegenden Arbeit zu erleichtern, ihr eigenes Urteil zu fällen.

Zu den **inhaltlichen Beschränkungen**: Die Human+ Technik hat sich als komplexes Zusammenspiel unterschiedlicher Komponenten erwiesen (Social-Media-Auswertungen, In-Situ-Auswertungen, Satellitendaten-Auswertungen, gemeinsames Echtzeitlagebild), welche für sich genommen jeweils eigene Probleme aufwerfen können, und zudem im Rahmen des Human+ Forschungsprojekt exemplarisch miteinander verwoben wurden. Für die vorliegende Analyse machte dies Abstraktionen notwendig, um bewältigbar zu sein. Die Entwicklungen wurden während der Projektlaufzeit außerdem „lediglich“ bis TRL4 betrieben, welches einen „Funktionsnachweis der Technologie im Labor(-maßstab) auf Systemebene“ (FFG 2019: 28) erfordert. Dies wurde während des Projekts in Form eines Demonstrator-tests im Rahmen einer künstlichen Umgebung mit teilweise speziell aufbereiteten Inputs zum Test der Algorithmen und Softwarelösungen eingelöst. Damit ist die Technik aber noch weit von einer tatsächlichen Anwendungsreife entfernt. Dies machte es schwierig, das eigentliche Bewertungsobjekt auszuwählen. Soll lediglich die Technik wie sie im Antrag skizziert ist, analysiert werden? Oder aber die Technik, wie sie während des Projekts entwickelt wurde? Oder aber die Technik, wie sie bei einer Weiterentwicklung in Richtung tatsächlicher Anwendungsreife zu erwarten wäre? Für den Rahmen der vorliegenden Arbeit wurde ein gemischter Weg gewählt: einerseits wurde stark auf den Antrag rekurriert, um über Ziele und Zwecke der Technik einen Anhaltspunkt zu haben, andererseits wurde für die Skizzierung der Technik selbst auf während der Projektlaufzeit verfasste Dokumente wie das Technische Pflichtenheft oder die Dokumentation zum Demonstrator-test zurückgegriffen, und mögliche Weiterentwicklungsperspektiven ebenfalls in den Blick genommen. Dieser Schwierigkeit ist grundsätzlich allerdings jede technikethische Beurteilung von gerade in Entwicklung begriffenen Techniken unterworfen. Dennoch unterliegen die Bewertungen der vorliegenden Arbeit damit einer gewissen Vorläufigkeit, da die Weiterentwicklung in Richtung eines höheren TRL außerhalb des Human+ Konsortiums stattfindet, und eine Arbeit wie die vorliegende nur Hinweischarakter für die ethische Beurteilung eines in einigen Jahren womöglich tatsächlich zur Anwendungsreife gebrachten Systems haben kann, das womöglich auch nur auf einigen Komponenten von Human+ basiert und einen völlig anderen Namen tragen wird. Dass die Ergebnisse wenigstens Hinweischarakter für ethisch verträglichere Weiterentwicklungen haben können, trifft zumindest dann zu, wenn (wie im hier vorliegenden Fall) ethische Probleme sich in einem solchen Rahmen halten, dass sie noch Fragen der Abwägung sein können, und die Technik nicht bereits inhärent solch gravierende ethische Schwierigkeiten aufwirft, dass diese auch

bei einer Weiterentwicklung durch technische oder organisationale Maßnahmen nicht in den Griff zu bekommen sind. In einem solchen Fall wäre Jonas' Credo nicht einhaltbar, so zu handeln (und stellen wir fest, auch Forschen ist Handeln), dass „[...] die Wirkungen deiner Handlungen verträglich sind mit der Permanenz echten menschlichen Lebens auf der Erde“ (Jonas 1979, S. 36; zitiert nach Grunwald 2010: 214). Solche gravierenden ethischen Probleme zeigten sich in der Behandlung der Hauptfragen für Human+ jedenfalls nicht.

Ebenfalls kaum berücksichtigt wurden rechtliche Aspekte, wenngleich das Schema von Bleisch und Huppenbauer (2014) dies als Teil der Projektarbeit vorsieht. Rechtliches kam allenfalls im Rahmen von grundrechtlichen Betrachtungen ins Bild.

