

Isabella Hermann

Georgios Kolliarakis

Fruzsina Molnár-Gábor

Timo Rademacher

Frauke Rostalski

1|2020

#VerantwortungKI – Künstliche Intelligenz und gesellschaftliche Folgen

## Vertrauenswürdige KI? Vorausschauende Politik!

Eine Schriftenreihe der interdisziplinären Arbeitsgruppe  
*Verantwortung: Maschinelles Lernen und Künstliche Intelligenz*



berlin-brandenburgische  
AKADEMIE DER WISSENSCHAFTEN





Berlin-Brandenburgische Akademie der Wissenschaften (BBAW)

VERTRAUENSWÜRDIGE KI? VORAUSSCHAUENDE POLITIK!





## VERTRAUENSWÜRDIGE KI ? VORAUSSCHAUENDE POLITIK !

---

Isabella Hermann  
Georgios Kolliarakis  
Fruzsina Molnár-Gábor  
Timo Rademacher  
Frauke Rostalski

## AUTORINNEN UND AUTOREN

**Isabella Hermann:** Wissenschaftliche Koordinatorin der interdisziplinären Arbeitsgruppe *Verantwortung: Maschinelles Lernen und Künstliche Intelligenz* der Berlin-Brandenburgischen Akademie der Wissenschaften.

**Georgios Kolliarakis:** Berater für Forschungsstrategie, Deutsche Gesellschaft für Auswärtige Politik, Berlin.

**Fruzsina Molnár-Gábor:** Gruppenleiterin an der Heidelberger Akademie der Wissenschaften im Bereich des Völker- und Europarechts sowie der Rechtsvergleichung unter besonderer Berücksichtigung von Datenschutz- und Medizinrecht.\*

**Timo Rademacher:** Juniorprofessor für Öffentliches Recht und das Recht der neuen Technologien an der Universität Hannover.\*

**Frauke Rostalski:** Professorin für Strafrecht, Strafprozessrecht, Rechtsphilosophie und Rechtsvergleichung an der Universität zu Köln.\*

\* Mitglied der interdisziplinären Arbeitsgruppe *Verantwortung: Maschinelles Lernen und Künstliche Intelligenz* der Berlin-Brandenburgischen Akademie der Wissenschaften.

Herausgeberin: Interdisziplinäre Arbeitsgruppe *Verantwortung: Maschinelles Lernen und Künstliche Intelligenz* der Berlin-Brandenburgischen Akademie der Wissenschaften.

Redaktion: Isabella Hermann

Grafik: Thorsten Probst/angenehme gestaltung

Druck: Union Sozialer Einrichtungen gGmbH, Berlin

© Berlin-Brandenburgische Akademie der Wissenschaften, 2020

Jägerstraße 22–23, 10117 Berlin, [www.bbaw.de](http://www.bbaw.de)

Nachdruck, auch auszugsweise, nur mit ausdrücklicher Genehmigung der Herausgeber.

ISBN: 978-3-939818-89-2

## INHALTSVERZEICHNIS

<b>Einführung</b> .....	7
Christoph Marksches	
<b>Stellungnahme zu den EU Ethics Guidelines for Trustworthy AI der High-Level Expert Group on Artificial Intelligence</b> .....	13
Fruzsina Molnár-Gábor, Timo Rademacher und Frauke Rostalski	
<b>Eine vorausschauende Politik für Europa – Ergebnisse des Workshops <i>Towards European Anticipatory Governance for Artificial Intelligence</i></b> .....	25
Isabella Hermann und Georgios Kolliarakis	





## EINFÜHRUNG

Zu den originären Aufgaben der Berlin-Brandenburgischen Akademie der Wissenschaften gehört die Gesellschafts- und Politikberatung. Solche Beratung geschieht vor allem durch Interdisziplinäre Arbeitsgruppen, ein Format, bei dem Mitglieder der Akademie und weitere entsprechend ausgewiesene Wissenschaftlerinnen und Wissenschaftler für eine gewisse Zeit zu einem Thema von entsprechender Bedeutung zusammenarbeiten. Zu den Interdisziplinären Arbeitsgruppen der Akademie gehört nun seit vergangenem Jahr eine Gruppe, die sich mit dem Thema *Verantwortung: Maschinelles Lernen und Künstliche Intelligenz* beschäftigt.<sup>1</sup> Ihr Spezifikum ist eine erfreulich große Zahl von Mitgliedern der Jungen Akademie.

Auf den ersten Blick scheint es nun aber so, als ob zum Themenfeld Künstliche Intelligenz und Maschinelles Lernen kaum mehr Bedarf an zusätzlicher Beratung durch die Hauptstadtakademie besteht. Nahezu täglich erscheinen in überregionalen Tageszeitungen kluge Übersichtsartikel und pointierte Stellungnahmen; immer wieder werden Tagungen, Konferenzen und Symposien veranstaltet und es stapeln sich einschlägige Publikationen. Der Deutsche Bundestag hat auf Antrag der Fraktionen von CDU/CSU, SPD, FDP und DIE LINKE vom 26. Juni 2018 (und also in breitem interfraktionellem Konsens) eine Enquete-Kommission zum Thema „Künstliche Intelligenz – Gesellschaftliche Verantwortung und wirtschaftliche, soziale und ökologische Potenziale“ eingerichtet<sup>2</sup> und verschiedene öffentliche und private Institutionen forschen ebenfalls auf diesem Gebiet. Eine für die Zwecke der genannten Enquete-Kommission erstellte Auswahlbibliographie umfasste – Redaktionsschluss war am 30. September 2018 – hundertfünfzig Seiten, und es ist schwer, angesichts der täglichen anwachsenden Menge an einschlägigen Ereignissen, Stellungnahmen und Veröffentlichungen den Überblick zu behalten. Mit der Künstlichen Intelligenz und dem Maschinellen Lernen beschäftigen sich derzeit aber nicht nur viele wissenschaftliche Einrichtungen, sondern vor allem auch Firmen, die in dem schnell expandierenden Gebiet entwickeln und produzieren. Insbesondere in dem Sektor der Medizintechnik, der Flugzeugtechnik und Automobilindustrie, aber auch in der Militär- und Sicherheitstechnik und

1 <http://www.bbaw.de/forschung/verantwort-maschinelles-lernen-und-kuenstliche-intelligenz> (letzter Zugriff am 12. Januar 2020).

2 <http://dip21.bundestag.de/dip21/btd/19/029/1902978.pdf>; vgl. auch die Homepage der Enquete-Kommission; [https://www.bundestag.de/ausschuesse/weitere\\_gremien/enquete\\_ki](https://www.bundestag.de/ausschuesse/weitere_gremien/enquete_ki) (letzter Zugriff am 12. Januar 2020).

selbst im Alltag von Bestellvorgängen bei großen Internetanbietern ist eine starke Entwicklungsdynamik zu beobachten.

Wer trägt die Verantwortung, wenn Künstliche Intelligenz und Maschinen Aufgaben übernehmen, die bislang dem Menschen vorbehalten waren? Diese Leitfrage hat sich die Interdisziplinäre Arbeitsgruppe der Berlin-Brandenburgischen Akademie der Wissenschaften gestellt, weil sie der Ansicht war, dass in den nahezu unübersehbar vielen Initiativen auf dem Gebiet diese Frage kaum, in jedem Fall zu wenig bedacht wird. Dabei kann sie sich für den Begriff der „Verantwortung“ auf gute Vorarbeiten aus der Akademie stützen.<sup>3</sup> Es geht also nicht allgemein um *die* Bedeutung von Ethik auf dem Feld von Künstlicher Intelligenz und Maschinellen Lernen, die vielfältig bedacht wird – so gibt es im Cluster „Science of Intelligence“ des Berliner Universitätsverbundes ein eigenes Teilprojekt „Ethics of Intelligence“.<sup>4</sup> Vielmehr geht es um ein spezifisches Problem, dessen Lösung verschiedene Disziplinen und neben der Philosophie beispielsweise auch die Rechts- und Technikwissenschaften betrifft. Wenn nicht alles täuscht, besteht dringender Bedarf für entsprechende Reflexion schon deswegen, weil die Politik hierzulande hinterherhinkt. Denn obwohl eigentlich vorgesehen war, dass im Laufe der Legislaturperiode für die Zulassung von entsprechenden Produkten und der Software, die für sie verwendet wird, ein standardisiertes Zulassungsverfahren entwickelt werden soll, wie es sich beispielsweise für Arzneimittel bewährt hat,<sup>5</sup> hält man jetzt im zuständigen Ministerium das Projekt in dieser Zeit nicht mehr für realisierbar. Damit wird die Frage, wer an welcher Stelle für was Verantwortung trägt, nur noch drängender und muss selbstverständlich auch im Zusammenhang gesehen werden.

Wenn im Zusammenhang mit Künstlicher Intelligenz und Maschinellen Lernen ethische Fragen diskutiert werden, oszilliert die öffentliche Debatte zwischen

3 Vgl. beispielsweise Julian Nida-Rümelin, Verantwortung, RUB 18829, Stuttgart 2011 sowie Weyma Lübke, Verantwortung in komplexen kulturellen Prozessen, Freiburg 1998. Über „Verantwortung“ wurde zudem in der BBAW schon in der IAG „Humanprojekt“ von 2006 bis 2011 länger nachgedacht: <http://humanprojekt.bbaw.de/Ueberblick> (letzter Zugriff am 12. Januar 2020).

4 <https://www.scienceofintelligence.de/research/projects/> (letzter Zugriff am 12. Januar 2020).

5 Vgl. die entsprechenden Formulierungen im Koalitionsvertrag: „Wir wollen zum Schutz der Verbraucherinnen und Verbraucher Algorithmen- und KI-basierte Entscheidungen, Dienstleistungen und Produkte überprüfbar machen, insbesondere im Hinblick auf mögliche unzulässige Diskriminierungen, Benachteiligungen und Betrügereien. Wir werden Mechanismen entwickeln, um bei bedenklichen Entwicklungen tätig werden zu können“ (S. 135, Z. 6354-6358; [https://www.cdu.de/system/tdf/media/dokumente/koalitionsvertrag\\_2018.pdf?file=1](https://www.cdu.de/system/tdf/media/dokumente/koalitionsvertrag_2018.pdf?file=1); letzter Zugriff am 12. Januar 2020).

apokalyptischen Szenarien und quasi-messianischen Heilsprophetien.<sup>6</sup> „Difficult, dirty, dull or dangerous work“ lautet ein Versuch, die gegenwärtigen Einsatzfelder der Künstlichen Intelligenz umfassend zu beschreiben, und wer wollte sich nicht freuen, wenn Arbeiten, die niemand tun will, oder solche, die nur schlecht erledigt werden, deutlich besser bewältigt werden können? In diesem Augenblick melden sich aber nicht nur die, die das goldene Zeitalter der Künstlichen Intelligenz mit bunten Farben auszumalen verstehen. Es melden sich auch die Apokalyptiker und warnen beispielsweise vor den Gefahren, die durch das automatisierte Automobil für Leib und Leben entstehen, weil beim Programmieren irgendein wichtiges Detail übersehen wurde und Menschen zu Schaden kommen, oder porträtieren einem in schrecklichen Farben das entmenschte Pflegeheim, in dem nur noch die Roboter tun, was sonst niemand mehr tun will.

