

# Ethik im Zeitalter der Künstlichen Intelligenz

## Eine Überblick internationaler Vorgehensweisen und Handlungsempfehlungen

Max Haarich, 18. Januar 2019

*"[...] wild speculation about the future of AI will be replaced by hard decisions about ethics and democracy; 2019 will be the year that AI grows up." - Wired Magazine*

**This is the draft of an internal  
report financed by applied AI  
initiative ([www.appliedai.de](http://www.appliedai.de)).  
NOT READY FOR  
PUBLICATION.**

## Inhalt

1	Einleitung.....	3
2	Ethik in er KI.....	4
2.1	Warum jetzt?.....	4
2.2	Was ist Ethik? .....	6
2.3	Wie kommt Ethik in die KI? .....	6
3	Ausgewählte Ethik-Vorgehensweisen in der KI.....	7
3.1	Staatliche Akteure .....	7
3.1.1	Die EU und ihre Staaten .....	7
3.1.2	Die "Weltführer" im KI-Bereich .....	13
3.2	Nicht-staatlicher Akteure .....	14
3.3	"Big Five" Technologie-Konzerne .....	16
3.4	Weitere bekannte KI-Ethik-Erklärungen.....	18
4	Handlungsempfehlungen .....	19
5	Anhang.....	22
6	Auswahl ethischer Prinzipien .....	22
6.1	Staatliche Akteure .....	22
6.1.1	EU-Staaten.....	22
6.1.2	EU .....	22
6.2	Nicht-staatliche Akteure.....	23
6.3	"Big Five" Technologie-Konzerne .....	24
6.4	Weitere Bekannte KI-Ethik-Erklärungen .....	24

# 1 Einleitung

Dieser Artikel soll jedem, der sich für die ethische Debatte um Künstlicher Intelligenz (KI) interessiert, einen schnellen Einstieg ermöglichen und weitergehende Recherchen erleichtern. Der Artikel gibt einen Überblick über internationale Ethik-Vorgehensweisen\* im Bereich KI: welche Akteure dominieren die öffentliche Debatte? Mit welchen Maßnahmen oder Prinzipien versuchen sie ethische Maßstäbe zu entwickeln und umzusetzen?

Recherchiert wurden hierfür die ethischen Vorgehensweisen

- der EU und ihrer Staaten<sup>1</sup>,
- der "Weltführer" im KI-Bereich USA und China,
- der nicht-staatlichen Akteure *Future of Life Institute*, *FAT/ML*, *Ethics in Artificial Intelligence*, *Open AI*, *Partnership on AI*, *AI4ALL*, *Ethics and Governance in Artificial Intelligence*, *AI Ethics Lab*, *MIT Stephen A. Schwarzman College of Computing* und
- der "Big Five" Technologie-Konzerne *Google*, *Amazon*, *Facebook*, *Apple* und *Microsoft*.
- Die bekannten Ethik-Erklärungen *Ethically Aligned Design*, *Asilomar Principles*, *Toronto Declaration* und *Montreal Declaration* werden ebenfalls kurz erläutert.

Abschließend wurden fünf Handlungsempfehlungen formuliert, die sich an alle Gestalter der digitalen Transformation richten. Die Empfehlungen unterstützen einen flexiblen und iterativen Lösungsprozess. Als Anhang sind alle im Dokument genannten Sets von Prinzipien noch einmal gesammelt.

**\*Hinweis zum Vorgehen:** Für diesen Artikel wurden ausschließlich online verfügbare Texte und Textpassagen untersucht. Bei den Nationalstaaten wurden nur die Dokumente untersucht, die von ihren Verfassern entweder als dediziertes Ethikkonzept bezeichnet werden, oder explizit ethische Fragen innerhalb einer offiziellen KI-Strategie thematisierten.<sup>2</sup> Entscheidend war, dass die Texte und Textpassagen eindeutig als offizieller Beitrag der Staaten zum aktuellen KI-Diskurs einzuordnen waren.<sup>3</sup> Zur Vereinfachung wurden alle einschlägigen Texte und Textpassagen desselben Verfassers als dessen "ethische Vorgehensweisen" bzw. "Ethik-Vorgehensweisen" zusammengefasst.

---

<sup>1</sup> Nur bei neun EU-Staaten konnten explizite Ethik-Vorgehensweisen gefunden werden: Dänemark, Italien, Frankreich, Schweden, Finnland, Vereinigtes Königreich, Niederlande, Österreich und Deutschland.

<sup>2</sup> Eine Übersicht internationaler KI-Strategien findet man z. B. beim *Future of Life Institute* (<https://futureoflife.org/national-international-ai-strategies/>), bei medium.com (<https://medium.com/politics-ai/an-overview-of-national-ai-strategies-2a70ec6edfd>) oder bei Oxford Insights (<https://www.oxfordinsights.com/insights/germanyai>). Die EU gibt eine allgemeine Übersicht über die europäische KI-Landschaft (<https://ec.europa.eu/digital-single-market/en/news/european-artificial-intelligence-landscape>).

<sup>3</sup> Es wurden auch solche Textpassagen einbezogen, bei denen statt von "Künstlicher Intelligenz" von ähnlichen Begriffen wie "Machine Learning" oder allgemeiner über "Algorithmen" gesprochen wurde. Man kann davon ausgehen, dass die mit diesen Begriffen assoziierten ethischen Themen so ähnlich sind, dass sie zum selben Diskurs beitragen.

## 2 Ethik in er KI

### 2.1 Warum jetzt?

Besonders in den letzten zwei Jahren haben so viele Unternehmen, Verbände und Staaten ethische Vorgehensweisen für KI erarbeitet, dass man von einem Hype<sup>4</sup> oder sogar einem internationalen Wettbewerb<sup>5</sup> sprechen möchte. Die möglichen Maßnahmen reichen von der Ernennung einer einzelnen für Ethik verantwortlichen Person<sup>6</sup> über die Entwicklung bias-identifizierender Software<sup>7</sup> bis hin zu Online-Konsultationen<sup>8</sup> zur Ermittlung der ethischen Vorstellungen potenziell aller Stakeholder eines gesamten Kontinents.

Wie kommt es, dass gerade jetzt so viel Akteure Maßnahmen zur Vermeidung ethischer Probleme vornehmen? Ethische Probleme mit KI sind streng genommen kein neues Thema. Bereits Anfang des 19. Jahrhunderts am Vorabend der industriellen Revolution erschienen Science-Fiction-Werke, die ethische Fragen im Umgang mit künstlich intelligenten Wesen stellen wie z. B. *Der Sandmann* (1817) von E.T.A. Hoffmann und *Frankenstein oder Der moderne Prometheus* (1818) von Mary Shelly.

Gerade in den letzten vier Jahren erhitzt sich die öffentliche Debatte an einer Vielzahl ethisch bedenklicher Nachrichtenmeldungen aus der KI-Welt. Die nachfolgende Grafik fasst die Entwicklung grob zusammen.

### Talk of AI and ethics is on the rise

Quarterly news mentions of ("AI OR artificial intelligence") AND "ethics" 2014 – Q3 2018



<sup>4</sup> <https://www.cs.ox.ac.uk/efai/developing-codes-of-ethics-for-ai/hype-in-ai/>

<sup>5</sup> Das Vereinigte Königreich formuliert z. B. in seiner KI-Strategie den Anspruch "moral and ethical leader on the global stage" zu werden.

<sup>6</sup> Salesforce stellte im Dezember 2018 Paula Goldman als "Chief Ethical and Humane Use Office" ein: <https://www.businessinsider.de/salesforce-hires-paula-goldman-as-chief-ethical-and-humane-use-officer-2018-12?r=US&IR=T>

<sup>7</sup> Facebook gab im Mai 2018 die Entwicklung der Software "Fairness Flow" bekannt, mit der möglicher Bias in ihrem System für Stellenempfehlungen identifiziert werden könne.

<sup>8</sup> Die Europäische Kommission bat im Januar 2019 online um Feedback zum KI-Ethik-Entwurf der High-Level-Expert-Group: <https://ec.europa.eu/futurium/en/ai-stakeholder-consultation/stakeholders-consultation-draft-ai-ethics-guidelines>

1. (06/2014) China kündigt ein nationales Social Credit System an zur Bewertung von Individuen und Regierungsinstitutionen.<sup>9</sup>
2. (06/2015) Die Google-App "Photos" identifiziert afro-amerikanische Menschen als Gorillas aufgrund unzureichendem Trainings der KI.<sup>10</sup>
3. (12/2015) Cambridge Analytica nutzt 50 Mio. Datensätze von Facebook-Nutzern, um deren politische Meinung in zahlreichen Wahlkämpfen durch gezielte Facebook-Beiträge zu beeinflussen.<sup>11</sup> Im März 2018 wird die Beteiligung von Cambridge Analytica im US-Präsidentenwahlkampf 2016 bekannt.<sup>12</sup>
4. (05/2016) In Florida stirbt der erste Mensch in einem autonom fahrenden PKW. An einer Kreuzung konnte sein Tesla einen weißen Truck nicht vom freien Himmel unterscheiden und beschleunigte noch kurz vor dem Aufprall.<sup>13</sup>
5. (12/2017) Vice Magazine berichtet über ein Deepfake-Video, bei dem das Gesicht einer Prominenten nachträglich über das Gesicht einer Pornodarstellerin gelegt wurde.<sup>14</sup> Deepfakes mit gefälschten Ansprachen von Obama<sup>15</sup> und Trump<sup>16</sup> folgen.
6. (03/2018) Die UN wirft Facebook vor durch die Verbreitung von Hate Speech die gewalttätigen Ausschreitungen in Myanmar begünstigt zu haben.
7. (05/2018) Google präsentiert die Software "Duplex", die so gut einen Menschlichen Gesprächspartner imitiert, dass sie mit einem Friseur-Salon telefonisch einen Termin ausmachen kann.<sup>17</sup>
8. (05/2018) Der Künstler Alexander Reben steuert eine Pistole über einen Smart Home Assistant. Er löst damit eine Debatte über die leichte Verfügbarkeit von Lethal Autonomous Weapon Systems (LAWS) aus.<sup>18</sup>
9. (06/2018) Amazon Anteilseigner bitten Jeff Bezos in einem offenen Brief seine "Rekognition" Software nicht der Straf-Verfolgung zur Verfügung zu stellen.<sup>19</sup>
10. (11/2018) Über 400 Google Mitarbeiter protestieren in einem offenen Brief gegen die Entwicklung von "Dragonfly", einer zensierten Suchmaschine, die staatliche Überwachung ermöglicht.<sup>20</sup>