Eine weitere inhaltliche Beschränkung ist der schwache Einbezug der Eigenperspektive der identifizierten Stakeholdergruppen. Während ein solcher Einbezug noch am ehesten für die potentiellen Human+ AnwenderInnen-Gruppen durch die Gespräche und Erhebungen während der Human+ Projektlaufzeit zutrifft, wurden mögliche Vorbehalte und Sichtweisen etwa von Geflüchteten, Social-Media-UserInnen und der Bevölkerung nur fiktiv oder durch Rekurs auf die Literatur erschlossen. Dadurch kann nicht das Anliegen einer partizipativen TA eingelöst werden. Die Substituierung über den Rekurs auf die Literatur stellt somit keinen Idealfall dar. Die Einholung direkter Rückmeldungen von StakeholderInnen könnte aber grundsätzlich in einer erneuten Anwendung des erarbeiteten Frameworks im Rahmen anderer Projekte bei Vorliegen entsprechender Forschungsmittel und frühzeitiger Berücksichtigung durchgeführt werden, und ist also kein grundsätzlicher Mangel des Frameworks, sondern lediglich der beschränkten Möglichkeiten im Rahmen dieser Masterarbeit (die großteils ja nach Abschluss des Human+ Projekts zwecks reflexionstheoretischer Vertiefung erstellt wurde). Das frühzeitige Einholen der Perspektive der Betroffenen während der Projektlaufzeit eines solchen Sicherheitsforschungsprojekts wird aber häufig dadurch erschwert, dass zu Beginn solcher Projekte meistens nur frühe grobe Konzepte feststehen, und technische Entscheidungen und Konkretisierungen erst während der Projektlaufzeit erfolgen, sodass hier eine Variante des „Collingridge-Dilemma“⁷⁶ (auch als Steuerungs- oder Kontrolldilemma bekannt) schlagend wird (vgl. Grunwald 2010: 165). Die vorliegende Variante dieses Dilemmas könnte so beschrieben werden: Wenn man die Perspektive von Betroffenen zu einer Technik einholen will, gelingt dies am besten, wenn die Technik bereits vorliegt – dann aber lässt sich diese Perspektive nur mehr begrenzt berücksichtigen. Wenn man allerdings versucht, frühzeitig Perspektiven (und dadurch Argumente von Betroffenen) zu erheben, steht man vor der Schwierigkeit, lediglich

⁷⁶ Das ursprüngliche Collingridge-Dilemma besagt, dass die Aussichten auf sicheres Folgenwissen umso besser werden, je weiter entwickelt eine Technik ist, je besser Produktionsbedingungen, Nutzungskontexte und Entsorgungsverfahren bekannt sind. Technikfolgen lassen sich am besten messen, wenn sie schon eingetreten sind, allerdings können die Technik und die Technikfolgen dann auch nicht mehr gestaltet werden (vgl. Grunwald 2010: 165). Sowohl sicheres Folgenwissen wie auch Gestaltungsspielraum sind damit nicht gleichzeitig zu haben.

vage konzeptionelle Beschreibungen der fraglichen Technik parat zu haben, und auf etwaige Rückfragen der Betroffenen nur konditional oder gar nicht antworten zu können.

Auf eine letzte inhaltliche Begrenzung sei noch verwiesen: TA erfordert (zumindest im Rahmen der sogenannten „probleminduzierten TA“) in der Bewertung von Technik, alle fraglichen technischen Lösungen für ein bestimmtes gegebenes (gesellschaftliches) Problem oder Bedürfnis zu berücksichtigen (vgl. Ropohl 2016: 183). Im Rahmen der vorliegenden Arbeit hätte das bedeutet, für den mittelbaren Zweck von Human+ (Verbesserung der humanitären Sicherheit) auch alternative Lösungsmöglichkeiten in Betracht zu ziehen, um die in Human+ angestrebten Lösungen zu kontrastieren und zu vergleichen. So ein Anspruch konnte hier nicht eingelöst werden, sodass bei der vorliegenden Analyse eher ein Fall von „technikinduzierter“ TA vorliegt, welche bereits bei einer konkreten oder sich abzeichnenden Lösung ansetzt und diese bewertet – mit einem ungleich eingeschränkteren Aussageanspruch.

8 LITERATURVERZEICHNIS

Akerma, Karim, 2000: Verebben der Menschheit? Neganthropie und Anthropozidee. Freiburg / München: Verlag Karl Alber.

Alvarez, Allen A.A., 2009: The cross-cultural importance of satisfying vital needs. *Bioethics*, Jg. 23, Heft 9, S. 486-496.

Baur-Ahrens, Andreas; Krüger, Marco, 2015: How Smart is „Smart Security“? Exploring Data Subjectivity and Resistance. Final Report. Tübingen: IZEW.

Beauchamp, Tom L.; Childress, James F., 2009: *Principles of Biomedical Ethics*. Sixth Edition. New York und Oxford: Oxford University Press.

Beier, Christoph; Messner, Dirk; Preuß, Hans-Joachim [Hrsg.], 2020: Globale Wanderungsbewegungen. Beiträge der internationalen Zusammenarbeit zum Umgang mit Flucht und Migration. Wiesbaden: Springer VS.

Belmont 1978a: The Belmont Report. Ethical Principles and Guidelines for the Protection of Human Subjects of Research. o.O.: The National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research. Online abrufbar unter https://videocast.nih.gov/pdf/ohrp_belmont_report.pdf, zuletzt besucht am 25.3.2021.

Belmont 1978b: The Belmont Report. Ethical Principles and Guidelines for the Protection of Human Subjects of Research. Appendix. Volume I. o.O.: The National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research. Online abrufbar unter https://videocast.nih.gov/pdf/ohrp_appendix_belmont_report_vol_1.pdf, zuletzt besucht am 25.3.2021.