Vereinfachungen unterlaufen allerdings nicht nur Laien. Auf einer Homepage des berühmten MIT – Massachusetts Institute of Technology findet sich eine „Moral Machine“, die erfassen will, „wie Menschen zu moralischen Entscheidungen stehen, die von intelligenten Maschinen, wie z.B. selbstfahrenden Autos, getroffen werden“.<sup>7</sup> Einmal abgesehen von der Frage, warum immer ein führerloses Auto *das* Beispiel für künstliche Intelligenz und Maschinelles Lernen sein muss (und beispielsweise nicht eine automatisierte Pumpe, die Insulin für einen individuellen Körper dosiert und zuführt), wundert man sich doch, warum der von Fachleuten formulierte einleitende Kurzttext nicht darüber orientiert, dass *Menschen* mittels Algorithmen die moralischen Entscheidungen programmiert haben, die dann scheinbar von Maschinen getroffen werden. Die „Moral Machine“ bilanziert, wie sich Menschen zu moralischen Entscheidungen stellen, die andere Menschen normiert haben und einer Technologie zugrunde legten. Noch präziser geht es bei dem Projekt darum, wie Menschen in moralischen Dilemmata entscheiden und wie andere das beurteilen. Aber man muss ja nur an das Dilemma denken, das in „Terror – Ihr Urteil“ sowohl in Form eines Theaterstücks von Ferdinand von Schirach als auch in Form eines Fernsehfilms von Lars Kraume öffentlichkeitswirksam auf die Bühne und ins Fernsehen gebracht wurde, um zu erkennen, dass

6 Die folgenden Ausführungen beruhen auf: Christoph Marksches, Warum sollte man einem Computer vertrauen? Die „Moral der Maschine“ und ihre ethischen Folgen: Die Richtlinien für Künstliche Intelligenz der Europäischen Union lassen viele Fragen offen, in: Frankfurter Allgemeine Zeitung Nr. 46, 23.02.2019, S. 11 und zitieren auch einige Wendungen aus: ders., Wer trägt eigentlich die Verantwortung? Die interdisziplinäre Arbeitsgruppe *Verantwortung: Maschinelles Lernen und Künstliche Intelligenz*, in: Jahresmagazin 2020 der Berlin-Brandenburgischen Akademie der Wissenschaften, hg. v. Martin Grötschel, Berlin 2017, S. 32–37.

7 Erläuterungen auf der deutschsprachigen Homepage des Projektes: <http://moralmachine.mit.edu/hl/de> (letzter Zugriff am 12. Januar 2020).

die Programmierung eines selbstfahrenden Autos vom Programmierer ähnliche Überlegungsgänge erfordert wie bei Zuschauern, die entscheiden sollen, ob ein von einem Terroristen entführtes Passagierflugzeug mit 164 Menschen auf dem Weg zum gezielten Absturz in die Münchener Allianz-Arena mit einer Rakete abgeschossen werden darf. Dilemmata zu studieren, kann dabei helfen, die Kompetenz beim Fällen moralischer Urteile zu schärfen.<sup>8</sup> Das ersetzt aber keine ethische Bildung und schon gar keine rechtlich normierten Verfahren. Ethische Bildung benötigen diejenigen, die Programme für Künstliche Intelligenz schreiben und die, die sich entscheiden, sie einzusetzen. Und rechtlich normierte Verfahren sind notwendig, wenn man Programme daraufhin überprüfen will, ob sie bestimmten ethischen Normen und rechtlichen Standards entsprechen.

Nun gelten in Europa an vielen Stellen auf den verschiedenen Ebenen sehr unterschiedliche Normen wie Standards. Deswegen ist es im Prinzip gut, dass die Europäische Kommission eine „High Level Expert Group on AI“ eingesetzt hat, die „ethische Richtlinien für vertrauenswürdige Künstliche Intelligenz“ erarbeiten soll, auch wenn das ganze Verfahren unter einem kaum akzeptablen Zeitdruck stand. Eine vorläufige Version aus dem Jahr 2018 wurde nach einer relativ knapp bemessenen Frist zur öffentlichen Diskussion im Jahre 2019 durch eine endgültigere Version ersetzt. Um diese Richtlinien geht es in der ersten Veröffentlichung der Interdisziplinären Arbeitsgruppe der Akademie, und sie wird auf den folgenden Seiten von Expertinnen und Experten ausführlich diskutiert.

Nur so viel als Vorbemerkung und Einleitung in das europäische Dokument: Die Idee, diese Richtlinien als eine Art lebendiges Dokument zu verstehen, das in Zukunft ständig „upgedated werden muss“, ist gut gemeint, hilft allerdings nicht.<sup>9</sup> Denn hier wird der textliche Status normativer Regelungen, die eine fest verabredete Stabilität brauchen, mit der gemeinsamen Textproduktion von Information beispielsweise bei einem Wikipedia-Artikel durch die Community verwechselt. Wer soll für die Qualitätskontrolle zuständig sein? Zu dieser Frage findet sich im Papier der „High Level Expert Group“ leider keine Idee.

8 Ergebnisse eines Votings bei Sendungen im deutschsprachigen Raum dokumentiert die Homepage des Dokumentarspiels: <https://www.daserste.de/unterhaltung/film/terror-ihr-urteil/voting/index.html> (letzter Zugriff am 12. Januar 2020).

9 Hochrangige Expertengruppe für künstliche Intelligenz, Ethik-Leitlinien für eine vertrauenswürdige Künstliche Intelligenz (Fassung vom April 2019), S. 4 (<https://ec.europa.eu/digital-single-market/en/news/ethics-guidelines-trustworthy-ai>; letzter Zugriff von 12. Januar 2020).

Außerdem versäumt der Text eine grundlegende Unterscheidung. Das zeigt sich schon im Titel der Richtlinien: Vertrauen kann ein Mensch jedenfalls nach dem in der Philosophie, Rechtswissenschaft und Theologie üblichen Sprachgebrauch,<sup>10</sup> zu einem anderen Menschen, seiner Person oder seinen Handlungen haben. Dabei basiert die Überzeugung, Vertrauen schenken zu wollen, natürlich auf rationalen Annahmen, stellt aber auch ein gewisses Wagnis dar, denn kein Mensch ist in seinen Handlungen so vorhersagbar wie eine KI-Maschine es allein aufgrund ihres Algorithmus sein sollte. Der Mensch hat beispielsweise Emotionen, die ihn eine einmal gefällte Entscheidung revidieren lassen. So wie ich natürlich keinem medizinischen Präparat vertraue, sondern dem Pharmazeuten, der es entwickelte, und der Ärztin, die es mir verschrieb, vertraue ich auch nicht letztlich der Künstlichen Intelligenz und Maschinellen Lernprozessen. Vertrauenswürdig sind *letztlich* nicht Systeme der KI, sondern die Menschen, die sie programmieren und die, die für rechtlich geordnete Zulassungsverfahren Verantwortung tragen. Von „vertrauenswürdiger Künstlicher Intelligenz“ kann man höchstens im übertragenen Sinne sprechen. Ähnliches gilt übrigens auch für Verantwortung: Natürlich kann, kommt es zu Fehlsteuerung, nicht eine Maschine zur Verantwortung gezogen und bestraft werden, sondern Programmierer (oder die Vorgesetzten bzw. in diversen Situationen auch die Anwender) sind verantwortlich, weil Verantwortung ein rationales Subjekt voraussetzt, das Verantwortung übernehmen kann.

Wird der kategoriale Unterschied zwischen rationalen Subjekten und Maschinen verwischt oder auch nur terminologisch unpräzise bestimmt, lässt sich der Dual von apokalyptischen Horrorszenarien oder pseudo-messianischen Heilsprophetien kaum mehr vermeiden. In Wahrheit gibt es Situationen, in denen ich dankbar dafür bin, dass automatisierte medizinische Diagnostik auf ungeheuer viele Patientendaten zurückgreifen kann, um meinen Befund zu erstellen. Gleichzeitig gibt es aber auch Situationen, in denen ich die Empathie eines medizinisch gebildeten Menschen brauche, um diesen Befund überhaupt zu verstehen und zu verkraften. Diesen kategorialen Unterschied sollte man nicht verwischen oder gar ignorieren. Auch nach dem Verarbeiten von weiteren hunderttausenden Patientendaten wird die Diagnose-App noch keine Empathie zeigen, sondern

10 Diskussionen in der Arbeitsgruppe haben immer ersichtlicher werden lassen, dass es hier deutliche Unterschiede gibt: In den *Technikwissenschaften* wird beispielsweise durchaus abgekürzt von „vertrauenswürdiger Technik“ geredet, wobei am Ende einer längeren Kausalitätskette natürlich auch wieder Menschen stehen, die als rational agierende Wesen letztllicher Grund des Vertrauens in von ihnen entworfene, hergestellte und zum öffentlichen Gebrauch zugelassene Technik samt Anwendungen sind. Ein Nachdenken über diese Zusammenhänge, die sich nicht auf unterschiedliche terminologische Konventionen beschränken, lässt das Dokument in seinen beiden Versionen von 2018 und 2019 eher nicht erkennen.

sie höchstens auf algorithmischer Basis simulieren. Wir brauchen standardisierte, gesetzlich geregelte Zulassungsformen für Systeme der Künstlichen Intelligenz. Für solche rechtlich geregelten Prozesse sind natürlich bestimmte ethische Normen notwendig. Gerade angesichts der Krise des europäischen Projektes wäre es wunderbar, wenn eine entsprechende Verständigung auf einen europäischen Wertekonsens gelänge. Diesen Konsens könnte man dann in der akademischen Ausbildung vermitteln und in die öffentliche Diskussion zu bringen versuchen. Dazu ist allerdings zunächst präzise Arbeit an den Begriffen notwendig und klare Unterscheidungen.

So lag es für die Interdisziplinäre Arbeitsgruppe der Berlin-Brandenburgischen Akademie der Wissenschaften nahe, die Guidelines von Anfang an einer kritischen Debatte zu unterziehen. Diese Debatte konnte im September 2019 direkt mit Mitgliedern der High Level Expert Group im Rahmen eines gemeinsam mit der Deutschen Gesellschaft für Auswärtige Politik (DGAP) organisierten Workshops zum Thema einer „vorausschauende Politik für Europa“ fortgesetzt werden. Das hier von der IAG vorgelegte erste Heft einer Schriftenreihe unter dem Titel „#VerantwortungKI – Künstliche Intelligenz und gesellschaftliche Folgen“ dokumentiert mit einer Stellungnahme der IAG-Mitglieder Fruzsina Molnár-Gábor, Timo Rademacher und Frauke Rostalski Teile der Diskussion um die europäischen Ethikrichtlinien aus einer rechtlichen Perspektive und liefert eine deutsche Version der Ergebnisse des gemeinsamen Workshops mit der DGAP. Es ist mir ein Bedürfnis, neben den Mitgliedern unserer Interdisziplinären Arbeitsgruppe, die hier ihre Beiträge aus Sitzungen zur Verfügung gestellt haben, auch denen zu danken, die diese vor allem aus rechtswissenschaftlicher Perspektive formulierten Stellungnahmen aus der Sicht anderer Disziplinen lebhaft diskutiert haben, ebenso auch meinen Mit-Sprechern Max Löhning und Jens Krause, aber auch Günter Stock. Christiane Wopen war so freundlich, uns den Text der Richtlinien und Hintergründe seines Zustandekommens ausführlich zu erläutern, auch dafür gebührt ihr unser herzlicher Dank. Ohne das große Engagement von Isabella Hermann hätte es allerdings gar nichts gegeben, das man hätte mit einem solchen Geleitwort versehen können. Ihr schulden wir daher alle auch besonders großen Dank!

Christoph Marksches

*Sprecher der interdisziplinären Arbeitsgruppe „Verantwortung: Maschinelles Lernen und Künstliche Intelligenz“ der BBAW*

## **STELLUNGNAHME ZU DEN EU ETHICS GUIDELINES FOR TRUSTWORTHY AI DER HIGH-LEVEL EXPERT GROUP ON ARTIFICIAL INTELLIGENCE**

### **ABSTRACT**

Die *Ethics Guidelines for trustworthy AI* der *High Level Expert Group on Artificial Intelligence* (AI HLEG) verlangen von einer KI-Anwendung, rechtmäßig, ethisch und robust zu sein, um das Prädikat „vertrauenswürdig“ zu erlangen. Eine rechtliche Analyse der Guidelines zeigt allerdings: Die rechtlichen Voraussetzungen für den vertrauenswürdigen Einsatz von KI werden mit Blick auf bereits existierendes Recht in den Guidelines bewusst ausgelassen. Es soll daher im Rahmen der folgenden Stellungnahme eine Methodik aufgezeigt werden, mithilfe derer die bisher fehlenden Leitlinien für eine rechtmäßige KI-Anwendung erarbeitet werden können. Die Methodik zeichnet sich durch ein kombiniertes deduktives und induktives Vorgehen aus. Es sollen, ausgehend von repräsentativen Fallstudien zu Anwendungsoptionen von KI in einem spezifischen Sektor, model rules aus einfachem sowie dem EU-Primärrecht entwickelt und u. a. ethisch und technisch beleuchtet werden. Die Methode ist anschließend auf verschiedene Sektoren auszuweiten. Durch einen Abgleich der Ergebnisse können sektorübergreifende Regeln für rechtmäßige KI-Anwendungen entwickelt werden, die die Richtlinien der AI HLEG sinnvoll ergänzen können. Aufgrund des induktiven Vorgehens werden die Besonderheiten der einzelnen Sektoren dabei hinreichend berücksichtigt. So können neben ethischen auch rechtliche Leitlinien zu einem vertrauensvollen Umgang mit KI-Anwendungen beitragen.