<sup>9</sup> [https://web.archive.org/web/20150503004438/http://news.xinhuanet.com/english/china/2014-06/27/c\\_133443776.htm](https://web.archive.org/web/20150503004438/http://news.xinhuanet.com/english/china/2014-06/27/c_133443776.htm)

<sup>10</sup> <https://www.theverge.com/2015/7/1/8880363/google-apologizes-photos-app-tags-two-black-people-gorillas>

<sup>11</sup> <https://www.theguardian.com/us-news/2015/dec/11/senator-ted-cruz-president-campaign-facebook-user-data>

<sup>12</sup> <https://www.theguardian.com/news/2018/mar/17/cambridge-analytica-facebook-influence-us-election>

<sup>13</sup> <https://www.theguardian.com/technology/2016/jun/30/tesla-autopilot-death-self-driving-car-elon-musk>

<sup>14</sup> [https://motherboard.vice.com/en\\_us/article/gydydm/gal-gadot-fake-ai-porn](https://motherboard.vice.com/en_us/article/gydydm/gal-gadot-fake-ai-porn)

<sup>15</sup> <https://www.theverge.com/tldr/2018/4/17/17247334/ai-fake-news-video-barack-obama-jordan-peelee-buzzfeed>

<sup>16</sup> <https://www.theguardian.com/technology/2018/nov/12/deep-fakes-fake-news-truth>

<sup>17</sup> <https://ai.googleblog.com/2018/05/duplex-ai-system-for-natural-conversation.html>

<sup>18</sup> <https://www.engadget.com/2018/05/30/google-assistant-fired-a-gun-we-need-to-talk/>

<sup>19</sup> <https://venturebeat.com/2018/06/18/shareholders-protest-selling-amazons-facial-recognition-software-to-law-enforcement/>

<sup>20</sup> <https://medium.com/@googlersagainstdragonfly/we-are-google-employees-google-must-drop-dragonfly-4c8a30c5e5eb>

## 2.2 Was ist Ethik?

Die Diskussion über die gesellschaftlichen Konsequenzen von Künstlicher Intelligenz (KI) ist durch zwei extreme Positionen gekennzeichnet: Schwarzmaler sehen die komplette Vernichtung der Menschheit durch den Terminator voraus, Enthusiasten erwarten hingegen unseren Einzug in ein volldigitales Smart Paradise. Beides wäre ein Ende der Welt, wie wir sie kennen. Für welche Option soll man sich entscheiden?

Ethische Diskussionen sind notwendig, um gesamtgesellschaftlich Alternativen zu entwickeln, die für alle erstrebenswert sind. Technologie zeigt nur, was wir machen könnten. Erst die Ethik sagt, was wir machen sollten. Je nachdem, welcher Form von Ethik man folgt, kann diese Antwort aber recht unterschiedliche ausfallen.

### Tugendethik

Die Tugendethik geht auf Aristoteles Nikomachische Ethik zurück und ist die älteste wissenschaftliche Ethik Europas. Nach der Tugendethik ist die richtige Handlung diejenige, die auch ein absolut tugendhafter Mensch mit besten Motiven und Gedanken vornehmen würde. Durch den regelmäßigen Vergleich seines Verhaltens mit dem Verhalten besonders tugendhafter Menschen könne man sich zunehmend einem universellen Ethik-Ideal annähern.

### Konsequenzialismus

Beim Konsequenzialismus spielt weder die konkrete Handlung noch die zugrundeliegende Motivation eine Rolle. Es kommt einzig und allein auf die Konsequenz an. Werden Konsequenzen angestrebt, die den Gesamtnutzen für alle erhöhen, spricht man vom *Utilitarismus*. Stehen positive Konsequenzen für die eigene Person im Vordergrund, geht man vom *ethischen Egoismus* aus.

### Deontologische Ethik

Bei der *deontologischen Ethik* sind die konkrete Handlung als auch deren Ergebnis nachgeordnet. Entscheidend ist nur, ob die Handlung an einer höheren Regel ausgerichtet war und ob diese korrekt befolgt wurde.

## 2.3 Wie kommt Ethik in die KI?

Für die folgende Untersuchung ist es hilfreich folgende drei Wege zu unterscheiden, auf denen man Ethik in eine KI bringen kann: Ethics by, in and for Design.

- *Ethics by Design*: die KI erhält die Fähigkeit ethische Entscheidungen zu treffen *aufgrund von* technischen oder algorithmischen Funktionen. Nach Meinung der meisten Experten gibt es das bisher nur in der Science Fiction.
- *Ethics in Design*: *im* Gestaltungsprozess benutzt der Entwickler Regeln oder Methoden, die eine ethisches Verhalten der KI sicherstellen. Hierzu zählen z. B. die Vermeidung von Bias in den Trainingsdaten.
- *Ethics for Design*: *für* Entwickler werden Regeln, Standards, Maßnahmen, etc. erarbeitet, um den Entwickler zu motivieren, eine KI ethische zu gestalten.

## 3 Ausgewählte Ethik-Vorgehensweisen in der KI

In diesem Abschnitt werden staatliche Akteure, nicht-staatliche Akteure und Tech-Konzerne präsentiert, die Ethik-Vorgehensweisen aufgestellt haben, um sich selbst bzw. ihre Mitarbeiter zu ethischem Verhalten zu motivieren. Sie sind in chronologischer Reihenfolge gemäß dem Veröffentlichungsdatum ihrer ethischen Vorgehensweise sortiert. Am Ende des Abschnitts folgen vier weitere ethische Leitsätze, die mehr oder weniger an den gesamten KI-Bereich adressiert wurden.

### 3.1 Staatliche Akteure<sup>21</sup>

#### 3.1.1 Die EU und ihre Staaten

##### Dänemark

Aktuelles Dokument mit Aussagen zu Ethik-Vorgehensweisen: *Strategy for Denmark 's Digital Growth*<sup>22</sup>

Veröffentlicht: Januar 2018

Dänemark beschreibt sich selbst als grundsätzlich gut darin, verantwortlich zu arbeiten. Entsprechend fallen die Aussagen zu ethischen Vorgehensweisen in der KI recht knapp aus. Die explizite Passage umfasst wenige Zeilen und nennt allgemeine Maßnahmen, wie die Etablierung von Daten-Ethik - möglicherweise inklusive eines Daten-Ethik-Codes - und Diskussionen zum Thema auf international relevanten Foren. Bis 2020 seien dafür Gesamtausgaben in Höhe von Fünf Mio. DKK (≈ 670.000 EUR) geplant.

##### Italien

Aktuelles Dokument mit Aussagen zu Ethik-Vorgehensweisen: *Artificial Intelligence. At the Service for Citizens*<sup>23</sup>

Veröffentlicht: März 2018

Italien besitzt unter den EU-Staaten eines der umfangreichsten Konzepte für Ethik in der KI und will diese gezielt zur Ausrichtung aller KI-Strategien einsetzen: Italien möchten mit bewussten Überlegungen darüber starten, wie wir uns die Welt zukünftig wünschen. Dabei berufen sie sich auf die Tugendethik von Aristoteles, aber betonen gleichzeitig, dass auch mit Anhängern anderer Ethiken wie der Kantischen Ethik, der Konfuzianischen Ethik, Shinto oder Ubuntu eine Einigung auf die Verbesserung menschlichen Wohlergehens als oberstes Ziel möglich sei. Wichtig sei die Bekräftigung des "Anthropozentrischen Prinzips", wonach KI immer vorrangig dem Bürger dienen solle. Grundsätzlich scheint die ethische Anwendung von KI in der öffentlichen Verwaltung ein zentraler Referenzpunkt der italienischen Überlegungen.

---

<sup>21</sup> Staatliche Vorgehensweisen sind nicht so einfach zu vergleichen, da 1. die Ausgangslagen und Vorgehensweisen der Akteure inhaltlich sehr unterschiedlich sind, 2. bei inhaltlich ähnlichen Vorgehensweisen oft unterschiedliche Begrifflichkeiten verwendet werden und 3. die Begrifflichkeiten auch innerhalb eines Ethikkonzeptes nicht immer stringent verwendet werden. Vermutlich ist dies darauf zurückzuführen, dass viele Strategien und Konzepte im Rahmen von groß angelegten Konsultationen entstanden, bei denen oft hunderte Menschen ihre unterschiedlichen Blickwinkel und Begriffsverständnisse einfließen ließen.

<sup>22</sup> [https://eng.em.dk/media/10566/digital-growth-strategy-report\\_uk\\_web-2.pdf](https://eng.em.dk/media/10566/digital-growth-strategy-report_uk_web-2.pdf)

<sup>23</sup> <https://ia.italia.it/en/>

Ethik in der KI ist in Italiens KI-Strategie Herausforderungen Nr. 1. Sie fordern die Formulierung allgemeiner Prinzipien der Gerechtigkeit als Basis für Freiheit sowie individuelles und kollektives Recht und schlagen vier "Elemente" für die öffentliche Debatte und die wissenschaftlichen Analyse vor. Eine konkret beschriebene Maßnahme ist der Aufbau eines *Trans-disciplinary Centre on AI*, mit den folgenden Zielen: Förderung einer Debatte über die Evolution der Ethik, Unterstützung kritischer Überlegungen zu aufkommenden ethischen Belangen und die erleichterte Einbindung von Experten und Bürgern bei der Ableitung von Regulierungen, Standards und technischen Lösung ausgehend von technischen und sozialen Aspekten.