Belmont 1978c: The Belmont Report. Ethical Principles and Guidelines for the Protection of Human Subjects of Research. Appendix. Volume II. o.O.: The National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research. Online abrufbar unter https://videocast.nih.gov/pdf/ohrp_appendix_belmont_report_vol_2.pdf, zuletzt besucht am 25.3.2021.

Bentham, Jeremy, 1948 [1823]: *A fragment on Government and An Introduction to the Principles of Morals and Legislation*. Oxford: Basil Blackwell.

Bleisch, Barbara; Huppenbauer, Markus, 2014: *Ethische Entscheidungsfindung. Ein Handbuch für die Praxis*. Zürich: Versus Verlag.

Boenink, Marianne; Swierstra, Tsjalling; Stemerding, Dirk, 2010: Anticipating the Interaction between Technology and Morality: A Scenario Study of Experimenting with Humans in Bionanotechnology. *Studies in Ethics, Law, and Technology*, Jg. 4, Heft 2, Artikel 4, S. 1-38.

Brey, Philip A.E., 2012: Anticipating ethical issues in emerging IT. *Ethics and Information Technology*, Jg. 14, Heft 4, S. 305-317.

Bröchler, Stephan; Simonis, Georg; Sundermann, Karsten, 1999: *Handbuch Technikfolgenabschätzung*. Band 1. Berlin: Edition Sigma.

Čas, Johann; Kaufmann; Mareile, 2012 [Hrsg.]: Deliverable 2.6 DESSI System of Criteria. Online abrufbar unter http://securitydecisions.org/download/9/system_of_criteria.pdf, zuletzt besucht am 23.3.2021.

Chadwick, Ruth [Hrsg.], 2012: *Encyclopedia of Applied Ethics*. London: Academic Press.

Cotton, Matthew, 2009: Evaluating the 'ethical matrix' as a radioactive waste management deliberative decision-support tool. *Environmental Values*, Jg. 18, Heft 2, S. 153-176.

Cusumano, Eugenio; Villa, Matteo, 2019: Sea Rescue NGOs: A Pull Factor of Irregular Migration? Florence: European University Institute. Online abrufbar unter <https://op.europa.eu/en/publication-detail/-/publication/66c3ef7b-64e0-11ea-b735-01aa75ed71a1>, zuletzt besucht am 14.5.2021.

Daase, Christopher, 2010: Wandel der Sicherheitskultur. Aus *Politik und Zeitgeschichte*, Jg. 50, S. 9-16.

Davey, Caroline L.; Wootton, Andrew B., 2017: *Design Against Crime. A Human-Centred Approach to Designing for Safety and Security*. Abdingdon und New York: Routledge.

de Jong, Sara; Messinger, Irene; Schütze, Theresa; Valchars, Gerd, 2017: Migrationsmanagement: Praktiken, Intentionen, Interventionen. *Journal für Entwicklungspolitik*. Jg. 33, Heft 1, S. 4-21.

Dekker, Rianne; Engbersen, Godfried;. Klaver, Jeanine; Vonk, Hanna, 2018: Smart Refugees: How Syrian Asylum Migrants Use Social Media Information in Migration Decision-Making. *New Media + Society*, Jg. 4, Heft 1, S. 1-11.

DerStandard 2015a: Online abrufbar unter <https://www.derstandard.at/story/2000020310006/erschuetternde-lage-in-traiskirchen-fluechtlinge-werden-nicht-mehr-aufgenommen>, zuletzt besucht am 11.5.2021.

DerStandard 2015b: Online abrufbar unter <https://www.derstandard.at/story/2000022034669/nickelsdorf-bis-zu-5-000-fluechtlinge-erwartet>, zuletzt besucht am 11.5.2021.

DerStandard 2015c: Online abrufbar unter <https://www.derstandard.at/story/2000023313571/flaute-bei-fluechtlingen-nur-von-kurzer-dauer>, zuletzt besucht am 11.5.2021.

DerStandard 2020a: Online abrufbar unter <https://www.derstandard.at/story/2000113739402/oes-terreich-und-italien-verhinderten-neustart-von-sophia>, zuletzt besucht am 14.5.2021.

DESSI 2011: o.V.: Deliverable 2.3 Dimensions in Security Investments. Online abrufbar unter http://securitydecisions.org/download/8/dimensions_in_security_investments.pdf, zuletzt besucht am 23.3.2021.

DHPol 2021: Eintrag zu Human+ auf der Seite der Deutschen Hochschule der Polizei. Online abrufbar unter https://www.dhpol.de/departements/departement_II/FG_II.1/projekt-human.php, zuletzt besucht am 17.7.2021.

DIVSI 2015: Deutsches Institut für Vertrauen und Sicherheit im Internet, 2015: Allgemeine Geschäftsbedingungen (AGB) von Kommunikationsdienstleistern. Internet-repräsentative quantitative Online-Panelbefragung. Online abrufbar unter https://www.divsi.de/wp-content/uploads/2015/10/2015-10-22_DIVSI_AGB-Umfrage_Charts.pdf, zuletzt besucht am 25.6.2021.