## EINFÜHRUNG

Die von der Europäischen Kommission eingesetzte AI HLEG hat sich mit ihren im April 2019 veröffentlichten *Ethics Guidelines for trustworthy AI* (im folgenden Guidelines) das Ziel gesetzt, „guidance for AI applications in general“<sup>1</sup> zu liefern. Bei deren Befolgung soll sich eine KI-Anwendung das Prädikat „trustworthy“ verdienen können. Um „trustworthy“ zu sein, so die zentrale Leitlinie auf höchster Abstraktionsebene, muss eine KI-Anwendung danach drei Voraussetzungen erfüllen: Sie muss rechtmäßig, ethisch und robust sein.

Die Autorinnen und der Autor wurden gebeten, die Guidelines einschließlich der drei Monate später nachgereichten Policy and Investment Recommendations for Trustworthy AI (im folgenden Policy Paper) aus einem *rechtlichen* Blickwinkel heraus zu analysieren. Nimmt man einen solchen ein, dann fällt als erstes auf: Rechtmäßige KI wird bewusst ausgeklammert, obwohl sie zunächst als eine von drei tragenden Säulen der angestrebten „trustworthy AI“ bezeichnet wird. Die AI HLEG konzentriert sich auf „4 ethical principles“ und deren „robuste“ Umsetzung durch „7 requirements“, an denen sich eine „trustworthy“ KI-Anwendung – jenseits der ausgesparten Rechtmäßigkeit – messen lassen müsse.

Diese methodische Prämisse der AI HLEG verwundert, denn sie kann (und will?) das Recht tatsächlich doch nicht ganz ausklammern. Die ethischen Prinzipien werden ihrerseits vollständig aus einer Synthese der existierenden Grundrechte der EU, also aus geltendem Recht entwickelt.<sup>2</sup> Das Recht ist damit Maß und Grenze der vorgeschlagenen „ethical principles“. Legt man neben die Guidelines noch die darauf aufbauenden, konkreteren Empfehlungen des Policy Papers, die die AI HLEG im Juni 2019 nachgereicht hat, nimmt die Verwirrung weiter zu. Dort heißt es: „many of the principles set out in the Guidelines reflect existing EU law“.<sup>3</sup> Geht es also doch um eine „lawful AI“ in einem lediglich reduzierten Zustand, dem die harte rechtliche Vollstreckbarkeit durch staatliche Gerichte fehlt, der also nur durch moralischen Appell und durch Bezugnahme auf die Grundrechte wirken soll? Eine der Empfehlungen der AI HLEG lautet dann aber, die EU möge die existierenden Rechtsnormen darauf überprüfen, ob sie die ethischen Prinzipien „fördern

1 High-Level Expert Group on Artificial Intelligence (2019): Ethics Guidelines for trustworthy AI, Brüssel, 8. April 2019, S. 5 [nachfolgend Guidelines].

2 Guidelines, S. 9 f.

3 High-Level Expert Group on Artificial Intelligence (2019): Policy and Investment Recommendations for trustworthy AI, Brüssel, 26 Juni 2019, S.37 [nachfolgend Policy Paper].



und sichern“,<sup>4</sup> und zwar „beyond voluntary guidance“. <sup>5</sup> Nimmt man dann noch hinzu, dass selbstverständlich alle Rechtsnormen der EU mit den EU-Grundrechten übereinstimmen müssen, aus denen die Guidelines entwickelt wurden – dann schließt sich der Kreis. Es scheint so, als wirke zunächst das existierende Recht in die Guidelines hinein, die dann wiederum zum Maßstab für das geltende Recht werden; die genauen Konturen bleiben dabei aber unklar. Welche Rolle das Recht und welche die Ethik hat, und wie ihre Rollen einander ergänzen oder ineinandergreifen und ob es Konfliktfälle gibt und wie diese zu behandeln sind, bleibt offen.

Methodisch ist fraglich, ob das bewusst fragmentarische Vorgehen der AI HLEG nicht mehr Unklarheit schafft, als es beseitigt. So schreibt nun auch die AI HLEG selbst in ihrem Policy Paper, dass die „guidance [...] on how to ensure the implementation of ethical and robust AI [is] already captured by a number of legal obligations, and as legal uncertainty exists regarding the extent to which such legal obligations are already in place, a guidance document on this first component [= lawful AI] would be welcomed.“<sup>6</sup>

Neben die rechtliche Säule, die nach Auffassung der AI HLEG schon unklar ist, ist mit den Guidelines nun eine ethische Säule gestellt, von der nicht ganz klar ist, bis zu welchem Grade sie aus dem geltenden Recht abgeleitet worden ist, oder davon abweicht oder darüber hinausgeht, und die zugleich den Anspruch erhebt, ihrerseits zum Maßstab für das Recht geworden zu sein.<sup>7</sup>

Es muss nun darum gehen, die zusätzlichen Unsicherheiten, die dieses bislang changierende Neben-<sup>8</sup> und Miteinander<sup>9</sup> verursachen könnte, rasch zu klären, bevor sie zu Verunsicherung der KI-Stakeholder führen kann. Die Aufforderung der AI HLEG, ein guidance document zu den bereits vorhandenen unionsrechtlichen

4 Policy Paper, Ziffer 27.1(3).

5 Policy Paper, S. 37.

6 Policy Paper, Ziffer 30.6.

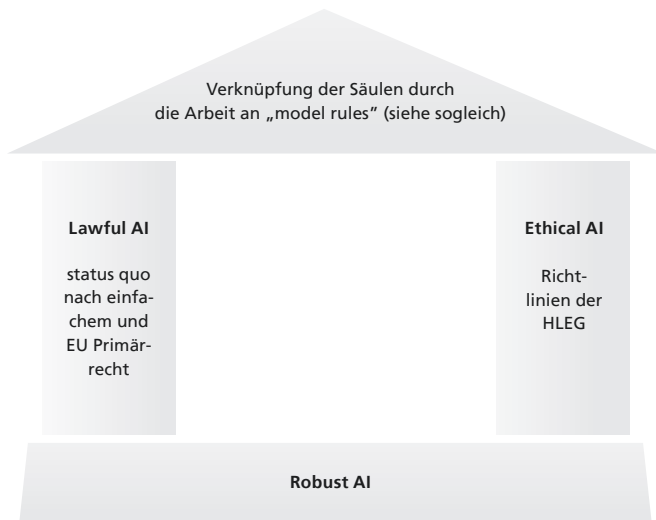
7 Policy Paper, Ziffer 27.1(33).

8 Für ein explizites Nebeneinander siehe etwa S. 6 der Guidelines, wonach die HLEG „recall[s] that it is the duty of any natural or legal person to comply with laws.“

9 Für ein Miteinander z. B. Ziffer 30.6 des Policy Papers, wonach die HLEG davon auszugehen scheint, dass die „guidance for stakeholders on how to ensure compliance with [lawful AI], for each of the seven requirements“ entwickelt werden müsse. Diese „seven requirements“ sollen also offenbar auch direkt für „lawful AI“ gelten, wobei die AI HLEG deren Entwicklung allein auf die ethischen Prinzipien gestützt hat (vgl. S. 14 der Guidelines), ohne sich zu deren auch rechtlicher Ableitbarkeit verhalten zu haben.

Rahmenbedingungen von KI zu erstellen,<sup>10</sup> ist daher nicht nur zu begrüßen, sondern sollte unseres Erachtens unbedingt der nächste Schritt sein.

Im Folgenden wollen wir eine Methodik skizzieren, mittels derer die bislang fehlende Säule der „rechtmäßigen KI“ errichtet werden könnte. Ethische Überlegungen können bei der gewählten Methode unmittelbar mit dem Recht verknüpft werden, sodass am Ende nicht nur zwei tragende Säulen, sondern auch ein gemeinsames, beide verbindendes Dach entsteht. Die prozedurale Komponente im Konzept der AI HLEG, die Robustheit,<sup>11</sup> wäre dann, um im Bild zu bleiben, das gemeinsame Fundament, das den beiden Säulen den *technisch* und *gesellschaftlich* sicheren Stand vermittelt:<sup>12</sup>



10 So das Policy Paper, Ziffer 30.6.

11 Vgl. Guidelines, S. 7: „[Robustness] is needed both from a technical perspective (ensuring the system’s technical robustness as appropriate in a given context, such as the application domain of life cycle phase), and from a social perspective (in due consideration of the context and environment in which the system operates).“

12 Wir glauben, damit ein Bild zu liefern, dass die Verknüpfungen nachvollziehbar aufzeigt. Vgl. demgegenüber auch das unserer Sicht nicht abschließend gelungene Schaubild der HLEG selbst (S. 8 der Guidelines), wo die „three components“ der Trustworthiness (rechtmäßig, ethisch, robust) scheinbar unverbunden neben den vier ethischen Prinzipien und sieben – für „lawful AI“ und „ethical AI“ formulierten – Requirements stehen.

## METHODISCHE ÜBERLEGUNGEN

Die AI HLEG hat eine deduktive Methode gewählt. Sie geht von vier allgemeinen Prinzipien aus, welche die – freilich selbst schon recht abstrakten – Grundrechte der EU und die zu beachtenden internationalen Menschenrechtsakte nochmals abstrahieren,<sup>13</sup> um auf diesem Weg gleichwohl zu einem Ergebnis zu gelangen, das einen hohen Konkretheitsanspruch erhebt: „guidance, on how such [ethical] principles can be *operationalised*“.<sup>14</sup> Ob der hohe Anspruch schon mit den Guidelines und dem Policy Paper erfüllt wird, erscheint uns zumindest zweifelhaft.<sup>15</sup> Das illustriert ein Blick in die „Examples of Opportunities and Critical Concerns Raised by AI“.<sup>16</sup> Die Begriffe, die die dort aufgeführten Beispiele aus den Bereichen Tracking, Covert AI Systems, Citizen Scoring und Autonomous Weapons beschreiben („proportionate use“, „targeted surveillance“, „meaningful consent“, „careful assessment“), sind Platzhalter für noch zu leistende Wertungen und vermitteln noch keine Ergebnisse einer Abwägung. Das entwertet für sich genommen nicht die Herangehensweise der AI HLEG. Aus dem Allgemeinen lässt sich das Konkrete ableiten, weshalb das Abstrakte besonders wichtig ist. Grundsätzlich ist das Abstrakte unerlässlich, um sachlich gleich gelagerte, konkrete Fälle auch gleich zu behandeln. Dabei ist es auch positiv hervorzuheben, wenn verstärkt mit der Sprache von Technikern und Informatikerinnen gearbeitet wird. Gerade die Begriffe der Sicherheit und der Robustheit werden in der Informatik zur Beschreibung von KI-Systemen verwendet. Stärker auf die Sprache der Technologie abzustellen, könnte dazu führen, dass Technologiebegriffe wie Robustheit eine normative Konnotation erhalten: Ihre Qualifizierung als Sammelbegriffe für ethische Wertungen könnte wiederum zu einer größeren Sensibilisierung für die normativen Implikationen der nur – vermeintlich neutralen – Technologie auf Seiten der Technikwissenschaften führen.

Diese Herangehensweise hat jedoch auch Nachteile. Das lässt sich am besten an einem Beispiel illustrieren: So stellen die Guidelines häufig auf Erklärbarkeit und Transparenz als zentrale Voraussetzungen einer „vertrauenswürdigen KI“ ab. Man sollte dabei aber beachten, dass Erklärbarkeit und Transparenz nicht für jede

13 Guidelines, S. 9 ff.: „Among the comprehensive set of indivisible rights set out in international human rights law, the EU Treaties and the EU Charter, the below families of fundamental rights are particularly apt to cover AI systems.“

14 Guidelines, S. 2. Hervorhebung hier.

15 Siehe auch die Kritik des AI HLEG-Mitglieds *Thomas Metzinger* an der – seines Erachtens – zu geringen normativen Substanz der Guidelines, abgedruckt im Tagesspiegel v. 8.4.2019, unter dem Titel „Nehmt der Industrie die Ethik weg!“

16 Guidelines, S. 32 ff.

KI-Anwendung realisierbar sind, zumindest nicht vollständig, und darüber hinaus auch nicht immer erforderlich sind (z.B. bei bestimmten Fällen des Marketings). Die wichtige Frage in diesem Kontext ist daher, *in welchem Maße* wir im konkreten Einzelfall Transparenz und/oder Erklärbarkeit verlangen wollen oder müssen.