## Frankreich

Aktuelles Dokument mit Aussagen zu Ethik-Vorgehensweisen: *AI for Humanity*<sup>24</sup>

Veröffentlicht: März 2018

Frankreichs KI-Strategie basiert auf dem Report<sup>25</sup> des Mathematikers und Politikers Cédric Villani. Im Abschnitt "Ethical Considerations of AI Recent" schlägt er z. B. eine Art Sondereinsatztruppe zertifizierter Experten vor, die bei verfassungsrechtlichen Fragen mit der Auditierung und dem Test von Algorithmen beauftragt werden. Für das Training von KI-Experten empfiehlt er *Ethics by Design* als obligatorischen Bestandteil, um von Anfang an für ethische Belange zu sensibilisieren. Weiterhin empfiehlt er ein "digital technology and AI ethics committee", um eine transparente öffentliche Debatte zu leiten. Bei dieser Debatte solle trotz aller dringlichen Fragen eine langfristige Perspektive entstehen.

Diesen und weiteren Empfehlungen folgt Präsident Macron. In seiner offiziellen Strategie findet man die Idee einer internationalen Expertengruppe ähnlich dem *Intergovernmental Panel on Climate Change (IPCC)*<sup>26</sup>. Die Gruppe soll unabhängige Erkenntnisse zu Transparenz und Fairness erarbeiten und in Bildungsprogramme einfließen lassen. Desweiteren sei Forschung zu besser verständlicher KI nötig: besser erklärbare Modell, besser interpretierbare Interfaces und insgesamt ein besseres Verständnis der zugrundeliegenden Mechanismen. KI-Entwicklern möchte er Ethik-Trainings und ein "discrimination impact assessment" für ihre Algorithmen bieten. Für den öffentlichen Sektor plant er die klare Regulierung von Vorhersage-Algorithmen in der Strafverfolgung und eine intensive internationale Debatte jeglicher Entwicklung von LAWS bis hin zu einem Observatorium zur Eindämmung ihrer Verbreitung. Als weitere Bedingungen ethischer KI benennt er Diversity und Inklusion. Konkret möchte er die Frauenquote in digitalen Ingenieurs-Studiengängen schon bis 2020 auf 40% heben. Begleitend möchte er Ausbilder durch Richtlinien und Trainings verstärkt auf die Bedeutung von Diversity hinweisen. Die Regierung solle den Einsatz von KI in Programmen für soziale Innovationen fördern, damit auch Arbeiter des Sozialbereichs von der Technologie profitieren.

## Schweden

---

<sup>24</sup> <https://www.aiforhumanity.fr/en/>

<sup>25</sup> [https://www.aiforhumanity.fr/pdfs/MissionVillani\\_Report\\_ENG-VF.pdf](https://www.aiforhumanity.fr/pdfs/MissionVillani_Report_ENG-VF.pdf)

<sup>26</sup> <https://www.ipcc.ch>



Aktuelles Dokument mit Aussagen zu Ethik-Vorgehensweisen: *National approach to artificial intelligence*<sup>27</sup>

Veröffentlicht: Mai 2018

Schweden nennt "Sustainable" AI als übergreifendes Motto und betont, dass Ethik, Schutz und Sicherheit keine nachgeschalteten Überlegungen sein dürfen, sondern integraler Bestandteil früher Phasen der KI-Entwicklung. Desweiteren wird der Aufbau interdisziplinären KI-Wissens für Ingenieure und Entscheider gewünscht und es werden transparente und verständliche Algorithmen gefordert für die Entwicklung von Regeln, Standards, Normen und ethischen Prinzipien. Die ethischen Bestrebungen müssten jedoch in Balance bleiben mit dem dringenden Bedarf an Daten, um eine die gesellschaftlichen Vorzüge von KI realisieren zu können.

## **Finnland**

Aktuelles Dokument mit Aussagen zu Ethik-Vorgehensweisen: *Work in the age of artificial intelligence Publications of the Ministry of Economic Affairs and Employment: Four perspectives on the economy, employment, skills and ethics*<sup>28</sup>

Veröffentlicht: Juni 2018

Finnland setzt sich das "ambitious goal" Technologie für wertvolle soziale Ziele einzusetzen und fokussiert dabei stark auf die Arbeitsgestaltung. Eine gesellschaftliche Vision sei eine Bedingung, damit die Regierung die Entwicklung und Anwendung von KI sinnvoll beeinflussen könne. Sie bemängeln, dass zu diesem Zeitpunkt noch in keinem der zahlreichen internationalen Reports eine Strategie oder ein Handlungsplan zur Erreichung der häufig gewünschten *good artificial intelligence societies* zu finden sei. Für eine solche Gesellschaft schlagen sie die Werte "Transparency", "Responsibility" und "Extensive societal benefits" vor. Sie betonen dabei, dass die rechtliche und moralische Verantwortung algorithmischer Anwendungen eindeutig beim Menschen bleiben müsse.

Sofern die zukünftigen Entwicklungen zu Arbeitsplatzverlusten führten, sei eine Unterstützung der betroffenen Menschen geboten, denn schließlich hätten sie den Fortschritt mit ihren Steuern bezahlt. Die Gesellschaft solle gleichermaßen von den Vorzügen der KI profitieren. Dafür sind Mechanismen zu Aufteilung dieser Vorzüge wie auch Einkommenstransfers genannt. Die Entwicklungen in der Arbeitswelt müssten auf der Ebene der Gesamtgesellschaft aber auch der individuellen Arbeitsorganisationen diskutiert werden.

Bei einzelnen Firmen würde der ethische Umgang mit KI zunehmend den Markenwert beeinflussen. Die großen Firmen hätten dies bereits verstanden, dennoch sei weitere Forschung aus Perspektive der Technologie, des Business Managements und der Politikwissenschaft nötig. Dies solle in internationalen Kooperationen ausgeweitet werden, wodurch Finnland auch eine Pionier-Funktion bei der Umsetzung der KI-Strategien der EU einnehmen möchte.

---

<sup>27</sup> <https://www.regeringen.se/4aa638/contentassets/a6488cceb6f418e9ada18bae40bb71f/national-approach-to-artificial-intelligence.pdf> Dieses Dokument basiert auf dem Report der schwedischen Agentur zur Innovationsförderung *Vinnova* ([https://www.vinnova.se/contentassets/29cd313d690e4be3a8d861ad05a4ee48/vr\\_18\\_09.pdf](https://www.vinnova.se/contentassets/29cd313d690e4be3a8d861ad05a4ee48/vr_18_09.pdf)) Ein finaler Report ist für Mai 2019 angekündigt.

<sup>28</sup> [http://julkaisut.valtioneuvosto.fi/bitstream/handle/10024/160980/TEMjul\\_21\\_2018\\_Work\\_in\\_the\\_age.pdf](http://julkaisut.valtioneuvosto.fi/bitstream/handle/10024/160980/TEMjul_21_2018_Work_in_the_age.pdf) Ein finaler Report ist für April 2019 angekündigt.

Das Gelingen der Strategie setze eine starke Bildung der Bevölkerung zum Thema AI voraus. Hier hat Finnland mit einem einzigartigen Programm zur Steigerung von Verständnis aber auch Diversität und Inklusion im AI-Bereich vorgelegt: Mit einer AI-Challenge unterstützt Finnland seit Mai 2017 insgesamt ca. 55.000 Bürger (1% der Gesellschaft), selber KI-Lösungen für ihre persönlichen Aufgaben zu entwickeln.<sup>29</sup>

Weitere geplante Maßnahmen sind der Aufbau einer "parliamentary monitoring group" zur Beobachtung des technologischen Fortschritts, die Formulierung einer auf das Gemeinwohl gerichteten Wertebasis (mit einem Exzellenzcenter speziell für den B2B-Bereich) sowie die Regulierung von Online-Plattformen nach dem Beispiel der Europäischen Datenschutzgrundverordnung (DSGVO)<sup>30</sup>. Insgesamt solle man trotz der kaum vorhersehbaren Entwicklungen im KI-Bereich weiterhin experimentieren und Erfahrungen sammeln. Dabei könne man sich darauf verlassen, dass die staatlichen Institutionen in kritischen Situationen unterstützend beistünden.

### **Vereinigtes Königreich**

Aktuelles Dokument mit Aussagen zu Ethik-Vorgehensweisen: *Data Ethics Framework*<sup>31</sup>

Veröffentlicht: Juni 2018

Das vereinigte Königreich formuliert unmissverständlich seinen Anspruch "world leader in ethical AI"<sup>32</sup> zu werden. Der Schwerpunkt liegt auf dem ethischen Umgang mit vertraulichen Daten, um deren Verfügbarmachung und Austausch zu erleichtern. Zentrale Instrumente zur Sicherung ethischer KI sind unter anderem *Data Trusts*<sup>33</sup> und ein *Centre for Data Ethics and Innovation (CDEI)*<sup>34</sup>. Letzteres scheint jetzt schon die offizielle Informationsstelle für die konkrete ethische Vorgehensweise des Vereinigten Königreichs zu sein. Es soll langfristig als Schnittstelle zwischen Regierung, Wirtschaft und Zivilgesellschaft fungieren. Konkret will es die Regierung bei der Gesetzgebung unterstützen, gemeinsam mit Forschungseinrichtungen und Think Tanks nationale

---

<sup>29</sup> Hier ein Bericht über einen teilnehmenden Zahnarzt: <https://www.politico.eu/article/finland-one-percent-ai-artificial-intelligence-courses-learning-training/>

<sup>30</sup> <https://eur-lex.europa.eu/legal-content/DE/TXT/?uri=celex%3A32016R0679>

<sup>31</sup> [https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment\\_data/file/737137/Data\\_Ethics\\_Framework.pdf](https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/737137/Data_Ethics_Framework.pdf) Konkretere Aussagen zur ethischen Vorgehensweise finden sich besonders in folgenden Dokumenten: *House of Lords Report AI in the UK: ready, willing and able?* (<https://publications.parliament.uk/pa/ld201719/ldselect/ldai/100/100.pdf>), die dazu verfasste Antwort der Regierung (<https://www.parliament.uk/documents/lords-committees/Artificial-Intelligence/AI-Government-Response2.pdf>), die Konsultation zum *Centre for Data Ethics and Innovation* (<https://www.gov.uk/government/consultations/consultation-on-the-centre-for-data-ethics-and-innovation/centre-for-data-ethics-and-innovation-consultation>), die dazu verfasste Antwort der Regierung ([https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment\\_data/file/757509/Centre\\_for\\_Data\\_Ethics\\_and\\_Innovation\\_-\\_Government\\_Response\\_to\\_Consultation.pdf](https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/757509/Centre_for_Data_Ethics_and_Innovation_-_Government_Response_to_Consultation.pdf)) und dem *AI Sector Deal* (<https://www.gov.uk/government/publications/artificial-intelligence-sector-deal/ai-sector-deal>).