Dudek, Karlheinz, 2020: Integrierte Stabsarbeit am Beispiel der Landespolizeidirektion Wien. In: Kern, Eva-Maria; Richter, Gregor; Müller, Johannes C.; Voß, Fritz-Helge [Hrsg.]: Einsatzorganisationen. Erfolgreiches Handeln in Hochrisikosituationen. Wiesbaden: Springer Gabler, S. 223-234.

Eberle, Jan, 2019: Schutzsuchende. Ein Konzept zur Quantifizierung des Bestands an Ausländerinnen und Ausländern, die sich aus humanitären Gründen in Deutschland aufhalten. Wirtschaft und Statistik, Ausgabe 1 / 2019, S. 19-34.

ETICA D.2.2: ETICA Projekt Deliverable D.2.2 Normative Issues Report. Online abrufbar unter <https://www.etica-project.eu/deliverable-files>, zuletzt besucht am 12.6.2021.

Europäische Kommission 2021: Website zur Ethical Appraisal Procedure. Online abrufbar unter https://ec.europa.eu/research/participants/docs/h2020-funding-guide/cross-cutting-issues/ethics_en.htm, zuletzt besucht am 26.3.2021.

EU 2012: Charta der Grundrechte der Europäischen Union. (2012/ C 326/ 02). Online abrufbar unter <https://eur-lex.europa.eu/legal-content/DE/TXT/PDF/?uri=CELEX:12012P/TXT>, zuletzt besucht am 9.5.2021.

EU 2014: VERORDNUNG (EU) Nr. 536/2014 des europäischen Parlaments und des Rates vom 16. April 2014 über klinische Prüfungen mit Humanarzneimitteln und zur Aufhebung der Richtlinie 2001/20/EG. Online abrufbar unter <https://eur-lex.europa.eu/legal-content/de/TXT/?uri=CELEX:32014R0536>, zuletzt besucht am 8.4.2021.

EU 2016: Verordnung (EU) 2016/679 des Europäischen Parlaments und des Rates vom 27. April 2016 zum Schutz natürlicher Personen bei der Verarbeitung personenbezogener Daten, zum freien Datenverkehr und zur Aufhebung der Richtlinie 95/46/EG (Datenschutz-Grundverordnung). Online abrufbar unter <https://eur-lex.europa.eu/legal-content/DE/TXT/PDF/?uri=CELEX:02016R0679-20160504&from=EN>, zuletzt besucht am 28.7.2021.

FFG 2018: Eintrag zu Human+ in der FFG Datenbank. Online abrufbar unter <https://projekte.ffg.at/projekt/2958460>, zuletzt besucht am 14.1.2022.

FFG 2019: Leitfaden für kooperative F&E-Projekte. Wien: Österreichische Forschungsförderungsgesellschaft. Online abrufbar unter https://www.ffg.at/sites/default/files/dok/il_kooperativfueprojekte_v32_bf.pdf, zuletzt besucht am 9.5.2020.

Fenner, Dagmar, 2010: Einführung in die Angewandte Ethik. Tübingen: Francke Verlag.

Feuerwehr Kernen 2018: <https://www.feuerwehr-kernen.de/2018/10/03/vostbw-gestartet-auch-feuerwehren-aus-dem-rems-murr-kreis-beteiligt/>, zuletzt besucht am 28.5.2021.

Frankena, William K., 1994: Analytische Ethik. Eine Einführung. München: Deutscher Taschenbuch Verlag.

Friedman, Batya; Kahn, Peter H.; Borning, Alan, 2006: Value Sensitive Design and Information Systems. In: Zhang, Ping; Galletta, Dennis [Hrsg.]: Human-computer interaction in management information systems: Foundations. Armonk, New York: M.E. Sharpe, S. 348-372.

Gethmann, Carl Friedrich; Sander, Thorsten, 1999: Rechtfertigungsdiskurse. In: Grunwald, Armin; Saupe, Stephan [Hrsg.]: Ethik in der Technikgestaltung. Berlin Heidelberg New York: Springer Verlag, S. 117-151.

GFDS 2015: <https://gfds.de/wort-des-jahres-2015/>, zuletzt besucht am 12.12.2021.

Gillon, Raanan, 1994a: Principles of Healthcare Ethics. London: Wiley.

Gillon, Raanan, 1994b: The Four Principles Revisited – A Reappraisal. In: Gillon, Raanan [Hrsg.]: Principles of Healthcare Ethics. London: Wiley, S. 319-333.

Gosh, Bimal, 2007: Managig migration: Whither the Missing Regime? In: Pécoud, Antoine; de Guchteneire, Paul [Hrsg.]: Migration without Borders. Essays on the Free Movement of People. Oxford: Berghan Books, S. 97-118.

Gough, Clair; Boucher, Philip, 2013: Ethical attitudes to underground CO2 storage: Points of convergence and potential faultlines. International Journal of Greenhouse Gas Control, Jg. 13, S. 156-167.

Grunwald, Armin, 1996: Die Bewältigung von Technikkonflikten. Theoretische Möglichkeit und praktische Relevanz einer Ethik der Technik in der Moderne. Zeitschrift für Philosophische Forschung, Band 50, Heft 3, S. 442-457.