Für die rechtliche Säule schlagen wir daher eine Methodik vor, die Deduktion und Induktion kombiniert und ausgehend vom Beispiel arbeitet. So ist am konkreten Beispiel zu prüfen, welches einfache Recht den Fall regelt, und welche höherrangigen Rechtssätze – insbesondere welche Grundrechte – ihrerseits Anforderungen an das einfache Recht stellen. Diese Methode lässt sich selbstverständlich nicht an allen denkbaren Anwendungskonstellationen von KI wiederholen, aber doch an einer hinreichend großen Zahl relevanter Fälle, sodass – induktiv – aus dem Konkreten auch Rückschlüsse für das Allgemeine gezogen werden können. Idealerweise strebt ein solches Vorgehen danach, sektorspezifische oder sogar allgemeine, sektorübergreifende Mustervorschriften (“model rules”) zu liefern. Diese Regeln eignen sich zur Diskussion der Stakeholder von KI und eventuell sogar zur Übernahme durch den Gesetzgeber. Ein gutes Beispiel für ein solches Vorgehen lieferte jüngst die ReNEUAL-Gruppe, die model rules für ein europäisches Verwaltungsverfahrensgesetz erstellt hat.<sup>17</sup> Der Vorteil des Arbeitens mit model rules, die dem juristischen Sprachgebrauch verpflichtet sind, ist es, dass sie zu einer sehr präzisen Darstellung zwingen (inkl. präziser Begriffsarbeit), und damit gut diskutierbar und kritisierbar sind.

Eine solche Arbeit ist sinnvoll nur auf Basis vorangegangener ethischer Diskussion zu leisten, wobei die model rules stets Raum für Anpassungen lassen müssen, sollten künftige gesellschaftliche Auseinandersetzungen um den Umgang mit KI zu abweichenden Ergebnissen gelangen. Aus rechtswissenschaftlicher Sicht lassen sich folgende Ansatzpunkte ausmachen, die für ein Zusammenspiel von Recht und Ethik im KI-Bereich in Betracht kommen: Zu denken ist dabei zunächst an das Potential ethischer Prinzipien, die Entstehung, Auslegung, Anwendung und Weiterentwicklung rechtlicher Normen zu begleiten und zu beeinflussen.<sup>18</sup> Insofern sind sie geeignet, sowohl Regulierungsakteuren als auch KI-Anwendern Handlungskorridore aufzuzeigen. Ethische Prinzipien können somit die Rolle eines sprachlichen Vermittlers übernehmen: Wo das Recht in seiner ggf. allzu

17 Siehe zur Methodik, dort für das europäische Verwaltungsverfahrenrecht, *Jens-Peter Schneider/Herwig C.H. Hofmann/Jacques Ziller*, ReNEUAL – Musterentwurf für ein EU-Verwaltungsverfahrenrecht, 2015, S. 2 f.

18 Siehe zum Verständnis der HLEG von Ethik und ihren Funktionen bes. S. 9 der Guidelines, sowie Policy Paper, Ziffer 27.1(3).

formalisierten Sprache den Bezug zu seinem Adressaten zu verlieren droht, können ethische Prinzipien die Kommunikation über die moralische Grundideen wieder eröffnen helfen, die hinter dem Recht stehen oder stehen sollten. Eine solche Kommunikation kann das Recht entlasten ebenso wie neue Regelungen anstoßen oder bestehende bestätigen. Auch kann sie zur Vergewisserung und Verteidigung der europäischen Wertestandards im Austausch mit anderen Rechtssystemen beitragen. Bei der Erfüllung dieser Aufgabe ist allerdings insoweit Vorsicht geboten, als dass bewusste Differenzierungen des Rechts nicht übersehen werden dürfen (vgl. unten, Meilensteine 2 und 3).

Ein wesentliches Element der Vertrauenswürdigkeit von KI ist schließlich in der Gewährleistung von Rechtssicherheit zu sehen (verstanden als Stabilisierung bestimmter Erwartungen gegenüber anderen Akteuren und als Basis von Vertrauen in der Gesellschaft). Es gilt, normative Parallelordnungen zu vermeiden sowie die Erfüllung und Ergebnisse dieses Kommunikationsauftrags weitestgehend zu legitimieren (de facto und de jure) und regulativ anzubinden. Insofern ist es richtig und wichtig, dass die AI HLEG bei der Entwicklung ihrer Aussagen zur "ethischen KI" das Recht abwechselnd als Quelle und dann wieder als Zielobjekt der eigenen Überlegungen mit heranzieht. Da die ethischen Prinzipien bei dieser Methode zwangsläufig über die bloße Vergewisserung der Grundrechte hinausgehen, ist es besonders wichtig – wie oben bereits dargelegt – darauf zu achten, dass die *bewussten* Differenzierungen, die das Recht anstellt, zumindest nicht unreflektiert übersehen werden.

Daraus ergeben sich folgende Arbeitsschritte:

<b>Meilenstein 1</b>	Auswahl einer gesellschaftlich relevanten Sektorenstudie mit repräsentativen Fallstudien
<b>Meilenstein 2</b>	Anwendbares einfaches Recht identifizieren und Lösung der Fallstudien skizzieren
<b>Meilenstein 3</b>	Anwendbares EU-Primärrecht identifizieren, Lösung skizzieren und mit Meilenstein 2 abgleichen
<b>Deliverable 1</b>	Sektorale model rules auf Ebene des einfachen Rechts entwickeln
<b>Weitere Meilensteine</b>	Wiederholung der Schritte 1–3 für andere ausgewählte Sektoren
<b>Deliverable 2</b>	Allgemein für KI-Anwendungen geltende model rules aus dem bestehenden Recht und den sektoralen model rules synthetisieren und zur Diskussion stellen

Im Detail:

### **Meilenstein 1: Entwurf einer Sektorenstudie anhand von repräsentativen Fallstudien**

Als erster Schritt wird ein Sektor identifiziert, in dem die Anwendung von KI gesellschaftlich kontrovers diskutiert wird und/oder wirtschaftlich bedeutend ist. Anhand von repräsentativen Fallstudien sollen die wesentlichen Anwendungsoptionen von KI in diesem Sektor demonstriert werden.

Als Beispiel kann an den Bildungsbereich und daraus, als repräsentative Fallstudie, an das Profiling von Kindern gedacht werden. Auch die AI HLEG bestimmt "Hochwertige Bildung und digitalen Wandel" als einen Sektor, in dem die Anwendung von KI große Chancen bietet.<sup>19</sup> Es verwundert daher nicht, dass die AI HLEG das Profiling von Kindern auch nicht prinzipiell für unzulässig hält, sondern (nur) anregt zu überlegen, ob Kinder zu einem bestimmten Zeitpunkt einen Anspruch auf eine „clean data slate“ haben könnten.<sup>20</sup> Auf den ersten Blick scheint das nicht nur ethisch, sondern auch rechtlich – wegen der besonderen Schutzbedürftigkeit von Kindern, vgl. Art. 24 EU-Grundrechtecharta – ein eher permissiver Umgang mit KI zu sein. „Profiling von Kindern“ stößt wahrscheinlich bei vielen auf starke Ablehnung. Bedenkt man aber, dass eben auch viele KI-Anwendungen im Bildungsbereich „Profiling“ darstellen (vgl. die Definition in Art. 4 Nr. 4 DS-GVO), dann ändert sich das Bild. Zum Beispiel kann es wünschenswert sein, Kinder in der Schule oder auch beim Übergang in ein Studium informationstechnisch zu unterstützen. Auch das könnte eine Form darstellen, die besondere staatliche Verantwortung für das Kindeswohl auszuüben. Entsprechende Überlegungen finden sich nicht nur (freilich in abschreckender Form) in China, sondern auch in den USA und in Kanada, wo sie ethisch, insbesondere hinsichtlich der Auswirkungen auf die Persönlichkeitsentwicklung, deutlich reflektierter erfolgen.<sup>21</sup> KI-Anwendungen im Bildungs- und Hochschulbereich bieten sich daher gerade wegen der ihnen innewohnenden rechtlichen Ambiguität als besonders vielversprechende Sektorenstudie an. Bei der Auswahl repräsentativer Fallstudien sollten an erster Stelle Expertinnen und Experten aus der Bildungsforschung und -praxis, der angewandten Ethik und der Informatik miteinbezogen werden.

19 Guidelines S. 33.

20 Ziffer 28.3 des Policy-Papers; vgl. auch 28.1.

21 Vgl. die Schilderungen bei *Karen Hao*, China has started a grand experiment in AI education, MIT Technology Review v. 2.8.2019, abrufbar unter <https://www.technologyreview.com/s/614057/china-squirrel-has-started-a-grand-experiment-in-ai-education-it-could-reshape-how-the> (letzter Zugriff, 10. Februar 2020).

## **Meilenstein 2: Identifikation und Anwendung des einschlägigen einfachen Rechts**

In einem zweiten Schritt ist zu identifizieren, welches einfache Recht auf die Fallstudien Anwendung findet. Im Bereich von KI-Anwendungen handelt es sich hierbei oftmals um Rechtsnormen zum Schutz der Daten Einzelner sowie von Organisationen. Gleichwohl darf keine vorschnelle Blickverengung auf das allgemeine Datenschutzrecht erfolgen. Bei näherem Hinsehen fällt in aller Regel auf, dass neben dem Datenschutz eine Vielzahl anderer Regelungsbereiche durch die jeweilige KI-Anwendung betroffen sind. Insofern ist bereits auf der Ebene der Identifikation des anwendbaren einfachen Rechts ein umfassender Maßstab anzulegen. Dabei zählt sich Akribie auf dieser Stufe in zumindest zweifacher Hinsicht aus. So gelingt auf diese Weise neben der Herausarbeitung der auf den konkreten Fall anwendbaren Rechtsnormen eine (kritische) Analyse des geltenden einfachen Rechts. Wer nach anwendbaren Vorschriften sucht, findet in aller Regel zugleich auch das, was (bislang) nicht in Gesetzesform gegossen wurde. Finden sich solche „Lücken“ im Gesetz, regt das zur Frage nach den Gründen an. Sofern die Lücke eine bewusste Entscheidung des europäischen Gesetzgebers war (sog. „planmäßige Lücke“), beugt diese Erkenntnis u. U. einer unreflektierten Ausdehnung des Rechts auf bislang nicht geregelte Bereiche vor. Ebenso verhält es sich im Hinblick auf eine unüberlegte Übernahme spezifischer Konzepte der Rechtserweiterung, beispielsweise in Gestalt der Rechtsfigur der „mittelbaren Drittwirkung von Grundrechten“.

## **Meilenstein 3: Identifikation und Anwendung des einschlägigen EU-Primärrechts**

Hier ist zu fragen, welche Vorgaben das auf die Fallstudien anwendbare Primärrecht der EU macht. Primärrecht ist gegenüber dem einfachen (europäischen und auch nationalen) Recht höherrangig, das letztere hat also dem erstgenannten zu entsprechen, ist andernfalls anzupassen oder – soweit möglich – primärrechtskonform auszulegen. Zum Primärrecht gehören insbesondere die Grundrechte der EU, wie sie in rechtsverbindlicher Weise seit 2009 in der EU-Grundrechtecharta niedergelegt sind.<sup>22</sup>

<sup>22</sup> Auch die von einem Konvent erarbeitete Grundrechtecharta lässt sich als model rule bezeichnen, da sie zunächst unverbindlich als Zusammenführung des grundrechtlichen Acquis in Europa gedacht war, um dann erst mit dem Vertrag von Lissabon im Jahr 2009 Verbindlichkeit zu erlangen.