<sup>32</sup> <https://slate.com/technology/2018/08/the-u-k-wants-to-be-the-world-leader-in-ethical-a-i.html> In diesem Artikel kann nicht bewertet werden, in wie weit derart kompetitiv anmutende Ambitionen möglicherweise einen Konflikt mit einer aristotelischen Tugend-Ethik auslösen können, die das gesunde Mittelmaß als Maßstab setzt.

<sup>33</sup> *Data Trusts* sollen einen geschützten und fairen Austausch von Daten ermöglichen. Es herrscht noch Unklarheit, wie das genau geschehen soll (<https://theodi.org/article/what-is-a-data-trust/>)

<sup>34</sup> <https://www.gov.uk/government/groups/centre-for-data-ethics-and-innovation-cdei> Die Board Members wurden bereits benannt und die Arbeit soll in Kürze beginnen.

Herausforderungen und globale Chancen identifizieren sowie führenden Unternehmen eine Wissensbasis und eine Austauschplattform für Best Practices bieten.

Die Aussagen des CDEI zu seinen grundsätzlichen Vorstellungen ethischer KI stehen weitgehend im Einklang mit dem allgemeinen Diskurs: Diversity unter den Entwicklern soll erhöht und Fairness, Transparenz, Haftung sowie Verantwortlichkeit sichergestellt werden. Vergleichsweise stärkere Aufmerksamkeit erhalten folgende Handlungsfelder: Targeting der Bürger durch Werbetreibende, Frameworks zum Austausch von Daten sowie Intellektuelles Kapital und Eigentumsrechte bei durch KI generierten Innovationen.

## **Niederlande**

Aktuelles Dokument mit Aussagen zu Ethik-Vorgehensweisen: *AI voor Nederland. Vergroten Versnellen en Verbinden*<sup>35</sup>

Veröffentlicht: Oktober 2018

Die public-private partnership *AINED*<sup>36</sup> hat gemeinsam mit der *Boston Consulting Group (BCG)* einen ersten Report für die Niederlande vorgelegt. Dieser enthält lediglich den Hinweis kontinuierlich Erfahrungen mit KI und seinen sozialen Folgen zu sammeln und daraus dann Regularien für die Technologie zu formulieren. Beobachter sehen für die Niederlande dringenden Handlungsbedarf, um sich von der ehemals guten Wettbewerbsposition nicht weiter zu entfernen.<sup>37</sup> *AINED* und *BCG* haben eine nationale Strategie in den nächsten Monaten angekündigt.

## **Österreich**

Aktuelles Dokument mit Aussagen zu Ethik-Vorgehensweisen: *Die Zukunft Österreichs mit Robotik und Künstlicher Intelligenz positiv gestalten. White Paper des Österreichischen Rats für Robotik und Künstliche Intelligenz*<sup>38</sup>

Veröffentlicht: November 2018

Österreich ist das einzige untersuchte Land, das in seiner ethischen Vorgehensweis zur KI gleichwertig die Robotik adressiert. Interessant ist die Aussage, dass der moralische Status von Robotern und KI "derzeit unklar" sei. Dem würden vermutlich die meisten anderen Autoren ethischer Vorgehensweisen aktuell widersprechen und jeglichen moralischen Status ausschließen. Österreich ist auch der einzige EU-Staat, der LAWS kategorisch verbieten möchte.<sup>39</sup> Abgesehen davon scheinen die weiteren Aussagen im Einklang mit dem allgemeinen Diskurs.

---

<sup>35</sup> <https://order.perssupport.nl/file/pressrelease/a253ded8-1863-4d07-b79d-72639baefc51/6be5dcdc-b9d8-4675-bd80-9425826c840c/AlvNL20181102final.pdf>

<sup>36</sup> Deren Website [ained.nl](http://ained.nl) ist leider noch nicht abrufbar.

<sup>37</sup> <https://www.nwo.nl/en/news-and-events/news/2018/11/companies-and-researchers-publish-roadmap-for-implementing-a-national-ai-strategy.html>

<sup>38</sup> [https://www.acrai.at/images/download/ACRAI\\_whitebook\\_online\\_2018.pdf](https://www.acrai.at/images/download/ACRAI_whitebook_online_2018.pdf) Eine nationale Strategie ist für Sommer 2019 angekündigt. (<https://derstandard.at/2000092010788/Oesterreichische-Robotik-und-KI-Strategie-fuer-Sommer-2019-geplant>)

<sup>39</sup> [https://www.stopkillerrobots.org/wp-content/uploads/2018/11/KRC\\_CountryViews22Nov2018.pdf](https://www.stopkillerrobots.org/wp-content/uploads/2018/11/KRC_CountryViews22Nov2018.pdf)

Österreich beruft sich in seiner ethischen Vorgehensweise von Anfang an auf die Werte der Europäischen Menschenrechts-Charta<sup>40</sup>. Dazu gehören z. B. Gerechtigkeit, Fairness, Diversität und Inklusion. Bedingung hierfür sei eine zu erreichende Transparenz und Sicherheit der Technologie. Aktuell müsse immer der Mensch jede ethische Entscheidung in Bezug auf KI vornehmen und verantworten. Eine Verlagerung der Entscheidungsbefugnis auf das KI-System erfordere einen klaren rechtlichen und moralischen Rahmen. Zusätzlich sollten "computationale Vorkehrungen für ethisches Verhalten" direkt in Maschinenarchitektur und -design integriert werden (*Ethics by design*). Um alle weiteren Entscheidungen am Gemeinwohl der Bevölkerung auszurichten, plant Österreich ebenfalls umfangreiche Expertenconsultationen und öffentliche Diskussionen.

## Deutschland

Aktuelles Dokument mit Aussagen zu Ethik-Vorgehensweisen: *Strategie Künstliche Intelligenz der Bundesregierung*<sup>41</sup>

Veröffentlicht: November 2018

Die deutsche KI-Strategie beinhaltet insgesamt die konkretesten Maßnahmen und Budgetangaben der verglichenen Staaten. Die Beschreibung der ethischen Vorgehensweise hingegen entspricht im Abstraktionsgrad ungefähr dem europäischen Durchschnitt.

Grundsätzlich sei der Ordnungsrahmen durch die deutsche Gesetzgebung und die DSGVO bereits sehr stabil, aber Lücken bei Algorithmen- und KI-basierten Entscheidungen, Dienstleistungen und Produkte würde man prüfen. Wie bei vielen anderen ethischen Vorgehensweisen fordert man Transparenz, Nachvollziehbarkeit und Überprüfbarkeit der Algorithmen als Schutz gegen Bias und Manipulation.<sup>42</sup> Es wird geprüft, ob zur Sicherung dieser Merkmale auch staatliche Stellen eingerichtet oder ausgebaut werden sollen. In jedem Fall sind Auditierungs-Standards und Standards für Folgenabschätzung geplant. Aber auch technologische Möglichkeiten zur Durchsetzung von Recht und ethischen Prinzipien sollen erkundet werden.

Ein Schwerpunkt liegt auf der digitalen Bildung der Bürger zu mündigen und selbstbestimmte Individuen in Bezug auf KI. Ethik-Fragen und soziologische Zusammenhänge sollen Teil von Lehre, Aus- und Weiterbildung werden. Es sollen Informations-Kampagnen gestartet werden, die Digital- und Medienkompetenz der Bürger fördern sowie zur kritischen Reflexion auf Basis realistischer Anwendungsszenarien befähigen und motivieren. Die Regierung betont die Notwendigkeit der Aufklärung der Gesellschaft und einer kontinuierlichen Diskussion über ethischen Rahmenbedingungen unter gleichberechtigtem Einbezug der Technik- und Naturwissenschaften sowie der Human-, Sozial- und Geisteswissenschaften. Hierfür sollen Foren und Debatten bereitgestellt werden. Bundesregierung beschreibt ihren Ansatz selbst als "ethics by, in and for design" und möchte damit einen Beitrag zum Markenzeichen der "AI made in Europe" leisten.

---

<sup>40</sup> [http://www.europarl.europa.eu/charter/pdf/text\\_de.pdf](http://www.europarl.europa.eu/charter/pdf/text_de.pdf)

<sup>41</sup> [https://www.bmbf.de/files/Nationale\\_KI-Strategie.pdf](https://www.bmbf.de/files/Nationale_KI-Strategie.pdf) Die in der Strategie benannten ethischen Vorgehensweisen basieren auf den *Empfehlungen der Datenethikkommission für die Strategie Künstliche Intelligenz der Bundesregierung* ([https://www.bmfv.de/SharedDocs/Downloads/DE/Ministerium/ForschungUndWissenschaft/DEK\\_Empfehlung\\_en.pdf?\\_\\_blob=publicationFile&v=2](https://www.bmfv.de/SharedDocs/Downloads/DE/Ministerium/ForschungUndWissenschaft/DEK_Empfehlung_en.pdf?__blob=publicationFile&v=2)) Eine überarbeitete Version der Nationalen KI-Strategie

<sup>42</sup> Bemerkenswert ist, dass dieser Schutz auch benachteiligten Bevölkerungsgruppen in Entwicklungsländern gelten soll.

## Europäische Union

Aktuelles Dokument mit Aussagen zu Ethik-Vorgehensweisen: Draft Ethics Guidelines for Trustworthy AI<sup>43</sup>

Veröffentlicht: Dezember 2018

Der aktuelle Entwurf der hochkarätig besetzten *High Level Expert Group (HLEG)*<sup>44</sup> wurde als "Working Document for stakeholders' consultation" veröffentlicht mit der Bitte um Feedback bis zum 1. Februar 2019.<sup>45</sup> Die finalen Guidelines sollen im März 2019 der EU Kommission vorgestellt werden. Der bisherige Entwurf betont besonders die notwendige Ausrichtung von KI auf den Menschen und seine Bedürfnisse ("human-centric design"). Nur wenn KI als "trustworthy", d. h. als gutartig und technisch robust wahrgenommen wird, würden sich die Menschen gerne darauf einlassen. Damit betont die EU die Notwendigkeit von Ethik in der KI für ihre Verbreitung und Weiterentwicklung.