Grunwald, Armin, 1999: Technology Assessment or Ethics of Technology? Reflections on Technology Development between Social Sciences and Philosophy. Ethical Perspectives, Jg. 6, Heft 2, S. 170-182.

Grunwald, Armin, 2010: Technikfolgenabschätzung – eine Einführung. Berlin: edition sigma.

Grunwald, Armin, 2012: Technikzukünfte als Medium von Zukunftsdebatten und Technikgestaltung. Karlsruhe: Karlsruher Institut für Technologie (KIT).

Grunwald, Armin; Saupe, Stephan [Hrsg.], 1999: Ethik in der Technikgestaltung. Berlin Heidelberg New York: Springer Verlag.

Höffe, Ottfried, 2008: Einleitung. In: Höffe, Ottfried [Hrsg.]: Einführung in die utilitaristische Ethik. Tübingen und Basel: A Francke Verlag, S. 7-51.

Höffe, Ottfried [Hrsg.], 2008: Einführung in die utilitaristische Ethik. Tübingen und Basel: A Francke Verlag.

Hoffmann, Sophia, 2017: Humanitarian security in Jordan's Azraq Camp. Security Dialogue, Jg. 48, Heft 2, S. 97-112.

Human+ 2016: Projektbeschreibung für Förderungsansuchen. Salzburg: Universität Salzburg.

Human+ 2018a: o.A.: Kumulativer Anforderungskatalog aller Bedarfsträger an ein Informations- und Darstellungssystem. Projektinternes Deliverable aus Arbeitspaket 2 (de) / 3 (at). o.O.

Human+ 2018b: o.A.: Arbeitspaket 2 / Task 2.1: Katalog der bei einer Flüchtlingslage relevanten Bedarfsträger. Projektinternes Deliverable aus Arbeitspaket 2 / Task 2.1. o.O.

Human+ 2018c: o.A.: Bedarfskatalog polizeiliche Gefahrenabwehr. Projektinternes Deliverable aus Arbeitspaket 2. o.O.

Human+ 2018d: o.A.: Pflichtenheft. Projektinternes Deliverable aus Arbeitspaket 3. o.O.

Human+ 2019a: o.A.: Bericht zur HUMAN+ Übungs- und Evaluierungsveranstaltung. Projektinternes Deliverable aus Arbeitspaket 5. o.O.

Human+ 2019b: o.A.: Transkript zum ExpertInnen-Interview mit einem Mitarbeiter am Institut für Technikfolgenabschätzung in Wien. o.O.

Innerhofer, Christian, 2018: Verhaltenssteuerung durch Sachzwänge. Handlungen analysieren, verstehen und beeinflussen. Wiesbaden: Springer Fachmedien.

- IOM 2019: Glossary on Migration. Geneva: International Organization for Migration.
- Jaspers, Karl, 1952: Vom Ursprung und Ziel der Geschichte. München: Piper Verlag.
- Jonas, Hans, 1979: Das Prinzip Verantwortung. Versuch einer Ethik für die technologische Zivilisation. Frankfurt am Main: Suhrkamp.
- Kaufmann, Mareile, 2012: 2.2 Rights and Ethics. In: Čas, Johann; Kaufmann; Mareile, 2012 [Hrsg.]: Deliverable 2.6 DESSI System of Criteria, S. 20-27. Online abrufbar unter http://securitydecisions.org/download/9/system_of_criteria.pdf, zuletzt besucht am 23.3.2021.
- Kaufmann, Arthur, 1994: Negativer Utilitarismus. Ein Versuch über das bonum commune. München: Verlag der Bayerischen Akademie der Wissenschaften.
- Kern, Eva-Maria, 2020: Wie funktionieren Einsatzorganisationen? In: Kern, Eva-Maria; Richter, Gregor; Müller, Johannes C.; Voß, Fritz-Helge [Hrsg.]: Einsatzorganisationen. Erfolgreiches Handeln in Hochrisikosituationen. Wiesbaden: Springer Gabler, S. 13-39.
- Kern, Eva-Maria; Richter, Gregor; Müller, Johannes C.; Voß, Fritz-Helge [Hrsg.], 2020: Einsatzorganisationen. Erfolgreiches Handeln in Hochrisikosituationen. Wiesbaden: Springer Gabler.
- KIRAS 2015: Ausschreibungsleitfaden. Version 1.5. Wien: FFG und BMVIT. Online abrufbar unter https://www.ffg.at/sites/default/files/allgemeine_downloads/thematische%20programme/Sicherheit/kiras_2015_ausschreibungsleitfaden.pdf, zuletzt besucht am 28.7.2021.
- KIRAS 2016: Ausschreibungsleitfaden. Version 1.6. Wien: FFG und BMVIT. Online abrufbar unter https://www.ffg.at/sites/default/files/allgemeine_downloads/thematische%20programme/Energie/kiras_2016_ausschreibungsleitfaden_03.10.2016_.pdf, zuletzt besucht am 28.7.2021.
- Knoepffler, Nikolaus; Kunzmann, Peter; Pies, Ingo; Siegetsleitner, Anne [Hrsg.], 2006: Einführung in die Angewandte Ethik. München: Verlag Karl Alber.
- Koch, Heiner, 2014: Grundfragen. In: Quinn, Regina A. [Hrsg.], 2014: Sicherheitsethik. Wiesbaden: Springer VS, S. 123-181.
- Koops, Bert-Jaap, 2021: The concept of function Creep. Law, Innovation and Technology, Jg. 13, Heft 1, S. 29-56.
- Kreissl, Reinhard; Čas, Johann, 2012: Chapter 1. A brief theoretical introduction. In: Čas, Johann; Kaufmann; Mareile [Hrsg.]: Deliverable 2.6 DESSI System of Criteria, S. 8-12. Online abrufbar unter http://securitydecisions.org/download/9/system_of_criteria.pdf, zuletzt besucht am 23.3.2021.
- Krüger, Marco, 2018: Gesellschaftsethisches Gutachten zum Projekt „Multisensorell gestützte Erfassung von Straftätern in Menschenmengen bei komplexen Einsatzlagen“. Tübingen: IZEW.