Die Einschätzung der AI HLEG, „that [EU] fundamental rights do not provide comprehensive legal protection in every case“,<sup>23</sup> ist streng genommen nicht falsch (vgl. Art. 51 EU-Grundrechtecharta zu deren begrenztem Anwendungsbereich), vermittelt aus unserer Sicht aber doch einen falschen Eindruck, da die Lücken des Rechts hier zumindest implizit überzeichnet werden. Denn das Feld von KI-Anwendungen, die *nicht* in den Anwendungsbereich des EU-Rechts und damit auch der verbindlichen EU-Grundrechte fallen, ist verschwindend gering; das gilt jedenfalls dann, wenn eine Anwendung personenbezogene Daten verwendet. Unter dieser Voraussetzung sind nämlich die DS-GVO und die JI-Richtlinie (Polizei, Sicherheit und Strafrecht) der EU anwendbar. Beide Regelwerke zusammen erfassen fast jede denkbare private oder staatliche KI-Anwendung – mit Ausnahme nur spezieller Bereichen wie Militär, Geheimdienste und Diplomatie – und unterwerfen sie verbindlichen Rechtsregeln. Damit ist dann auch der gesamte Grundrechtebestand der EU-Grundrechtecharta, und nicht etwa nur das Recht auf Privatheit, Familie und Kommunikation (Art. 7) oder das Grundrecht auf Datenschutz (Art. 8 Grundrechtecharta) anwendbar. Die Arbeit anhand von Fallstudien hat also an dieser Stelle den Vorteil, dass sie dabei hilft, den Anwendungsbereich des verbindlichen Rechts nicht zu unterschätzen bzw. den „Freiraum“ für ethisch motivierte Abwägungen nicht zu überschätzen.

### **Deliverable 1: Entwicklung eines Vorschlags für sektorale model rules auf Ebene des einfachen Rechts**

Das erste Produkt des hier vorgeschlagenen Vorgehens würde in *sektorspezifischen* model rules für den ausgewählten Sektor, zum Beispiel für „KI und Bildung“, liegen. Solche „modellhaften“ Regeln können Anleitung bieten für eine gegebenenfalls vorzunehmende Revision des einfachen, sektorspezifischen Rechts durch den Gesetzgeber.

Dafür ist in einem ersten Schritt eine Synthese vorzunehmen aus dem an den Fallstudien „getesteten“ einfachen Recht (Meilenstein 2) und den Vorgaben, die das Primärrecht (Meilenstein 3) an die einfachrechtliche „Lösung“ stellt. Ergebnis dieser Synthese könnte dann ein kohärenter Vorschlag dafür sein, wie „lawful AI“ für einen bestimmten Sektor unter Berücksichtigung des rechtlichen Status quo aussehen würde.

<sup>23</sup> Guidelines, S. 10, Absatz um Fn. 18.



Da es im Umgang mit neuen Technologien aber nicht nur darum geht, den rechtlichen Status quo zu ermitteln oder gar festzuschreiben, müssen die model rules in einem zweiten Schritt ethisch, technisch, wirtschaftlich usw. hinterfragt werden. Diese Reflexion sollte im Rahmen einer zeitlich nicht zu kurz bemessenen Konsultationsphase mit KI-Stakeholdern erfolgen. Die Stakeholder können etwa nach den Kriterien, die von der AI HLEG hierfür entwickelt wurden, bestimmt werden (z. B. Universitäten, Studentenorganisationen, Wissenschaftsvertreter, NGOs<sup>24</sup>).

#### **Meilensteine 4 ... n: Weitere Sektorenstudien**

Nunmehr sind weitere Sektoren zu identifizieren, die sich neben dem Bildungsbereich für weitere Fallstudien eignen. Die Meilensteine 1–3 bis hin zum Deliverable 1 sind daran zu wiederholen: Das einschlägige einfache Recht sowie das Primärrecht sind zu identifizieren und anzuwenden und anhand ihrer Synthese in der Falllösung und deren kritischer Reflexion in einer Konsultationsphase kann ein Vorschlag für die kohärente Regulierung des jeweiligen Bereichs entwickelt werden. Für diese weiteren Sektorenstudien bieten sich unter anderem die Bereiche an, welche die AI HLEG als Beispiele für die Darstellung der Chancen und Risiken von KI-Anwendungen identifiziert hat: Klimaschutz und nachhaltige Infrastruktur, Gesundheit und Wohlergehen.<sup>25</sup>

#### **Deliverable 2: model rules für ein allgemeines KI-Recht**

Ein Abgleich der Ergebnisse der Sektorenstudien kann schließlich dazu dienen, Regulationsstrukturen aufzuzeigen, die allen oder zumindest einer Mehrzahl von Sektoren gemeinsam sind. Ein daraus geformtes allgemeines „KI-Gesetzrecht“ könnte *sektorübergreifend* passende Regeln zusammenführen, um rechtliche Lösungen für gleichgelagerte Fälle an zentraler Stelle zu „speichern“. Ob und auf welcher Ebene ein solcher KI-Rechtsentwurf legislativ umgesetzt wird, bleibt weiterer, auch rechtspolitischer Diskussion vorbehalten und erfordert vor allem eine gründliche Prüfung der einschlägigen Kompetenznormen des EU-Vertrags.<sup>26</sup>

<sup>24</sup> Guidelines, S. 5.

<sup>25</sup> Policy Paper, S. 32 ff.

<sup>26</sup> Insbesondere die Forderung der AI HLEG, ein einheitliches „data regulatory framework“ auf EU-Ebene zu schaffen und nationale Regulierung in diesem Bereich weitestgehend auszuschließen (Ziffern 30.1, 30.2 des Policy Papers), mit der EU-Kompetenzordnung vereinbar ist, bleibt zu prüfen.

Der Mehrwert eines solchen kohärenten Entwurfes zeigt sich an erster Stelle darin, dass entsprechendes Wissen über rechtliche Lösungen nicht mehr in allen spezifischen Sektoren immer wieder neu erworben werden müsste. Auch können sektorale Regelungen, sofern sie von den allgemeinen Strukturen abweichen, auch in Zukunft darauf überprüft werden, ob diese Abweichungen immer noch gerechtfertigt sind. So verstanden stellt die hier skizzierte Methode keinen einmaligen, sondern einen regelmäßig zu wiederholenden Prozess dar, an dem das Allgemeine und das Spezifische immer wieder aneinander gemessen werden können.

## AUSBLICK

Die AI HLEG hat einen wertvollen Reflexionsprozess in Sachen KI-Nutzung und -Regulierung angestoßen. Diesen Prozess gilt es nun rechtlich abzurunden. Die Konsultations- und Pilotphase bietet dafür Gelegenheit und kann wichtige Impulse für die weitere Konkretisierung von „lawful AI“ liefern. Der Mehrwert der hier vorgeschlagenen Methodologie besteht darin, dass der politisch-gesellschaftliche Prozess der Konkretisierung in einer Weise strukturiert würde, die den Blick für das Allgemeine wahrt, ohne die Besonderheiten einzelner Sektoren zu vernachlässigen. Ethik und Recht stehen in diesem Konzept keineswegs unverbunden nebeneinander. Vielmehr will die Beschreibung des Prozesses hin zu einer „lawful AI“ Räume öffnen, in die der ethische Diskurs vorstoßen und gemeinsam mit dem Recht auf den vertrauenswürdigen Einsatz<sup>27</sup> von KI hinwirken kann.

<sup>27</sup> Um nicht der – unseres Erachtens ungerechtfertigten – weiteren Vermenschlichung der KI Vorschub zu leisten, sollte nicht von vertrauenswürdiger KI, sondern besser von einem vertrauenswürdigen Einsatz von KI durch den Menschen gesprochen werden. Vgl. zur entsprechenden Kritik an der Wortwahl auch schon *Christoph Marksches*, Warum sollte man einem Computer vertrauen? F.A.Z. v. 23.2.2019, S. 11.

## **EINE VORAUSSCHAUENDE POLITIK FÜR EUROPA – ERGEBNISSE DES WORKSHOPS TOWARDS EUROPEAN ANTICIPATORY GOVERNANCE FOR ARTIFICIAL INTELLIGENCE**

Im September 2019 veranstalteten die interdisziplinäre Arbeitsgruppe *Verantwortung: Maschinelles Lernen und Künstliche Intelligenz* der Berlin-Brandenburgischen Akademie der Wissenschaften und der Forschungsbereich *Technologie und Internationale Politik* der Deutschen Gesellschaft für Auswärtige Politik den Workshop *Towards European Anticipatory Governance for Artificial Intelligence*. Der Workshop brachte führende Experten aus Forschung und Wissenschaft mit politischen Entscheidungsträgern aus den Bundesministerien, Vertretern von Normungsbehörden sowie Technologieunternehmen zusammen. Ziel war es, Rahmenbedingungen für eine vorausschauende europäische Governance zu diskutieren, damit Technologien, die auf Künstlicher Intelligenz (KI) basieren, keinen Schaden anrichten, sondern der Gesellschaft nützen.

Technologien haben immer sowohl wünschenswerte als auch unerwünschte Auswirkungen auf die Gesellschaft. Da KI als „technology of everything“ in fast allen Bereichen zum Einsatz kommen kann und somit jeden direkt oder indirekt betrifft, sollten die Möglichkeiten und Grenzen der Anwendung in einem breiten gesellschaftlichen und interdisziplinären Rahmen ausgehandelt werden. Dabei ist es wichtig, dass die Debatte nicht durch alarmierende oder zu euphorische Narrative verzerrt wird, sondern sich auf Fakten stützt. Mit dem Workshop strebten wir an, KI als technische Anwendung zu entmystifizieren und die tatsächlichen Herausforderungen zu formulieren. Unser konkretes Ziel war es, auf politische Optionen hinzuweisen, die die Handlungsfähigkeit Europas in der Zukunft vor dem Hintergrund der geopolitischen Veränderungen, die KI-basierte Technologien derzeit auslösen, stärken und konsolidieren.

### **EINFÜHRUNG**

Hoffnungen und Bedenken bei der Anwendung von KI-basierten Technologien heizen seit geraumer Zeit öffentliche Kontroversen an. Die Debatte verläuft

zwischen optimistischen Innovationsversprechen und der Angst vor möglichen nicht beabsichtigten oder unerwarteten negativen Auswirkungen. Letztere könnten Risiken für die Grundrechte und Sicherheit, fehlende Bestimmungen für Verantwortung und Rechenschaftspflicht sowie mangelnde Durchsetzung von Gesetzen oder untaugliche Regelungen bedeuten. Zusätzliche Brisanz erhält die KI-Diskussion angesichts des zunehmenden Tempos von Forschung und Entwicklung (F&E) und der möglichen „Dual-Use“-Anwendungen.

Wie in den meisten Fällen neu aufkommender Technologien gibt es also keinen Konsens über mögliche Risiken und Chancen, wohingegen sich deren Steuerung und Kontrolle als sehr komplex herausstellt. Daher ist für KI-Anwendungen eine *European Anticipatory Governance* vonnöten, eine antizipierende Steuerungsform auf europäischer Ebene. Was bedeutet dieser Begriff? Erstens verwenden wir den Begriff *Governance*, weil er auf die Vielfalt der Instrumente und Akteure verweist, die an der Gestaltung von Politik, Gesellschaft und Technologie beteiligt sind. Konkret müssen wir den Begriff *Governance* breiter diskutieren als nur im Zusammenhang mit Gesetzen und industriellen Selbstbeschränkungen in Form ethischer Prüfungen. Es gibt mehr und bereits etablierte Instrumente, um die Gegenwart und Zukunft der Technologie in, mit und für die Gesellschaft zu gestalten. Zweitens weist *Anticipatory* auf die Tatsache hin, dass nicht nur eine einzige, deterministische Variante der Zukunft vor uns liegt, sondern dass mehrere Varianten denkbar sind, die sich je nach unseren Entscheidungen unterscheiden. Es ist unsere gegenwärtige Aufgabe, mögliche positive und unerwünschte Auswirkungen von KI-basierter Technologie in der Gesellschaft abzuschätzen. Auf diese Weise schaffen wir Bewusstsein und im Idealfall gemeinsame Visionen, um Ressourcen zu mobilisieren und Wege in eine für die Gesellschaft vorteilhafte Zukunft zu erarbeiten. Deshalb sollten wir drittens *European* denken, also die Macht der Europäischen Union und anderer europäischer Länder nutzen, um auf der Grundlage europäischer Werte einen strategischen und materiellen Vorteil auf internationaler Ebene zu schaffen. Dadurch kann Europa die Vorteile von KI-Technologien nutzen, aber die Risiken für die europäische Bevölkerung vermeiden. Diese Aufgabe umfasst nicht nur die Technologie- und Industriepolitik, sondern die gesamte Zivilgesellschaft.