Kern des Drafts ist ein Prozessvorschlag zur Umsetzung von "trustworthy" KI: ausgehend von den ethischen Vorsätzen (z. B. den fundamentalen Menschenrechten) werden fünf ethische Prinzipien<sup>46</sup> abgeleitet, diese werden in 10 konkrete Anforderungen übersetzt, für die dann technische und nicht-technische Implementierungsmethoden vorgestellt werden. Für die Überprüfung der 10 Anforderungen schlägt die HLEG abschließend einen Fragenkatalog vor.

Mit ihrem Draft bietet die HLEG unter den untersuchten ethischen Vorgehensweisen das vollständigste Gerüst für die Implementierung ethischer KI. Es stellt einen Top-down-Implementierungsvorgang dar (vom ethischen Vorsatz bis zur konkreten Umsetzung in der Situation), der aber selbst bottom-up erstellt wurde (von der konkreten Situation des Menschen zum ethischen Vorsatz). Diesen dialektischen und dialogischen Prozess beabsichtigt die EU Kommission kontinuierlich fortzuführen. Die HLEG schlägt ihrerseits die Vermittlung ihres Ansatzes in Trainings und Bildungsangeboten für KI-Entwickler und -Nutzer vor. Komplementär dazu plant die Europäische Kommission eine führende Rolle bei der globalen Debatte über ethische Prinzipien einzunehmen und einschlägige Forschung zu unterstützen.

### 3.1.2 Die "Weltführer" im KI-Bereich

#### USA

Aktuelles Dokument mit Aussagen zu Ethik-Vorgehensweisen: FUTURE of Artificial Intelligence Act<sup>47</sup>

---

<sup>43</sup> [https://ec.europa.eu/futurium/en/system/files/ged/ai\\_hleg\\_draft\\_ethics\\_guidelines\\_18\\_december.pdf](https://ec.europa.eu/futurium/en/system/files/ged/ai_hleg_draft_ethics_guidelines_18_december.pdf)  
Daneben gibt es ein Statement der European Group on *Ethics, Science and New Technologies (EGE)*, in dem als Kernergebnis neun ethische Prinzipien vorgestellt werden ([http://ec.europa.eu/research/ege/pdf/ege\\_ai\\_statement\\_2018.pdf](http://ec.europa.eu/research/ege/pdf/ege_ai_statement_2018.pdf)). Ethisch relevant ist auch die DSGVO (<https://eur-lex.europa.eu/legal-content/DE/TXT/?uri=celex%3A32016R0679>). Die übergreifende KI-Strategie der EU ist zusammengefasst in der *Communication Artificial Intelligence for Europe* (<https://ec.europa.eu/digital-single-market/en/news/communication-artificial-intelligence-europe>).

<sup>44</sup> <https://ec.europa.eu/digital-single-market/en/high-level-expert-group-artificial-intelligence>

<sup>45</sup> <https://ec.europa.eu/futurium/en/ai-stakeholders-consultation/stakeholders-consultation-draft-ai-ethics-guidelines>

<sup>46</sup> In der vorausgegangenen *Communication Artificial Intelligence for Europe* hatte die EU u. a. auf die 10 Prinzipien der *UNI Global Union* ([http://www.thefutureworldofwork.org/media/35420/uni\\_ethical\\_ai.pdf](http://www.thefutureworldofwork.org/media/35420/uni_ethical_ai.pdf)) verwiesen. Auffällig ist, dass diese Prinzipien noch ein Verbot von LAWS vorsahen. Weder die HLEG noch die EGE haben diese Forderung bisher in ihren Prinzipien übernommen.

<sup>47</sup> <https://www.congress.gov/bill/115th-congress/house-bill/4625/text>

Veröffentlicht: Dezember 2017

In den USA liegt die staatliche Verantwortung für KI aktuell vorrangig beim Pentagon. Januar 2019 haben die USA die Entwicklung ethischer Prinzipien für den militärischen Einsatz von KI begonnen.<sup>48</sup> Das bislang zentrale Dokument der staatliche KI-Strategie, der *FUTURE of Artificial Intelligence Act*, empfiehlt Ethik-Trainings für KI-Entwickler, aber empfiehlt auch zu prüfen, ob (!) und wie ethische Standards in die Entwicklung und Anwendung von KI eingebunden werden können. Dem in der öffentlichen Wahrnehmung noch gering erscheinenden staatliche Engagement der USA im Bereich KI-Ethik steht dort eine Vielzahl nicht-staatlicher Einrichtungen mit internationaler Reichweite gegenüber. Einige davon werden in Kapitel **Fehler! Verweisquelle konnte nicht gefunden werden.** skizziert.

## China

Aktuelles Dokument mit Aussagen zu Ethik-Vorgehensweisen: *New Generation Artificial Intelligence Development Plan*<sup>49</sup>

Veröffentlicht: Juli 2017

China wird aufgrund des für 2020 angekündigten *Social Credit System*<sup>50</sup> häufig als ethisches Negativ-Beispiel dargestellt. Aber auch dort intensiviert sich die ethische Debatte ausgehend vom Thema moralischer Agenten.<sup>51</sup> Und zumindest hinsichtlich der Menge und Konkretheit von Angaben zu ethischen Vorgehensweisen ist der 2017 vorgestellte KI-Entwicklungsplan absolut vergleichbar mit europäischen Strategien. Wie bei allen anderen Strategien werden ethische Normen, Gesetze und Regulierungen angestrebt. Deren kontinuierliche Optimierung durch Forschungsaktivitäten ist bis 2030 geplant. Zusätzlich soll ein ethisches Framework für Mensch-Computer-Kollaborationen entstehen und ein ethischer Verhaltenskodex für Forschung und Entwicklung zu KI-Produkten. Potenzielle Risiken von KI sollen bewertbar gemacht werden und es soll Lösungen für Notfall-Szenarien geben. Bei Industrie und Unternehmen setzt China auf die Förderung von Selbstdisziplin und die Stärkung der Manager-Moral. Letzteres wurde in keiner anderen untersuchten Ethik-Vorgehensweise so formuliert.

## 3.2 Nicht-staatlicher Akteure

Auch im Privatsektor gibt es eine Vielzahl teils wettbewerblicher teils gemeinnütziger Institutionen, die sich um ethische KI bemühen. Nachfolgend sind einige der bekanntesten Institutionen aufgelistet in chronologischer Reihenfolge ihrer Gründung.

### Future of Life Institute

Das 2014 in Cambridge gegründete *Future of Life Institute* ist eine außeruniversitäre Forschungseinrichtung und befasst sich mit den existenzbedrohenden Menschheitsrisiken

---

<sup>48</sup> <https://www.defenseone.com/technology/2019/01/pentagon-seeks-list-ethical-principles-using-ai-war/153940/>

<sup>49</sup> Dieser Plan existiert bisher nicht in englischer Sprache aber hier findet sich eine Übersetzung: <https://chinacopyrightandmedia.wordpress.com/2017/07/20/a-next-generation-artificial-intelligence-development-plan/>

<sup>50</sup> <https://www.businessinsider.de/china-social-credit-system-punishments-and-rewards-explained-2018-4?r=US&IR=T>

<sup>51</sup> [https://www.fhi.ox.ac.uk/wp-content/uploads/Deciphering\\_Chinas\\_AI-Dream.pdf](https://www.fhi.ox.ac.uk/wp-content/uploads/Deciphering_Chinas_AI-Dream.pdf)

Atomkriege, Biotechnologie, Klimawandel und Künstliche Intelligenz.<sup>52</sup> Mit finanzieller Unterstützung von Elon Musk förderte das Institut seit 2015 insgesamt 47 Projekte für nutzbringende KI mit Summen zwischen 100.000 und 1,5 Mio. USD (Future of Life Grants). Im Rahmen der 2017 abgehaltenen Asilomar Conference hat das Institut die Asilomar Principles (s. 0) entwickelt.

## **FAT/ML**

*FAT/ML* steht für *Fairness, Accountability and Transparency in Machine Learning*.<sup>53</sup> Die seit 2014 jährlich stattfindende Konferenz richtet sich an internationale KI-Forscher, die sich mit informatischen Methoden zur Sicherstellung ethischer KI befassen (Ethics by design). Die *Principles for Accountable Algorithms and a Social Impact Statement for Algorithms* wurden im Juli 2016 formuliert und sollen KI-Entwickler und Produktmanager konkrete ethische Orientierung bieten.<sup>54</sup>

## **Ethics in Artificial Intelligence**

Die *Ethics in Artificial Intelligence*<sup>55</sup> Initiative wurde 2015 mit Geldern der *Future of Life Grants* an der *Oxford University* gegründet. Die Initiative fokussiert weniger inhaltlich auf die ethischen Herausforderungen im Zusammenhang mit KI, sondern sie befasst sich explizit mit der Frage, wie ethische Codes erarbeitet werden können und aufgebaut sein sollten. Online wird eine umfangreiche Übersicht ethischer Codes sowie weiterführender Links und Dokumente über die Erstellung ethischer Codes angeboten.<sup>56</sup>

## **Open AI**

*Open AI*<sup>57</sup> ist eine gemeinnützige Forschungseinrichtung, die Ende 2015 von Elon Musk und Sam Altman in San Francisco gegründet wurde. Ihr Ziel ist die Untersuchung von Sicherheit und Ethik und die Entwicklung von Produkten in Richtung starker KI.<sup>58</sup> Im Februar 2018 trat Musk aufgrund potentieller Zielkonflikte mit seiner Arbeit bei Tesla aus dem Vorstand zurück.<sup>59</sup>

## **Partnership on AI**

Die *Partnership on AI*<sup>60</sup> ist eine der größten KI-Initiativen mit über 80 Partnern unterschiedlicher Branchen. Sie wurde im September 2016 unter anderem von Facebook, Amazon und Google gegründet. Ihr Ziel ist gemeinsam mit den Partnern Best Practices für KI zu etablieren und die Bevölkerung über Möglichkeiten und Reichweite der Technologie zu informieren. Die Partner haben sich selbst zu acht ethischen Leitsätzen<sup>61</sup> verpflichtet.