Kunzmann, Peter, 2006: Technikethik. In: Knoepffler, Nikolaus; Kunzmann, Peter; Pies, Ingo; Siegetzleitner, Anne [Hrsg.]: Einführung in die Angewandte Ethik. München: Verlag Karl Alber, S. 249-266.

Lingner, Stephan, 2013: Rationale Technikfolgenbeurteilung. In: Simonis, Georg: Konzepte und Verfahren der Technikfolgenabschätzung. Wiesbaden: Springer VS, S. 75-90.

Löffler, Winfried; Runggaldier, Edmund [Hrsg.], 1999: Vielfalt und Konvergenz der Philosophie. Vorträge des V. Kongresses der Österreichischen Gesellschaft für Philosophie. Wien: Verlag Holder-Pichler-Tempsky.

Menlo 2012: The Menlo Report. Ethical Principles Guiding Information and Communication Technology Research. o.O.: U.S. Department of Homeland Security Science and Technology Directorate, Cyber Security Division. Online abrufbar unter https://www.caida.org/publications/papers/2012/menlo_report_actual_formatted/menlo_report_actual_formatted.pdf, zuletzt besucht am 25.3.2021.

Menlo 2013: The Menlo Report. Applying Ethical Principles to Information and Communication Technology Research. A Companion to the Menlo Report. o.O.: U.S. Department of Homeland Security Science and Technology Directorate, Cyber Security Division. Online abrufbar unter https://www.caida.org/publications/papers/2013/menlo_report_companion_actual_formatted/menlo_report_companion_actual_formatted.pdf, zuletzt besucht am 25.3.2021.

Mepham, Ben, 2000: A Framework for the Ethical Analysis of Novel Foods. The Ethical Matrix. Journal of Agricultural and Environmental Ethics, Jg. 12, S. 165-176.

Mepham, Ben; Kaiser, Matthias; Thorstensen, Erik; Tomkins, Sandy; Millar, Kate, 2006: Ethical Matrix. Manual. The Hague: LEI.

Mill, John Stuart, 2014 [1864]: Utilitarianism. Cambridge: Cambridge University Press.

Moula, Payam; Sandin, Per; 2015: Evaluating Ethical Tools. Metaphilosophy, Jg. 46, Heft 2, S. 263-279.

Mumford, Lewis, 1974: Mythos der Maschine. Kultur, Technik und Macht. Wien: Europaverlag.

Nagenborg, Michael, 2014: Ethik als Partnerin in der Technikgestaltung. In: Quinn, Regina A. [Hrsg.]: Sicherheitsethik. Wiesbaden: Springer Verlag, S. 241-252.

Naujoks, Daniel, 2020: Flucht und Migration: Trends, Faktoren, Dynamik. In: Beier, Christoph; Messner, Dirk; Preuß, Hans-Joachim [Hrsg.]: Globale Wanderungsbewegungen. Beiträge der internationalen Zusammenarbeit zum Umgang mit Flucht und Migration. Wiesbaden: Springer VS, S. 15-41.

OECD 1980: Organisation for Economic Co-operation and Development, 1980: Guidelines for the Transborder Flows of Personal Data. Online unter <https://www.oecd.org/digital/ieconomy/oecdguidelinesontheProtectionofPrivacyandTransborderFlowsofPersonalData.htm>, zuletzt besucht am 9.5.2021.

Palm, Elin; Hansson, Sven O., 2006: The case for ethical technology assessment (eTA). *Technology Forecasting & Social Change*, Jg. 73, Heft 5, S. 543-558.

Pécoud, Antoine; de Guchteneire, Paul [Hrsg.], 2007: *Migration without Borders. Essays on the Free Movement of People*. Oxford: Berghan Books.

Petermann, Thomas, 1999: Technikfolgen-Abschätzung. Konstituierung und Ausdifferenzierung eines Leitbilds. In: Bröchler, Stephan; Simonis, Georg; Sundermann, Karsten, 1999: *Handbuch Technikfolgenabschätzung*. Band 1. Berlin: Edition Sigma, S. 17-48.