Technologien schaffen keine deterministischen Zwänge für Gesellschaft und Politik, sondern umgekehrt, es sind gesellschaftliche und politische Entscheidungen, die einschränkende bzw. begünstigende Bedingungen dafür schaffen, dass Technologien den von Menschen gesetzten Zielen dienen. Die

beiden wichtigsten Fragen in diesem Zusammenhang sind: Was ist unsere Vision einer spezifisch europäischen Entwicklung und Anwendung von KI-basierten Technologien? Was ist die richtige Mischung aus Gesetzgebung, Zertifizierung und Selbstregulierungsansätzen, damit KI einen gesellschaftlichen Nutzen bringt und gleichzeitig unerwünschte Nebenwirkungen verhindert werden? Wir sind diesen Fragen in vier Workshop-Sitzungen zu folgenden Themen nachgegangen: 1. Verhaltenskodizes, Ethik und Moratorien (codes of conduct, ethics, and moratoriums); 2. Standardisierung und Zertifizierung (standardization and certification); 3. nationale, europäische und internationale Gesetzgebung, Verträge und Vereinbarungen (national, European and international legislation, treaties and agreements) sowie 4. Europäische Handlungsfähigkeit (Europe's capacity act).

## **KERNPUNKTE**

Die Entwicklung KI-basierter Technologie schreitet rasant voran und die Einsatzfelder weiten sich stetig aus. Dies stellt eine dreifache politische Herausforderung dar: erstens im Hinblick auf die hohe Unsicherheit bei der Beurteilung künftiger nützlicher oder bedenklicher KI-Anwendungen, zweitens in Bezug auf Wert- und Interessenkonflikte zwischen den beteiligten Akteuren, aber auch Gesellschaften und Staaten, drittens in Bezug auf ein hohes Maß an Komplexität aufgrund der Einbeziehung mehrerer Politikfelder wie Forschungs-, Industrie-, Verbraucher-, Wettbewerbs-, Arbeits-, Verteidigungs- und Außenpolitik. Um diesen Herausforderungen zu begegnen, existiert eine ganze Reihe von politischen Instrumenten – von selbstregulierenden wie beispielsweise Verhaltenskodizes (Codes of Conduct – CoCs), über „weiche“ wie etwa Investitionen in Forschung und Zertifizierung, bis hin zu „harten“ und verbindlichen wie gesetzliche Verordnungen und internationale Vereinbarungen. Die Diskussionen auf dem Workshop führten zur Formulierung folgender Kernpunkte:

### **Politik setzt den Rahmen für KI-Innovationen**

Wenn Europa den Wert KI-basierter Technologien in den europäischen Gesellschaften freisetzen will, brauchen wir ein neues Narrativ, das KI nicht als unbeeinflussbare „Disruption“ – als grundlegenden Umbruch unserer Wirtschaft und Gesellschaft – mystifiziert. Technologien sind kein natürliches Phänomen, das strukturelle Beschränkungen für die Entscheidungsfindung auferlegt. Vielmehr

werden Technologien von Menschen entwickelt und genutzt und ermöglichen dadurch Handlungsspielräume. Politik und Gesetze scheinen oft hinterherzuhinken, weil die technologischen Fortschritte in den Labors und Start-ups gemacht werden und nicht in der Bürokratie. Neu entstehende Technologien und ihre vielfältigen Anwendungen finden jedoch immer in einem politischen, organisatorischen und kulturellen Kontext statt und werden stets von ihnen geprägt. Die gesellschaftliche Einbettung von KI-basierten Technologien ist eine Chance, sowohl mit den verfügbaren regulatorischen Vorgaben als auch mit Anreizen, frühzeitig in KI-Wertschöpfungsketten einzugreifen

### **Die europäischen Werte verteidigen, statt US-amerikanischen oder chinesischen Pfaden zu folgen**

Auch wenn die europäischen Werte (Demokratie, Freiheit, Rechtsstaatlichkeit) nicht unbedingt ein Modell für alle Staaten darstellen, beeinflussen europäische Rechtseingriffe wie etwa die Datenschutzgrundverordnung (DSGVO) zunehmend Regulierungsansätze in anderen Regionen der Welt, einschließlich den USA. Vor diesem Hintergrund kann eine „European branded AI“, die auf hohe Qualitätsstandards und die Einhaltung bestehender gesetzlicher Bestimmungen zu Grundrechten und Nichtdiskriminierung setzt sowie auf exzellenter Pionierforschung basiert, global ausstrahlen. Eine mit einem EU-Label versehene KI könnte einen Wettbewerbsvorteil schaffen, und Europa einen Handlungsrahmen jenseits der bloßen Kopien von US-amerikanischen oder chinesischen Methoden zur Entwicklung und Nutzung von KI ermöglichen. Das heißt auch, dass Europa bei der Schaffung geeigneter Technologien, die die europäischen Werte fördern, mutiger, ehrgeiziger und durchsetzungsfähiger werden kann.

### **Das Potenzial des EU-Modells für „Responsible Research and Innovation“ freisetzen**

Die Forderungen nach einer „ethischen KI“ sind in der öffentlichen Debatte angekommen. Dabei handelt es sich allerdings zunächst um einen abstrakten Anspruch, der operationalisiert werden muss, um handlungsrelevant zu werden. Dabei soll der Fokus nicht nur auf Algorithmen gelegt werden, sondern auch auf die Daten und den gesellschaftspolitischen Kontext, in dem KI-basierte Technologien entwickelt und angewendet werden. Dieser Prozess steht erst am Anfang. Darüber hinaus

ist zu sehen, dass viele der ethischen Prinzipien im Zusammenhang mit KI von der Industrie und der Wirtschaft vorgelegt oder weitgehend beeinflusst werden, ohne ausreichend Experten der (angewandten) Ethik oder der Zivilgesellschaft einzubeziehen. Nicht zuletzt ist die breite Einbindung von Stakeholdern eine der Kernvoraussetzungen für verantwortungsvolle Forschung und Innovation (Responsible Research and Innovation – RRI). Die Umsetzung von RRI, d. h. die Einbeziehung mehrerer Interessengruppen, wechselseitige Ansprechbarkeit und gegenseitige Verbindlichkeit, ist die wichtigste Voraussetzung, um integrative, verantwortbare und akzeptierte Innovationen zu schaffen, die für die Menschen von Nutzen sind. Nur wenn sich diese Bedingungen in einer in Europa entwickelten KI materialisieren, können wir von einer ethischen und menschenzentrierten „European branded AI“ sprechen.

### **Vertrauen gegenüber Institutionen fördern und Verantwortlichkeiten definieren**

Eine der grundlegenden Fragen ist, wer bei der Entwicklung von KI und technologiebasierten Produkten dafür verantwortlich und haftbar ist, dass diese unseren ethischen Maßstäben genügen. Denn insbesondere KI-basierte Anwendungen führen zu einer Diffusion der Verantwortung entlang der Entwicklungsprozesse, was es immer schwieriger macht, bestimmten (personalen) Akteuren moralische und rechtliche Verantwortung zuzuschreiben. In diesem Zusammenhang existiert eine breite Diskussion darüber, ob KI an sich vertrauenswürdig sein kann oder nicht. In dieser Diskussion werden zwei Elemente durcheinandergebracht: Die „Vertrauenswürdigkeit“ einer Technologie ist an technische Normen, Standards und Zertifikate gebunden, die einen Raum akzeptabler technischer Leistung definieren. Das Vertrauen der Bürgerinnen und Bürger wird stattdessen in erster Linie gegenüber öffentlichen Institutionen aufgebaut, die aus Personen (in Behörden und Regierungen) bestehen, die das gesellschaftliche Wohlergehen und die Sicherheit von KI-basierten Technologien garantieren.

### **Technische Standards und Zertifizierungen einführen**

Standards können ein wirksames „weiches“ Regulierungsinstrument für die Entwicklung neu entstehender Technologiebereiche sein, in denen die Gesetzgebung noch nicht alle Aspekte der Technologienutzung erfassen kann, wie die laufenden Bemühungen um Standards für autonome Fahrzeuge zeigen.

Gegenwärtig gibt es auf nationaler, europäischer und internationaler Ebene Bemühungen, gemeinsame technische Terminologien im Bereich der KI zu definieren, technische Spezifikationen zu prüfen, Akzeptanzrisiken zu bewerten sowie rechtliche, gesellschaftliche und ethische Aspekte in die Standards zu integrieren. Ein großer Vorteil des Verfahrens zur Erstellung von Standards ist, dass es von der F&E-Gemeinschaft angetrieben und kontrolliert wird und daher über Rückkopplungsschleifen verfügt, die es anpassungsfähig an technologische Veränderungen machen. Die Einhaltung der Standards könnte zu einer KI-Qualitätszertifizierung für Produkte, die auf den Markt kommen, führen.

### **Vorausschauende Technikfolgenabschätzung und demokratische Aufsicht in die Politikgestaltung integrieren**

Da die technologischen Entwicklungen im Zusammenhang mit KI einen unterschiedlichen Reifegrad haben und sich die Anwendungen auf dem Markt schnell weiterentwickeln, sind die Auswirkungen ihrer gegenwärtigen und zukünftigen Anwendungen nicht klar. Dies erfordert eine Stärkung von zukunftsorientierten Analysen, einschließlich institutioneller, organisatorischer und kultureller/wertpolitischer Fragen. Im Kontext von RRI und Technikfolgenabschätzung sollte die Politik auf nationaler, regionaler und internationaler Ebene mögliche negative Auswirkungen – inklusive gesellschaftliche und geopolitische Aspekte – antizipieren. Diese Erkenntnisse müssen in die politische Entscheidungsfindung, aber auch in die demokratische Kontrolle systematisch einfließen.

### **Ein bewusstes Gleichgewicht zwischen dem Innovations- und dem Vorsorgeprinzip herstellen**

Oftmals mag die Spannung zwischen Innovation und den Vorsorgeprinzipien als unvereinbar erscheinen. Innovationen müssen jedoch nicht unbedingt durch unnötige Verbote eingeschränkt werden. Das Vorsorgeprinzip, das seit 2005 in den EU-Verträgen verankert ist, schreibt eine proaktive Vorsicht vor, wenn es um Risiken für den Verbraucher bzw. Bürger geht, die sich mit den verfügbaren Lösungen nicht verhindern oder mindern lassen. Bei der Vorsorge geht es nicht um Verbote, sondern um die Festlegung von „Verkehrsregeln“ oder auch um die Verhängung von Moratorien, wenn mehr Zeit zur Bewältigung der Risiken erforderlich ist. Dieser Ansatz ist besonders wichtig, wenn es um Technologien geht, die Bedenken



eines möglichen Missbrauchs aufwerfen, indem sie Bürgerrechte einschränken könnten, z. B. bei der Gesichtserkennung, oder für eine „Doppelnutzung“ (Dual Use) im zivilen und militärischem Bereich in Frage kommen.

### **Fähigkeiten stärken, auf nationaler und europäischer Ebene strategisch zu handeln**

Die Maßnahmen auf EU-Ebene erfolgen oftmals langsam und zu vorsichtig, was in den meisten Fällen auf den Unwillen der Mitgliedsländer zurückzuführen ist, an einem Strang zu ziehen. Der Anspruch an Forschung, Entwicklung und Innovationsprozesse im KI-Bereich reicht vom Wohlergehen und Schutz des einzelnen Bürgers über die industrielle Wettbewerbsfähigkeit, den Schutz kritischer Infrastrukturen und die nationale Sicherheit bis hin zur europäischen Handlungsfähigkeit in einer vernetzten Welt. Eine kritische Masse und Skalierungspotenzial, um in der KI-Entwicklung an Fahrwasser zu gewinnen, können nur gemeinsam auf europäischer Ebene erreicht werden. Darüber hinaus brauchen wir innerhalb eines umfassenden EU-Governance-Ansatzes, der sich an RRI-Prinzipien orientiert, eine permanente Dialogplattform, die alle Interessengruppen in der gesamten KI-Wertschöpfungskette einbezieht. Damit könnte die derzeitige Entkopplung zwischen den Akteuren auf der Ebene der KI-Entwicklung einerseits und den Interessenvertretern auf der Ebene der Nutzer sowie den Regulierungsbehörden und öffentlichen Verwaltungen andererseits, aufgehoben werden.