## **AI4ALL**

---

<sup>52</sup> <https://futureoflife.org/background/existential-risk/>

<sup>53</sup> <http://www.fatml.org/>

<sup>54</sup> <http://www.fatml.org/resources/principles-for-accountable-algorithms>

<sup>55</sup> <https://www.cs.ox.ac.uk/efai/>

<sup>56</sup> <https://www.cs.ox.ac.uk/efai/resources/>

<sup>57</sup> <https://openai.com/>

<sup>58</sup> <https://openai.com/systems/>

<sup>59</sup> <https://www.theverge.com/2018/2/21/17036214/elon-musk-openai-ai-safety-leaves-board>

<sup>60</sup> <https://www.partnershiponai.org/>

<sup>61</sup> <https://www.partnershiponai.org/tenets/>

*AI4ALL*<sup>62</sup> ist eine gemeinnützige Bildungsinitiative, die Anfang 2017 an der Stanford University gegründet wurde. Ihr Ziel ist Diversity und Inklusion in der Lehre, Forschung, Entwicklung und Gesetzgebung im Bereich KI zu erhöhen durch Qualifizierung und die Etablierung von Role Models. Ihr Bildungsangebot richtet sich speziell an Bevölkerungsgruppen, die an der Universität unterrepräsentiert sind.

### **Ethics and Governance of Artificial Intelligence**

Die in Harvard ansässige *Ethics and Governance of Artificial Intelligence Initiative*<sup>63</sup> ist seit 2017 tätig. Sie wurde in Kooperation des *Berkman Klein Center* und des *MIT Media Lab* mit 27 Mio. USD Förderung der *Knight-Foundation* gegründet.<sup>64</sup> Sie erforscht für private und öffentliche Entscheidungsträger Möglichkeiten zur gemeinwohlorientierten Nutzung von KI. In einem offenen Wettbewerb bietet die Initiative insgesamt 750.000 USD für Projekte zu vier ethischen Herausforderungen.<sup>65</sup>

### **AI Ethics Lab**

Das *AI Ethics Lab*<sup>66</sup> wurde im Januar 2018 ebenfalls an der *Harvard University* gegründet und bietet Forschern und Praktikern Beratung, Trainings, Workshops und Vorträge rund um die ethische Implementierung von KI. Ihr Mapping-Workshop hilft KI-Entwicklern ethische Konfliktquellen in ihren Projekten zu identifizieren und Lösungswege zu erarbeiten, ohne dabei die technische Integrität zu riskieren.<sup>67</sup>

### **MIT Stephen A. Schwarzman College of Computing**

Im Oktober 2018 gab das *Massachusetts Institute of Technology* (MIT) die Investition von 1 Mrd. USD bekannt für die Adressierung globaler Chancen und Herausforderungen durch KI.<sup>68</sup> Im Zentrum steht die Gründung des *Stephen A. Schwarzman College of Computing*, benannt nach dem CEO des Vermögensverwalters Blackstone, der persönlich 350 Mio. USD gespendet hatte. Die Förderung von Ethik in der KI soll ein wichtiger Bestandteil der Aktivitäten des College sein. Beispielsweise sollen neue Lehrpläne erarbeitet werden, die KI mit anderen Disziplinen verbinden, oder es sollen Foren ausgerichtet werden, bei denen Führungspersönlichkeiten aus Politik, Wirtschaft, Forschung und Journalismus über die möglichen Folgen neuester KI-Fortschritte diskutieren und Handlungsbedarfe ableiten. Die inhaltliche Arbeit beginnt voraussichtlich im September 2019.<sup>69</sup>

## **3.3 "Big Five" Technologie-Konzerne**

Die "Big Five" Tech-Konzerne *Amazon*, *Apple*, *Facebook*, *Google* und *Microsoft* nehmen eine entscheidende Vorbildfunktion im ethischen Diskurs um KI ein. Sie alle sind Mitglieder der *Partnership on AI* und der Industriegruppe *Information Technology Industry Council* (ITI)<sup>70</sup>, die im

---

<sup>62</sup> <http://ai-4-all.org/>

<sup>63</sup> <https://aiethicsinitiative.org/>

<sup>64</sup> <https://robotics.mit.edu/mit-media-lab-participate-27-million-initiative-ai-ethics-and-governance>

<sup>65</sup> <https://aiethicsinitiative.org/challenge/>

<sup>66</sup> <http://aiethicslab.com>

<sup>67</sup> <http://aiethicslab.com/the-mapping/>

<sup>68</sup> <http://news.mit.edu/2018/mit-reshapes-itself-stephen-schwarzman-college-of-computing-1015>

<sup>69</sup> <http://news.mit.edu/2018/faq-mit-stephen-schwarzman-college-of-computing-1015>

<sup>70</sup> <https://www.itic.org/>



Oktober 2017 *AI Policy Principles*<sup>71</sup> veröffentlicht hat.<sup>72</sup> Nachfolgend werden die zusätzliche, selbstinitiierten Maßnahmen der Firmen dargestellt. Das ethische Engagement von *Apple* und *Amazon* scheint sich hinsichtlich KI-Bereich auf die Mitgliedschaft in den beiden Vereinigungen zu beschränken,<sup>73</sup> bzw. sie besitzen allgemeine Ethik-Grundsätze, aber eben keine KI-spezifischen.

## Google

*Google* hat bereits im Jahr 2014 ein *AI Ethics Board* eingerichtet. Sowohl Mitglieder als auch Aufgaben sind bisher unbekannt.<sup>74</sup> Im Juni 2018 hat *Google* seine Prinzipien<sup>75</sup> für ethische KI veröffentlicht und kurz darauf folgten Empfehlungen für *Responsible AI Practices*<sup>76</sup>. Im Dezember 2018 kündigte *Google* über seinen Blog einen umfangreichen Maßnahmenkatalog an.<sup>77</sup> Dieser beinhaltet unter anderem Trainings, eine *AI Ethics Speaker Series*, ein Lehrmodul zu Fairness im Machine Learning Crashkurs sowie ein interdisziplinäres *Responsible Innovation Team*, eine Gruppe erfahrener Seniorberater und einen Beirat erfahrener Führungskräfte für komplexere Fragestellungen, die gleich mehrere Produkte und Technologien betreffen.

## Facebook

*Facebook* hat auf seiner *F8* Entwicklerkonferenz im Mai 2018 bekanntgegeben, ein eigenes Ethics Team gegründet zu haben.<sup>78</sup> Dieses Team kooperiert unternehmensweit mit Entwicklerteams und hat die Software *Fairness Flow* entwickelt, die Bias einer anderen KI erkennen könne.<sup>79</sup> *Fairness Flow* wurde in die Software *FB Lerner Flow* integriert, mit dem *Facebook* seine Algorithmen trainiert. Wissenschaftliche Publikationen dazu sollen folgen. *Facebook* plant den Bias seiner Software zusätzlich zu minimieren, indem es die Diversität seiner Belegschaft erhöht, z. B. durch Büroeröffnungen weit entfernt vom Silicon Valley.

## Microsoft

*Microsoft* hat 2017 die *Fairness Accountability Transparency and Ethics group* (FATE)<sup>80</sup> gegründet. Die Gruppe möchte informatische Methoden entwickeln, die sowohl ethisch als auch innovativ sind, und plant darüber zu publizieren. Im selben Jahr kam *AI and ethics in engineering and research* (AETHER) dazu, ein unternehmensweit rekrutierter Beirat erfahrener Führungskräfte, der ethische Probleme frühzeitig erkennen und angehen soll.<sup>81</sup> *Microsoft* hat im Januar 2018 eigene Ethik-Prinzipien veröffentlicht und hat im November desselben Jahres ethische Guidelines für die Entwicklung von KI-

---

<sup>71</sup> <https://www.itic.org/resources/AI-Policy-Principles-FullReport2.pdf>

<sup>72</sup> Welche Vorteile oder Konflikte daraus entstehen, wenn sich ein Unternehmen gleich mehreren Sets von Prinzipien verpflichtet, wird in diesem Artikel nicht untersucht.

<sup>73</sup> <https://phys.org/news/2018-11-tech-companies-ethical.html>

<sup>74</sup> <https://www.forbes.com/sites/samshead/2018/04/27/googles-mysterious-ai-ethics-board-should-be-as-transparent-as-axons/#3b4a425d19d1>

<sup>75</sup> <https://ai.google/principles/>

<sup>76</sup> <https://ai.google/education/responsible-ai-practices>

<sup>77</sup> <https://www.blog.google/technology/ai/google-ai-principles-updates-six-months/>

<sup>78</sup> <https://www.forbes.com/sites/samshead/2018/05/03/facebook-reportedly-has-a-dedicated-ai-ethics-team/#5690de1d2e5c>

<sup>79</sup> <https://www.forbes.com/sites/samshead/2018/05/03/facebook-reportedly-has-a-dedicated-ai-ethics-team/#607f61d02e5c>

<sup>80</sup> <https://www.microsoft.com/en-us/research/group/fate/>

<sup>81</sup> <https://www.businessinsider.de/microsoft-research-forms-new-ai-unit-2017-7?r=US&IR=T>

Chatbots formuliert.<sup>82</sup> Im Dezember 2018 kamen ethische Prinzipien für den Einsatz von Gesichtserkennung hinzu.<sup>83</sup>

### 3.4 Weitere bekannte KI-Ethik-Erklärungen

#### **Ethically Aligned Design**

Die Broschüre *Ethically Aligned Design: A Vision for Prioritizing Human Well-being with Autonomous and Intelligent Systems* basiert auf dem Input mehrerer hundert Experten von sechs Kontinenten. 2016 wurde die erste Version<sup>84</sup> veröffentlicht. 2017 folgte eine zweite Version mit der Bitte um Feedback der Öffentlichkeit.<sup>85</sup> 2019 soll die überarbeitete Variante finalisiert werden. Im Zentrum stehen fünf Prinzipien, die einen ethischen Rahmen für die Gestaltung, Entwicklung und Anwendung von KI. Die Broschüre wird herausgegeben von der *Ethics in Action Initiative*<sup>86</sup> des amerikanischen *Institute of Electrical and Electronics Engineers* (IEEE), dem weltgrößten Verband technischer Berufe. Seit Juni 2016 arbeitet dieser Verband außerdem an der Standard-Serie *P7000*<sup>87</sup>, bei der Prozessmodelle zur Adressierung ethischer Überlegungen bei der Konzeption und Gestaltung von IT-Systemen entwickelt werden.