Popper, Karl, 2003: *Die offene Gesellschaft und ihre Feinde*. Band I. Der Zauber Platons. Tübingen: Mohr Siebeck Verlag.

Quinn, Regina A. [Hrsg.], 2014: *Sicherheitsethik*. Wiesbaden: Springer Verlag.

Rauprich, Oliver, 2012: Principlism. In: Chadwick, Ruth [Hrsg.], 2012: *Encyclopedia of Applied Ethics*. London: Academic Press, S. 590-598.

Reijers, Wessel; Wright, David; Brey, Philip; Weber, Karsten; Rodrigues, Rowena; O'Sullivan, Declan; Gordijn, Bert, 2018: Methods for Practising Ethics in Research and Innovation: A Literature Review, Critical Analysis and Recommendations. *Science and Engineering Ethics*, Jg. 24, Nr. 5, S. 1437-1481.

Roth, Florian; Prior, Timothy, 2019: Utility of Virtual Operation Support Teams. An international survey. *Australian Journal of Emergency Management*, Jg. 34, Heft 2, S. 53-59.

Ropohl, Günter, 2016: *Ethik und Technikbewertung*. Frankfurt am Main: Suhrkamp.

Sasikumar, Karthika; Dudley, Danijela [Hrsg.], 2019: *Political and Military Sociology. The European Refugee Crisis. An Annual Review*. Volume 45. New York: Routledge.

Simonis, Georg, 2013a: Einführung. In: Simonis, Georg [Hrsg.]: *Konzepte und Verfahren der Technikfolgenabschätzung*. Wiesbaden: Springer VS, S. 11-19.

Simonis, Georg [Hrsg.], 2013b: *Konzepte und Verfahren der Technikfolgenabschätzung*. Wiesbaden: Springer VS.

Singer, Peter, 1994: *Praktische Ethik*. Neuausgabe. Stuttgart: Reclam Verlag.

Smart, Roderick N., 1958: Negative Utilitarianism. *Mind*, Jg. 67, Heft 268, S. 542-543.

Stahl, Bernd Carsten, 2011: IT for a better future: how to integrate ethics, politics and innovation. Journal of Information, Communication & Ethics in Society, Jg. 9, Heft 3, S. 140-156.

Swierstra, Tsjalling; Rip, Arie, 2007: Nano-ethics as NEST-ethics: Patterns of Moral Argumentation About New and Emerging Science and Technology. Nanoethics, Jg. 1, Heft 1, S. 3-20.

Tangen, Reidun, 2014: Balancing Ethics and Quality in Educational Research – the Ethical Matrix Method. Scandinavian Journal of Educational Research, Jg. 58, Heft 6, S. 678-694.

Taylor, Paul. W., 2011 [1987]: Respect for Nature. A Theory of Environmental Ethics. Princeton und Oxford: Princeton University Press.

Twitter AGB 2019: Online abrufbar unter <https://twitter.com/de/tos>, zuletzt besucht am 15.6.2019.

Ulbert, Cornelia; Werthes, Sascha, 2008a: Menschliche Sicherheit – Der Stein der Weisen für globale und regionale Verantwortung? Entwicklungslinien und Herausforderungen eines umstrittenen Konzepts. In: Ulbert, Cornelia; Werthes, Sascha [Hrsg.]: Menschliche Sicherheit. Globale Herausforderungen und regionale Perspektiven. Baden-Baden: Nomos Verlagsgesellschaft, S. 13-27.

Ulbert, Cornelia; Werthes, Sascha [Hrsg.], 2008b: Menschliche Sicherheit. Globale Herausforderungen und regionale Perspektiven. Baden-Baden: Nomos Verlagsgesellschaft.

UNESCO 2019: Immaterielles Kulturerbe und Tierethik: Online abrufbar unter <https://www.unesco.at/kultur/immaterielles-kulturerbe/artikel/article/immaterielles-kulturerbe-und-tierethik/>, zuletzt besucht am 26.7.2020.

UNHCR 2015: Online abrufbar unter <https://web.archive.org/web/20150812071308/http://www.unhcr.at/presse/pressemitteilungen/artikel/daa75e69c2cb7c397532a5e2b42a3e00/unhcr-appelliert-keine-weiteren-asylsuchenden-mehr-nach-traiski.html>, zuletzt besucht am 11.5.2021.

UNHCR (o.D.): Flüchtlinge und Integration. Begriffe einfach erklärt. Wien: UNHCR in Zusammenarbeit mit dem ÖIF. Online abrufbar unter https://www.integrationsfonds.at/fileadmin/content/AT/Fotos/Publicationen/Broschuere/Fluchtglossar_zweite_Auflage.pdf, zuletzt besucht am 15.5.2021.

VDI 1991: VDI-Richtlinie 3780. Düsseldorf.