### **Erwünschte und unerwünschte Disruption aus politischer Sicht überdenken**

KI-basierte Technologien haben sowohl in Experten- als auch in Laienkreisen mehrere Debatten über bevorstehende „Disruptionen“ ausgelöst. Nur ein aufmerksames Monitoring kann aufzeigen, was lediglich ein Hype und was der tatsächlich zu erwartende Nutzen ist, aber auch was die Kosten und nicht beabsichtigten Folgen sind. Es ist die Aufgabe der politischen Akteure, fundierte Entscheidungen über wünschenswerte Ziele zu treffen und mit Gesetzen und Anreizen einzugreifen, um diese zu erreichen. Dabei sind ökologische, wohlfahrtsstaatliche, grundrechtliche und sozioökonomische Auswirkungen zu berücksichtigen, ebenso Argumente in Bezug auf den technologischen Vorsprung, den wirtschaftlichen Gewinn oder die Wettbewerbsfähigkeit. Disruption durch Technologie kann, wenn sie nicht aufgefangen wird, zu politischen Turbulenzen und gesellschaftlichen Unruhen führen.

## **KI-basierte Technologien auf intelligente und nachhaltige Weise regulieren**

Eine „European branded AI“ hat das Ziel, die europäischen Werte mithilfe KI-basierter Technologien zu fördern und zu verbessern sowie auf Wohlstand und Fairness hinzuarbeiten. Dafür müssen wir erstens prüfen, wo es bereits eine Regulierung gibt, die für KI-basierte Technologien gilt, und zweitens müssen wir entscheiden, welche Art von neuer Regulierung sinnvoll ist. Gleichzeitig können wir von anderen sehr wichtigen RRI-Bereichen, wie dem medizinischen und pharmazeutischen Sektor, lernen, wie rechtliche Rahmen der Selbstregulierung letztlich eine Möglichkeit zur Durchsetzung von Verhaltenskodizes sein könnten. Um wettbewerbsfähig zu bleiben, kann es ein gangbarer Weg sein, nicht die Technologie an sich zu regulieren, sondern zu definieren, wie wir KI-basierte Technologien nutzen wollen und welche Art von Anwendung wir nicht tolerieren werden. Eine technologieneutrale Regulierung ermöglicht es auch, technische Weiterentwicklungen an etablierten KI-Anwendungen auszurichten.

## **Investitionen in Faktoren, die KI-Innovation ermöglichen**

Governance-Maßnahmen sollten die Aufnahme und Verbreitung von KI in Unternehmen, Behörden und in der Forschung ansprechen und fördern, indem sie die Vertreter dieser Sektoren und die Bürger im Allgemeinen in die Lage versetzen, fundierte Entscheidungen zu treffen. Ohne die Ausbildung von digitaler und KI-Kompetenz, die über ein grundlegendes Verständnis der Logik von Algorithmen und der Rolle von Daten hinausgehen muss, wird es keine Verbreitung innovativer technologischer Lösungen geben – und auch keine kritische Aufsicht. Gleichzeitig müssen Philosophen, Ethiker und Sozialwissenschaftler in den Besonderheiten von KI-Anwendungen geschult werden, um das Potenzial einer „European branded AI“ zu realisieren.

## **Europäische Champions aufbauen**

Dies wird erfordern, die Kräfte auf EU-Ebene zu bündeln, anstatt getrennte nationale KI-Strategien zu verfolgen. Darüber hinaus müssen die europäischen Regierungen im neuen mehrjährigen Finanzrahmen Geld in die Hand nehmen, nachdem die Ende 2019 neu eingesetzte Europäische Kommission die

herausragende Rolle der digitalen Technologien erklärt hat. Anstatt einem rückwärtsgerichteten Verteilungsmodell zu folgen, müssen die Dossiers Forschung und Entwicklung, Digitalisierung und Wettbewerb angesichts der geopolitischen Herausforderungen, denen Europa sich im Hinblick auf die Dynamiken in China, Indien, den USA und Russland gegenübersteht, gestärkt werden. Eine enge europäische Koordinierung der Wettbewerbs-, Innovations-, Handels- und Grundrechtspolitik ist der Schlüssel zu einem wirksamen, kohärenten und nachhaltigen Instrument zur Mobilisierung des ungenutzten Potenzials KI-basierter Technologien für Europa. Entscheidend für die Verbreitung von KI-gestützten B2B- oder B2C-Lösungen ist eine Infrastruktur, die Konnektivität und Interoperabilität ermöglicht. Die Innovation sollte auf einer zweck- und regelgesteuerten gemeinsamen Nutzung von Daten beruhen und gleichzeitig die in den Europäischen Verträgen und der Europäischen Charta festgeschriebenen Grundrechte schützen.

## **ZUSAMMENFASSUNG DER WORKSHOP-SITZUNGEN**

### **Verhaltenskodizes, Ethik und Moratorien im Bereich F&E**

Die erste Sitzung begann mit einer Debatte über das aktuelle Ethik-Mainstreaming im KI-Diskurs. Es wurde die Frage aufgeworfen, ob verantwortungsvolle Forschung und Innovation (Responsible Research and Innovation – RRI) ein konkreter Weg sein könnte, ethische Standards in der Praxis umzusetzen, und ob selbst auferlegte Verhaltenskodizes (Codes of Conduct – CoCs) funktionieren können. Hier dienten die Ethikrichtlinien für vertrauenswürdige KI der Hocharangigen Expertengruppe für Künstliche Intelligenz (AI HLEG) der Europäischen Kommission vom April 2019 als Referenzdokument.<sup>1</sup> Die Kernfrage ist, wie wir eine inhärente europäische Ethik, die auf dem europäischen Erbe, den europäischen Verträgen und der Europäischen Grundrechtecharta basiert, im Zeitalter von neuen KI-gestützten Technologien geltend machen.

#### **Die Ethik-Diskussion mit RRI-Ansätzen zusammenführen**

Als Ausgangspunkt kann die Ethikdiskussion über KI am besten mit anderen normativen Ansätzen der technologischen Entwicklung zusammengeführt werden.

<sup>1</sup> Die Ethikrichtlinien für vertrauenswürdige KI sind unter folgendem Link in verschiedenen Sprachen abrufbar: <https://ec.europa.eu/digital-single-market/en/news/ethics-guidelines-trustworthy-ai> (letzter Zugriff am 10. Februar 2020).

Ein solcher Ansatz ist RRI, „der potenzielle Auswirkungen und gesellschaftliche Erwartungen in Bezug auf Forschung und Innovation antizipiert und bewertet, mit dem Ziel, die Gestaltung einer integrativen und nachhaltigen Forschung und Innovation zu fördern“.<sup>2</sup>

Basierend auf der Methodik der Technologiefolgenabschätzung (TA) wurde RRI im Rahmen des EU-Forschungs- und Innovationsprogramms „Horizon 2020“ als Querschnittspriorität für alle Forschungsmaßnahmen mit einer Laufzeit von sieben Jahren (2014 bis 2020) umgesetzt. Der Ausgangspunkt von RRI ist, dass es zwar allgemeine Prinzipien gibt, denen die technologische Entwicklung folgen muss – z. B. Menschenrechte – die aber in bestimmten Situationen je nach sozialem Kontext unterschiedlich interpretiert werden. Technologie ist immer kontextabhängig, da soziotechnische Systeme und soziale Gefüge unterschiedlich sind. RRI konzentriert sich auf soziale Herausforderungen, die Einbeziehung von Stakeholdern, Risikovermeidung durch Antizipation, Inklusivität und Verantwortung.

RRI thematisiert dabei viele der Herausforderungen in der aktuellen Diskussion über KI, nämlich der Entkopplung zwischen der Ethikdiskussion über KI und der tatsächlichen Entwicklung und Anwendung von KI-basierten Technologien. Die Ethikrichtlinien für vertrauenswürdige KI versuchen, diese Diskrepanz zu überwinden, aber dennoch könnte man sich von RRI vieles abschauen, da es bereits greifbare Maßstäbe und Lehren gibt und der Ansatz außerhalb Europas sehr gut bekannt ist und angewendet wird. Gegenwärtig wird RRI in den noch weiter gefassten Begriff „Open Science“ integriert. Ein „RRI/Open Science“-Ansatz zur Entwicklung der KI könnte einen praktischen Versuch darstellen, international zusammenzuarbeiten. Darüber hinaus könnte er zur Entmystifizierung von KI beitragen, um sie als das zu sehen, was sie ist: ein technologischer Wegbereiter (Technology Enabler), der in Zeiten der Digitalisierung die Analyse von großen Datenmengen (Big Data) ermöglicht.

2 Im Original auf der Seite des EU-Förderprogramms Horizon 2020: “Responsible research and innovation is an approach that anticipates and assesses potential implications and societal expectations with regard to research and innovation, with the aim to foster the design of inclusive and sustainable research and innovation. Online unter: <https://ec.europa.eu/programmes/horizon2020/en/h2020-section/responsible-research-innovation>“ (letzter Zugriff am 10. Februar 2020).

## **Einhaltung von Verhaltenskodizes**

Wenn wir CoCs im Geiste von RRI und Open Science umsetzen wollen, stellt sich die Frage nach deren Verbindlichkeit. Hier empfiehlt es sich, einen Schritt zurückzutreten und zwei entscheidende Aspekte in der Diskussion zu berücksichtigen: Zum einen besteht ein Ungleichgewicht in der Beteiligung der Industrie an der derzeitigen Ausarbeitung zahlreicher Ethikrichtlinien, zum anderen führen KI-basierte Technologien zu Unklarheiten bezüglich der menschlichen Verantwortung.

Industrielle Akteure spielen eine herausragende Rolle bei der Definition ethischer Prinzipien für die Entwicklung und Anwendung von KI, sei es intern, in Zusammenarbeit mit anderen Unternehmen, in politischen Gremien oder auch im akademischen Umfeld. Microsoft und Facebook können als prominente Beispiele genannt werden. Das gilt auch für die EU-Ethikrichtlinien für vertrauenswürdige KI, da die Industrie in der AI HLEG, die die Richtlinien vorlegte, stark vertreten ist. Generell setzt die Industrie auf Selbstregulierung, denn je verbindlicher bestimmte Regeln sind, desto mehr werden sie als innovations- und geschäftshemmend angesehen. Es ist bemerkenswert, dass bei der Festlegung ethischer Regeln die Experten auf diesem Gebiet – Philosophen und Ethiker – unterrepräsentiert zu sein scheinen. Wenn wir wirklich davon überzeugt sind, dass Europa eine intellektuelle Führungsrolle bei der Definition und Umsetzung einer ethischen KI hat, müssen wir auch eine neue Generation von Experten auf dem Gebiet der (angewandten) Ethik ausbilden. In diesem Zusammenhang verbirgt der Begriff „Vertrauenswürdigkeit von Technologien“ ein wichtiges Element, nämlich dass die Bürger Vertrauen zu öffentlichen Institutionen aufbauen, die sich aus Menschen zusammensetzen, die das gesellschaftliche Wohlergehen und die Sicherheit von KI-basierten Technologien garantieren, und nicht zur Technik an sich.

Zweitens führen KI-Anwendungen zu einer Diffusion der Verantwortung zwischen menschlichen und nichtmenschlichen Akteuren sowie entlang von Entwicklungsprozessen, was es immer schwieriger macht, bestimmten (personalen) Akteuren moralische und rechtliche Verantwortung zuzuschreiben. Eine der grundlegenden Fragen, die es zu klären gilt, ist, wer verantwortlich und damit haftbar ist, dass die Entwicklung und die Produkte mit KI-Technologie unseren ethischen Standards entsprechen. Ist es das Unternehmen, das die Anwendung entwickelt, der Coder oder die Datenwissenschaftlerin, die das KI-System bauen, eine staatliche Stelle, die die Anwendung als Dienstleistung anbietet oder die Person, die sie nutzt? Obwohl das Problem der verteilten Verantwortlichkeiten in der technologischen

Entwicklung schon immer bestand, sind wir jetzt mit einer neuen Schwere des Problems konfrontiert, denn die Technologien, die die KI nutzen, werden nicht nur in einzelnen Teilbereichen angewendet, sondern sie schaffen ein ganzes technisch-soziales System, das unsere Gesellschaften in grundlegender Weise beeinflussen wird.