#### **Asilomar Principles**

Die 23 *Asilomar Principles*<sup>88</sup> sind die bekannteste Auflistung ethischer Prinzipien für nutzbringende KI. Formuliert wurden sie im Januar 2017 auf der Asilomar Konferenz in Californien von ca. 100 Experten aus KI, Wirtschaft, Industrie, Philosophie und Sozialwissenschaften. Zu den mittlerweile fast 4.000 Unterzeichnern der Prinzipien gehören z. B. Elon Musk, Ray Kurzweil und Stephen Hawking.

#### **Toronto Declaration**

Im Rahmen der RightsCon Konferenz wird im Mai 2018 auf Initiative von *Amnesty International* und *AccessNow* die *Toronto Declaration*<sup>89</sup> für Menschenrechte und Künstliche Intelligenz erstellt.<sup>90</sup> Sie fordert eine stärkere Zusammenarbeit des öffentlichen und privaten Sektors zur Sicherung der Menschenrechte im Zeitalter Künstlicher Intelligenz und erinnern dafür z. B. an die *Vienna Declaration*<sup>91</sup> des *UN Human Rights Committee* oder das *International Covenant on Economic, Social and Cultural Rights*<sup>92</sup>. Weitere Unterstützer sind *Human Rights Watch* und die *Wikimedia Foundation*.

#### **Montreal Declaration**

Die *Montreal Declaration for Responsible Development of Artificial Intelligence*<sup>93</sup> wurde von der *Université de Montréal* gemeinsam mit dem *Fonds de recherche du Québec* erarbeitet. Sie wurde im

---

<sup>82</sup> <https://www.microsoft.com/en-us/ai/our-approach-to-ai>

<sup>83</sup> <https://www.reuters.com/article/us-microsoft-facial-recognition/microsoft-adopts-ethical-principles-aiming-to-bar-misuse-of-face-recognition-technology-idUSKBN1O52NS>

<sup>84</sup> <https://standards.ieee.org/industry-connections/ec/ead-v1.html>

<sup>85</sup> <https://ethicsinaction.ieee.org/>

<sup>86</sup> ebd.

<sup>87</sup> <https://standards.ieee.org/project/7000.html>

<sup>88</sup> <https://futureoflife.org/ai-principles/>

<sup>89</sup> [https://www.accessnow.org/cms/assets/uploads/2018/08/The-Toronto-Declaration\\_ENG\\_08-2018.pdf](https://www.accessnow.org/cms/assets/uploads/2018/08/The-Toronto-Declaration_ENG_08-2018.pdf)

<sup>90</sup> <https://www.rightscon.org/access-now-amnesty-international-launch-toronto-declaration/>

<sup>91</sup> <https://www.ohchr.org/EN/ProfessionalInterest/Pages/Vienna.aspx>

<sup>92</sup> <https://www.ohchr.org/EN/ProfessionalInterest/Pages/CESCR.aspx>

<sup>93</sup> <https://www.montrealdeclaration-responsibleai.com/the-declaration>

Dezember 2018 in der Society for Arts and Technology präsentiert. Sie ist das Ergebnis mehr als ein Jahr andauernder Konsultation von Philosophen, Soziologen, Juristen und KI-Forschern sowie Online- und Face-to-face-Befragungen der Bevölkerung.<sup>94</sup> Ihre zehn Prinzipien sollen einen ethischen Rahmen für die Entwicklung und Anwendung von KI bieten, als Orientierung in der digitalen Transformation der Gesellschaft dienen und Basis für weitere Debatten sein. Auch zukünftig sind kontinuierliche Überarbeitungen der Declaration geplant.

## 4 Handlungsempfehlungen

Der Vergleich der Akteure offenbart große Unterschiede zwischen den einzelnen Ethik-Vorgehensweisen. Dies lässt vermuten, dass eine ebenso große Unsicherheit bezüglich des richtigen Vorgehens herrschen könnte. Keine der Ethik-Vorgehensweisen scheint sich als die eine beste hervorzutun, kein Set ethischer Prinzipien wird langfristig oder in allen Fällen nützlich sein

Daher werden nachfolgend Maßnahmen empfohlen, die nicht auf eine starre ethische Lösung abzielen, sondern auf einen flexiblen und dauerhaften Lösungsprozess, der den spezifischen Kontext berücksichtigt und auf dem Wissen aller Beteiligten aufbaut.

### 1. JETZT handeln!

Es darf keine weitere Zeit verloren gehen. Wir brauchen klare Ziele, wirksame Gesetze und ethisch perfekte Entwicklungsprozesse. Bis ein Gesetz alle Hürden genommen hat und in Kraft getreten ist, hat sich die Technologie unter Umständen schon wieder grundlegend verändert.<sup>95</sup> Die Anwendung wartet nicht. Deswegen heißt es: Ban the arms race, start the ethics race! Eine gemeinwohlorientierte KI-Gesellschaft steht in Aussicht, wenn es uns jetzt gelingt, die knappen Humanressourcen an KI- und Ethik-Experten für das Thema zu gewinnen und die gesellschaftliche Transformation gemeinsam zu gestalten.

Diese Debatte hat oberste Priorität. Es mag wertvoll und inspirierend sein, sich mit ethischen Fragen für starke<sup>96</sup> Künstliche Intelligenzen oder sogar Superintelligenzen zu beschäftigen.<sup>97</sup> Viel dringlicher ist die Diskussion über schwache künstliche Intelligenz in Politik, Wirtschaft, Wissenschaft und Gesellschaft.

### 2. Alle einbeziehen!

KI ist das, was wir daraus machen. Deswegen ist jeder gefragt sich mit seinen Ideen, Ängsten und Träumen in die gesellschaftliche Debatte einzubringen: je unterschiedlicher die Perspektiven, umso

---

<sup>94</sup> <https://theconversation.com/the-montreal-declaration-why-we-must-develop-ai-responsibly-108154>

<sup>95</sup> <https://www.pcwelt.de/a/ethik-und-ki-die-zeit-draengt-wir-muessen-handeln,3451885>

<sup>96</sup> Meist unterscheidet man schwache KI (ANI= Artificial Narrow Intelligence), starke KI (AGI=Artificial General Intelligence) und Künstliche Superintelligenz (ASI= Artificial Superintelligence). Mit schwacher KI erzielt man menschenähnliche oder sogar übermenschliche Leistungen *auf einem Spezialgebiet* wie z. B. in der Bilderkennung, beim Schachspiel oder dem Social Scoring von Staatsangehörigen. (Alle bisher existierenden KI-Anwendungen gehören dieser Klasse an.) Starke KI soll hingegen *auf allen Gebieten* menschenähnliche Leistungsfähigkeit besitzen. Eine künstliche Superintelligenz besäße sogar übermenschliche Leistungsfähigkeit *auf allen Gebieten*. Starke KI und Künstliche Superintelligenz existieren bisher nur in Science-Fiction-Werken wie z. B. der Fernsehserie "Westworld" oder Mary Shelly's "Frankenstein".

<sup>97</sup> Siehe hierzu z. B. "The Ethics of Artificial Intelligence" von Nick Boström und Eliezer Yudkowsky. (<https://nickbostrom.com/ethics/artificial-intelligence.pdf>)

nachhaltiger das Ergebnis. Dabei geht es nicht um die Einigung auf *ein* Zukunftsszenario, sondern die Bereitschaft und Kompetenz pluralistische Vorstellungen zu verhandeln - "in Vielfalt geeint"<sup>98</sup>.

Die gesamte Bevölkerung sollte zur ethischen Debatte über KI befähigt werden: durch Information, Dialog und indem KI, Ethik und KI-Ethik feste Bestandteile von Bildungsangeboten auf allen Ebenen werden, von der Grundschule bis zum Management-Seminar. Die Demokratisierung von KI-Wissen ist der beste Weg, um algorithmischem Bias, Macht-Asymmetrien und sozialer Ungleichheit vorzubeugen - und unsere heutige Technologie macht es leichter denn je.

### **3. Begriffe klären!**

Um einen ethischen Diskurs zwischen Individuen, Organisationen und Nationen effizient führen zu können, sollte eine Einigung über die grundlegenden Begrifflichkeiten vorgenommen werden.<sup>99</sup> In welchen Fällen spricht man von Prinzipien, Werten, Regeln, etc.? Wann spricht man von Auditability, Traceability, Explicability, etc.? Ein entsprechender Vorschlag sollte von einer Institution mit ausreichender Aufmerksamkeit, Reputation und Reichweite in der Ethik-Debatte kommen (wie z. B. dem Future of Life Institut oder den United Nations) und von den Verfassern ethischer Vorgaben beherzigt werden.

### **4. Iterativ vorgehen!**

Die Einsatzmöglichkeiten von KI wachsen kontinuierlich. Gleichzeitig unterscheiden sich ethische Vorstellungen von Person zu Person und von Zeit zu Zeit. In einer Ethik der KI multipliziert sich das Problem. Diese Komplexität und Dynamik erfordern iterative Prozessmodelle der Ethik-Entwicklung, die sich ständig der aktuellen Situation anpassen, um die Wartezeit zwischen Identifikation und Beantwortung neuer ethischer Fragen zu minimieren.

Nicht die einzelne Antwort ist das Ziel, sondern die Fähigkeit schnell und immer wieder gemeinsame Antworten finden zu können. Dies setzt zum einen effiziente Kommunikationsstrukturen- und Prozesse zwischen Gesellschaft, Wissenschaft, Wirtschaft und Politik voraus, aber vor allem Moderationskompetenz und Kreativität des Individuums - das sind die Skills, die in Zukunft zählen.