VDI 2002: Ethische Grundsätze des Ingenieurberufs. Verein deutscher Ingenieure. Online abrufbar unter https://www.vdi.de/ueber-uns/presse/publikationen/details?tx_vdipublications_publication-details%5Bpublication%5D=269&cHash=4ae5da4f7372a8158cc891eaf38c9860, zuletzt besucht am 26.7.2020.

Von Silva-Tarouca Larsen, Beatrice, 2011: Setting the watch: privacy and the ethics of CCTV surveillance. Oxford and Portland, Oregon: Hart Publishing.

VOST Europe 2021: <https://vosteuropa.org/>, zuletzt besucht am 28.5.2021.

Vrgovic, Petar; Kapor-Stanulovic, Nila, 2019: Who Are the Refugees? A Demographic Analysis. In: Sasikumar, Karthika; Dudley, Danijela [Hrsg.]: Political and Military Sociology. The European Refugee Crisis. An Annual Review. Volume 45. New York: Routledge, S. 4-17.

Williams, Bernard, 1979: Kritik des Utilitarismus. Frankfurt am Main: Vittorio Klostermann GmbH.

WinFuture 2013: Türkei: Verhaftungswellen gegen Social Media-User: <https://winfuture.de/news,76381.html>, zuletzt besucht am 12.6.2021.

Wolkenstein, Andreas F.X., 2014: Akzeptanz und Akzeptabilität im Kontext der Angewandten Ethik. In: Quinn, Regina A. [Hrsg.]: Sicherheitsethik. Wiesbaden: Springer Verlag, S. 225-239.

Wright, David, 2011: A framework for the ethical impact assessment of information technology. Ethics and Information Technology, Jg. 13, Heft 3, S. 199-226.

Zecha, Gerhard, 1999: Kann ein negativer Utilitarismus die Schwächen des klassischen Utilitarismus überwinden? In: Löffler, Winfried; Runggaldier, Edmund [Hrsg.], 1999: Vielfalt und Konvergenz der Philosophie. Vorträge des V. Kongresses der Österreichischen Gesellschaft für Philosophie. Wien: Verlag Holder-Pichler-Tempsky, S. 421-426.

Zhang, Ping; Galletta, Dennis [Hrsg.], 2006: Human-computer interaction in management information systems: Foundations. Armonk, New York: M.E. Sharpe.

Zweck, Axel, 2013: Technikbewertung auf Basis der VDI-Richtlinie 3780. In: Simonis, Georg: Konzepte und Verfahren der Technikfolgenabschätzung. Wiesbaden: Springer VS, S. 145-160.

9 ANHANG

9.1 ABSTRACT IN DEUTSCHER SPRACHE

Die vorliegende Arbeit befasst sich aus technikethischer Perspektive mit dem deutsch-österreichischen Sicherheitsforschungsprojekt Human+, welches darauf abzielt, durch die Auswertung und Integration verschiedener Datentypen (z.B. Social-Media-Daten, In-Situ-Daten, Satellitenfernerkundungsdaten) Migrationsprognosen zu erstellen und ein Echtzeitlagebild für Behörden und Organisationen mit Sicherheitsaufgaben bereitzustellen. Dies soll im Falle von erneuten Fluchtereignissen wie in den Jahren 2015/2016 dazu dienen, die Informationslage für die beteiligten Organisationen zu verbessern und so die Gewährleistung humanitärer Sicherheit zu unterstützen.

Im Rahmen der vorliegenden Arbeit werden zwei Ziele verfolgt: ein inhaltliches und ein methodisches Ziel. Das inhaltliche Ziel ist die technikethische Bewertung von Human+. Das methodische Ziel ist die Erarbeitung eines Bewertungsframeworks hierzu, welches einerseits auf Human+ Anwendung finden können soll, andererseits aber generisch genug ist, um auch für andere Sicherheitsforschungsprojekte adaptierbar zu sein. Hierzu werden nach einer Standortbestimmung der vorliegenden Arbeit zwischen Technikethik und Technikfolgenabschätzung durch Rückgriff auf ein Principlism-Fundament und verschiedene etablierte Ansätze der Technikethik, sowie auf den Negativen Utilitarismus, Grundlagen erarbeitet, und ein Bewertungsframework daraus erstellt. Dieses wird anschließend exemplarisch für die technikethische Analyse von Human+ herangezogen.

9.2 ABSTRACT IN ENGLISH

This master thesis examines the German-Austrian security research project Human+. The goal of Human+ is the analysis and integration of a broad spectrum of different data (e.g. social media, in-situ, satellite remote sensing), in order to predict migration and to provide a real time overview of the situation for authorities and organizations responsible for security. Both (predictions and overview of the situation) are intended to improve situational awareness for involved organizations in cases similar to the refugee crisis in 2015/2016 in order to support the provision of human security.

Thus, this thesis has two objectives: the ethical evaluation of Human+, and as a prerequisite for this, the development of an evaluation framework that is both applicable to Human+ and sufficiently generic to be adapted for other security research projects as well. Essential constituents of the evaluation framework will be found in Principlism, in distinct approaches to technology ethics, and also in negative utilitarianism. The framework will then be applied for the ethical evaluation of Human+.