### **5. Kunst integrieren!**

Ein ressourcensparender Weg zur ethischen KI ist die Integration der Kunst in den Technologieentwicklungsprozess. Die Einbindung künstlerischer Perspektiven<sup>100</sup> begünstigt

---

<sup>98</sup> [https://europa.eu/european-union/about-eu/symbols/motto\\_de](https://europa.eu/european-union/about-eu/symbols/motto_de)

<sup>99</sup> Die Diskussion über LAWS bei einer UN-Konferenz im August 2018 ist ein Beispiel für die Notwendigkeit gemeinsamer Begriffsbestimmungen. Laut der Campaign to Stop Killer Robots (<https://www.stopkillerrobots.org/>) hätten die Befürworter solcher Waffensysteme die ethische Debatte alleine dadurch ausbremsen können, dass sie keiner Definition für LAWS zustimmten. (<https://www.independent.co.uk/life-style/gadgets-and-tech/news/killer-robots-un-meeting-autonomous-weapons-systems-campaigners-dismayed-a8519511.html>) Ein positives Beispiel ist die Erarbeitung der Montreal Declaration. Zunächst wurde ein Begriffsglossar definiert und anschließend der Erklärung beigefügt. (<https://www.montrealdeclaration-responsibleai.com/the-declaration>)

<sup>100</sup> Diese Formulierung soll betonen, dass hierfür nicht zwingend ausgebildete Künstler notwendig sind. Mit der richtigen Kreativmethode kann jeder zu künstlerischen Sichtweisen befähigt werden.

Innovationen, die sowohl kreativer als auch ethischer.<sup>101</sup> Kunst ermöglicht komplexe Sachverhalte greifbar und damit effektiv diskutierbar zu machen.<sup>102</sup> Zum einen sollten KI-Akteure mehr Residencies für Künstler anbieten zum anderen sollten Future Labs<sup>103</sup> gezielt für Kunst und KI eingerichtet werden - um Antworten zu finden, die bisher undenkbar sind.

*"Und würden alle Menschen nach dem Schönen und Edlen wetteifern und ihre Kraft anspannen, das Edelste zu tun, dann wäre der allgemeine Zustand so, wie er sein sollte, und jedem einzelnen würden die höchsten Güter zu teil [...]" - Aristoteles*

---

<sup>101</sup> Dies ist die Argumentation des Programms S+T+ARTS der EU (<https://www.starts.eu/>), dass seit 2016 Projekte an der Schnittstelle von Kunst und Technologie fördert. Die EU empfiehlt dies insbesondere im Bereich KI. (<https://ec.europa.eu/digital-single-market/en/blogposts/artificial-intelligence-next-frontier>)

<sup>102</sup> Siehe hierzu das Beispiel des nach der mystischen Figur des "Kratt" benannte KI-Gesetz Estlands: <https://medium.com/e-residency-blog/estonia-starts-public-discussion-legalising-ai-166cb8e34596>

<sup>103</sup> Vorbild könnte das Ars Electronica Future Lab in Linz sein: <https://ars.electronica.art/futurelab/>

## 5 Anhang

### 6 Auswahl ethischer Prinzipien

Nachfolgend werden alle in diesem Artikel genannten ethischen Prinzipien aufgelistet. Es werden auch solche aufgelistet, die vom Verfasser nicht als Prinzipien bezeichnet werden, aber als solche eingesetzt werden.<sup>104</sup> Zur besseren Vergleichbarkeit wurden bei komplexeren Prinzipien nur die Anfangsbegriffe ausgewählt wie z. B. im Falle *Asilomar Principles*.

#### 6.1 Staatliche Akteure

##### 6.1.1 EU-Staaten

###### Italien<sup>105</sup>

1. Data quality and Neutrality
2. Accountability and Liability
3. Transparency and Openness
4. Protection of private sphere

##### 6.1.2 EU

###### Ethics, Science and New Technologies (EGE)<sup>106</sup>

1. Human Dignity
2. Autonomy
3. Responsibility
4. Justice, equity, and solidarity
5. Democracy
6. Rule of law and accountability
7. Security, safety, bodily and mental integrity
8. Data protection and privacy
9. Sustainability

###### High Level Expert Group (HLEG)<sup>107</sup>

1. Beneficence
2. Non-Maleficence
3. Autonomy of humans
4. Justice
5. Explicability

###### (UNI Gobar Union)<sup>108,109</sup>

---

<sup>104</sup> Italien bezeichnet seine Prinzipien beispielsweise als "Central elements in the public debate and scientific analysis".

<sup>105</sup> <https://ia.italia.it/assets/whitepaper.pdf>

<sup>106</sup> [http://ec.europa.eu/research/ege/pdf/ege\\_ai\\_statement\\_2018.pdf](http://ec.europa.eu/research/ege/pdf/ege_ai_statement_2018.pdf)

<sup>107</sup> [https://ec.europa.eu/futurium/en/system/files/ged/ai\\_hleg\\_draft\\_ethics\\_guidelines\\_18\\_december.pdf](https://ec.europa.eu/futurium/en/system/files/ged/ai_hleg_draft_ethics_guidelines_18_december.pdf)

<sup>108</sup> [http://www.thefutureworldofwork.org/media/35420/uni\\_ethical\\_ai.pdf](http://www.thefutureworldofwork.org/media/35420/uni_ethical_ai.pdf) Diese Prinzipien sind keine offiziellen Prinzipien der EU, aber wurden in der *Communication Artificial Intelligence for Europe*

1. Demand That AI Systems Are Transparent
2. Equip AI Systems With an “Ethical Black Box”
3. Make AI Serve People and Planet
4. Adopt a Human-In-Command Approach
5. Ensure a Genderless, Unbiased AI
6. Share the Benefits of AI Systems
7. Secure a Just Transition and Ensuring Support for Fundamental Freedoms and Rights
8. Establish Global Governance Mechanisms
9. Ban AI Arms Race

## 6.2 Nicht-staatliche Akteure

### Partnership on AI<sup>110</sup>

1. We will seek to ensure that AI technologies benefit and empower as many people as possible.
2. We will educate and listen to the public and actively engage stakeholders to seek their feedback on our focus, inform them of our work, and address their questions.
3. We are committed to open research and dialogue on the ethical, social, economic, and legal implications of AI.
4. We believe that AI research and development efforts need to be actively engaged with and accountable to a broad range of stakeholders.
5. We will engage with and have representation from stakeholders in the business community to help ensure that domain-specific concerns and opportunities are understood and addressed.
6. We will work to maximize the benefits and address the potential challenges of AI technologies [...].
7. We believe that it is important for the operation of AI systems to be understandable and interpretable by people, for purposes of explaining the technology.
8. We strive to create a culture of cooperation, trust, and openness among AI scientists and engineers to help us all better achieve these goals.

### FAT/ML<sup>111</sup>

1. Responsibility
2. Explainability
3. Accuracy
4. Auditability
5. Fairness

### ITI<sup>112</sup>

---

(<https://ec.europa.eu/digital-single-market/en/news/communication-artificial-intelligence-europe>) als gutes Beispiel genannt.

<sup>109</sup> Diese Prinzipien sind keine offiziellen Prinzipien der EU, aber wurden in der *Communication Artificial Intelligence for Europe* (<https://ec.europa.eu/digital-single-market/en/news/communication-artificial-intelligence-europe>) als gutes Beispiel genannt.

<sup>110</sup> <https://www.partnershiponai.org/tenets/>

<sup>111</sup> <http://www.fatml.org/resources/principles-for-accountable-algorithms>

1. Industry's Responsibility in Promoting Responsible Development and Use
2. The Opportunity for Governments to Invest In and Enable the AI Ecosystem
3. The Opportunity for Public-Private Partnerships (PPPs)

### 6.3 "Big Five" Technologie-Konzerne

#### Google<sup>113</sup>

1. Be socially beneficial
2. Avoid creating or reinforcing unfair bias (Fairness)
3. Be built and tested for safety (Safety)
4. Be accountable to people (Accountability)
5. Incorporate privacy design principles (Privacy and transparency in the use of data for training)
6. Uphold high standards of scientific excellence (Share AI knowledge by publishing educational materials, best practices, and research that enable more people to develop useful AI applications.)
7. Be made available for uses that accord with these principles.

#### Microsoft<sup>114</sup>

1. Fairness
2. Reliability & Safety
3. Privacy & Security
4. Inclusiveness
5. Transparency
6. Accountability

### 6.4 Weitere Bekannte KI-Ethik-Erklärungen

#### Ethically Aligned Design<sup>115</sup>

1. Human Rights
2. Well-being
3. Accountability
4. Transparency
5. Awareness of misuse

#### Asilomar Principles<sup>116</sup>

1. Research Goal
2. Research Funding
3. Science-Policy Link
4. Research Culture
5. Race Avoidance

---

<sup>112</sup> <https://www.itic.org/resources/AI-Policy-Principles-FullReport2.pdf>

<sup>113</sup> <https://ai.google/principles/>

<sup>114</sup> <https://www.microsoft.com/en-us/ai/our-approach-to-ai>

<sup>115</sup> <https://ethicsinaction.ieee.org/>

<sup>116</sup> <https://futureoflife.org/ai-principles/>



6. Safety
7. Failure Transparency
8. Judicial Transparency
9. Responsibility
10. Value Alignment
11. Human Values
12. Personal Privacy
13. Liberty and Privacy
14. Shared Benefit
15. Shared Prosperity
16. Human Control
17. Non-subversion
18. AI Arms Race
19. Capability Caution
20. Importance
21. Risks
22. Recursive Self-Improvement
23. Common Good

#### **Toronto Declaration**<sup>117</sup>

1. The right to equality and non-discrimination
2. Protecting the rights of all individuals and groups: promoting diversity and inclusion
3. State use of machine learning systems

#### **Montreal Declaration**<sup>118</sup>

1. Well-being Respect for Autonomy
2. Protection of privacy and intimacy
3. Solidarity
4. Democratic participation
5. Equity
6. Diversity
7. Inclusion
8. Prudence
9. Responsibility
10. Sustainable
11. Development

---

<sup>117</sup> [https://www.accessnow.org/cms/assets/uploads/2018/08/The-Toronto-Declaration\\_ENG\\_08-2018.pdf](https://www.accessnow.org/cms/assets/uploads/2018/08/The-Toronto-Declaration_ENG_08-2018.pdf)

<sup>118</sup> <https://www.montrealdeclaration-responsibleai.com/the-declaration>