### **Technologieanwendungen regulieren, nicht (nur) die Technologieentwicklung**

Ein komplementärer, laufender Diskurs ist die Frage, ob oder wie Technologien angesichts der rasch fortschreitenden technologischen Entwicklung auf angemessene Weise reguliert werden können. Ein grundlegender Anspruch ist es, nicht die Technologie an sich zu regulieren, sondern zu definieren, wie wir wollen, dass KI-basierte Technologien in Übereinstimmung mit unseren Werten eingesetzt werden, und welche Art von Anwendung wir nicht zu tolerieren bereit sind. Eine technologieneutrale Regulierung ermöglicht es, die Weiterentwicklung von Technologien anhand der bewährten gesellschaftlichen Nutzung zu konkretisieren. An dieser Stelle ist es wichtig, darauf hinzuweisen, dass es bereits Gesetze gibt, die ein bestimmtes gewünschtes soziales Ergebnis vorsehen, wie z. B. das europäische Antidiskriminierungsgesetz und seine nationalen Umsetzungen. Um jedoch neue Gesetze zu erlassen oder bestehende Gesetze durchzusetzen, ist ethisches, rechtliches, soziales und technisches Fachwissen bezüglich der Entwicklung und Anwendung von KI auf Seiten von Behörden und Beamten erforderlich.

Schließlich sollte ein europäischer KI-Ansatz menschenzentriert sein. Das bedeutet konkret, dass eine gute Zusammenarbeit zwischen Mensch und Maschine gewährleistet werden muss. Der Schwerpunkt ist darauf zu legen, was für den Menschen am besten ist, nicht darauf, was den effizientesten und optimiertesten Prozess darstellt. Wenn wir die Regulierung von KI-Anwendungen diskutieren, müssen wir daher auch daran denken, Grenzen zu definieren. Diese sollten nicht als Innovationsbeschränkungen, sondern vielmehr als Innovationsförderer wahrgenommen werden, die sicherstellen, dass technologische Neuerungen den beabsichtigten Nutzen bringen, und verhindern, dass sie dem Menschen Schaden zufügen. Beispiele hierfür könnten zum Beispiel der Einsatz von KI im Krieg oder spezifische Anwendungen wie Gesichts- oder Emotionserkennung sein. Denkbar sind auch Moratorien in diesen Bereichen, bis geeignete Mechanismen zur Risikobewältigung etabliert sind, wie die jahrzehntelange Erfahrung im Bereich der Gentechnologie und der pharmazeutischen Forschung gezeigt hat.

## **Standardisierung und Zertifizierung**

Die zweite Sitzung nahm mit Aspekten von Standardisierung und Zertifizierung auch Bedingungen für die Wettbewerbsfähigkeit im globalen KI-Bereich vor dem Hintergrund der europäischen Grundrechte und Werte mit auf. Gegenstand der Debatte waren Internationalisierungstendenzen im Gegensatz zu Politikmaßnahmen auf nationaler Ebene.

### **KI-Finanzierung**

Europa kann mit Spitzenforschung an Universitäten und Exzellenzzentren im KI-Bereich aufwarten. Forschungseinrichtungen und Innovationszentren im europäischen Kontext sollten jedoch zusammengeführt werden, um Synergien zu nutzen und dadurch eine wettbewerbsfähige kritische Masse in der KI-Entwicklung zu schaffen. Das ist umso wichtiger, da sich die großen US-amerikanischen „GAFAM“ (Google, Amazon, Facebook, Apple, Microsoft)-Technologieunternehmen immer mehr in „KI-Unternehmen“ wandeln. Um eine höhere europäische Wettbewerbs- und Innovationsfähigkeit zu erreichen, sollte Europa nicht in Bereichen konkurrieren, in denen es fast unmöglich ist, neben den USA und China noch Fuß zu fassen – beispielsweise der Aufbau einer europäischen Suchmaschine. Stattdessen könnten die EU und ihre Mitgliedsstaaten eine Nische in der KI-Branche finden, um eigene innovative Produkte mit Marktpotenzial und im Einklang mit den europäischen Werten zu schaffen. In diesem Zusammenhang sollte Europa die Stärken der industriellen Basis bündeln und diese Vorteile mit KI-Technologien verbinden. Auch wenn es die Auffassung zu geben scheint, dass nur große Länder und Organisationen im KI-Bereich einen Innovationsdurchbruch erzielen können, gibt es auch viel Potenzial für kleine Akteure.

### **Auf dem Weg zu internationalen Normen und technischen KI-Standards**

Standards werden eine Schlüsselrolle spielen. Auf nationaler und internationaler Ebene gibt es in diesem Bereich eine gewisse Arbeitsteilung: Die politischen Entscheidungsträger definieren die spezifischen gesellschaftspolitischen Anforderungen, aber die Standardisierungsgremien legen die konkreten technischen Standards fest. Um KI-Standards zu entwickeln, sind gemeinsame Konzepte, Operationalisierung und Vokabular erforderlich. ISO/IEC JTC 1 entwickelt derzeit

eine internationale Norm zur KI-Terminologie. Der erste Entwurf dieser Norm wird voraussichtlich bereits in diesem Jahr verfügbar sein.

In diesem Prozess muss auch ein ausgewogener Ansatz zwischen Innovation und Standardisierung mit marktorientiertem Denken erfolgen. Wenn Standards zu früh gesetzt werden, können sie Innovationen blockieren. Es ist zu bedenken, dass es verschiedene Arten von Standards gibt, z.B. in der Terminologie, die Innovation durch Verbesserung der Interoperabilität fördern können, oder in der Umsetzung von Technologieanwendungen, die während der Nutzungsphase eingeführt werden. Die Integration von EU-Werten in KI-Systeme könnte sich positiv auf den Markt auswirken und ein gutes Verkaufsargument für die EU weltweit darstellen. Zusammen mit der hohen Qualität der europäischen F&E- und den internationalen Standardisierungsbemühungen könnte eine „European branded AI“ geschaffen werden, die nicht nur ein Fenster der Gelegenheit eröffnet, die USA und China aufzuholen, sondern auch mit diesen Ländern zu konkurrieren.

### **Public-Private-Partnerships**

Die Zusammenarbeit zwischen öffentlichen und privaten Einrichtungen bei der Entwicklung von KI wird zunehmend schwierig, da Erstere nicht über die gleichen Kenntnisse verfügen wie Letztere. Das kann dazu führen, dass öffentliche Organisationen nicht immer in der Lage sind, zu beurteilen, womit sie es zu tun haben. Eine Investitionsmöglichkeit besteht daher darin, das Wissen und die Kompetenz des Staates und der Öffentlichkeit im Allgemeinen gegenüber KI-basierten Technologien zu erhöhen. Darüber hinaus muss die Beziehung zwischen Verbrauchern und der Industrie auf einem greifbaren Nutzen der Technologien basieren. Europäische KI-Systeme sollten so entwickelt werden, dass sie benutzerfreundlich und gleichzeitig standardkonform sind.

Viele der aktuellen europäischen KI-Projekte haben Pilot- und Vorbereitungscharakter. Es wurde die Frage aufgeworfen, ob EU-Initiativen zur Entwicklung von Forschungs- und Innovationsrahmenwerken wie Horizon Europe, KI in ihrer künftigen wichtigen Rolle ausreichend berücksichtigen. Die EU scheint bisher einen vorsichtigen Ansatz zu verfolgen, was sich jedoch im Rahmen des bevorstehenden mehrjährigen Forschungsprogramms „Horizon Europe 2021–2027“ ändern könnte. Die EU hat immer noch eines der größten Bruttoinlandsprodukte weltweit und kann auf eine enorme Marktgröße zurückgreifen, daher sollte sie mutig genug sein,



in starke Innovationszentren zu investieren. Dennoch scheint ein Missverhältnis zwischen den über die nationalen KI-Strategien vermittelten hochgesteckten Zielen und dem, was kurzfristig realistisch ist, zu bestehen.

## **Gesetzgebung, Verträge und Vereinbarungen auf nationaler, europäischer und internationaler Ebene**

Nach den weichen Regulierungsinstrumenten ging es in der dritten Sitzung des Workshops um die Gesetzgebung als hartes Regulierungsinstrument, um die Rahmenbedingungen für vorteilhafte KI-basierte Technologien zu gestalten. Was könnte eine gute Regulierung in einem so aufstrebenden und breiten Technologiebereich sein, ohne Innovationen zu ersticken?

### **Neue Regulierungen vs. bessere Umsetzung bestehender Regulierungen**

KI wird zwangsläufig Politikbereiche wie Wettbewerb (Kartellrecht), Industrie und Handel (inkl. Dual Use), Justiz (inkl. GDPR), Verbraucherschutz, Verteidigung und Außenpolitik beeinflussen. Es gibt zwar Bereiche, die neue Rechtsvorschriften erfordern, wie der sich rasch entwickelnde Anwendungsbereich der tödlichen autonomen Waffensysteme (LAWS), aber wir müssen von Fall zu Fall sorgfältig prüfen, wann und wo sich KI-Anwendungen in einem Maße entwickeln, dass die bestehende Regulierung nicht alle potenziellen Anwendungsfälle berücksichtigen kann. Es erscheint sinnvoll, die Technik nicht als solche zu regulieren, sondern spezifische Anwendungen, die mit KI arbeiten. Daher sollten wir anhand von Anwendungsfällen beurteilen, welche Art von Anwendungen und entsprechender Regulierung es bereits gibt, bevor wir neue verabschieden. Diese Position wurde durch Ansätze in Estland und anderen nördlichen EU-Ländern bestätigt, die zeigen, dass es keinen großen Bedarf an neuen Vorschriften gibt, die sich speziell mit KI befassen, auch um die Gefahr einer Überregulierung zu vermeiden. Dies könnte insbesondere für kleinere Neugründungen gelten, da große (US-)Unternehmen neue bzw. mehr Gesetze leichter erfüllen können, aber kleinere europäische Unternehmen ihre Wettbewerbsfähigkeit verlieren könnten.

Eine Schlüsselfrage, an der sich neue politische Initiativen orientieren sollten, ist, ob wir es mit einem neuen Problem zu tun haben oder ob das Problem durch eine bereits bestehende Regulierung angegangen werden kann. Im Zusammenhang

mit den Debatten der letzten Jahre über „Überregulierung“ sollten wir darauf achten, dass wir nicht unnötig neue Gesetze hinzufügen, deren Durchsetzung Ressourcen binden und die Umsetzung von Initiativen erschweren. Ein strittiger Punkt in der Diskussion war, ob wir spezialisierte, maßgeschneiderte Gesetze für bestimmte KI-Anwendungen brauchen oder ob dies zu einer Explosion der Regulierung führen würde. Die gegenteilige Ansicht war, dass eine intelligente Regulierung die Anzahl der Gesetze reduzieren und stattdessen effizienter bei der Definition von Rahmenbedingungen werden könnte, die mehrere Aspekte von KI-Anwendungen auffangen.

Im Bereich der neu entstehenden und konvergierenden Technologien, mit denen wir es zu tun haben, ist nicht vorherzusehen, wie sie sich auf die Gesellschaft auswirken werden. Als Präzedenzfall kann die Regulierung von Nukleartechnologien für die zivile und militärische Nutzung gelten, die zu jahrzehntelangen und noch immer andauernden internationalen (schwierigen und nicht immer erfolgreichen) Verhandlungen über Entwicklungsbeschränkungen, Testverbote und bedingte Nutzungsdefinitionen geführt hat. Angesichts der probabilistischen Natur von KI können Verfahren so gestaltet werden, dass die rechtlichen Rahmenbedingungen für bestimmte Anwendungsbereiche, statt für die Technologie selbst, getestet werden. Dies könnte zu einer besseren Vorhersehbarkeit der Ergebnisse führen und mittelfristig Sicherheit für die Bürgerinnen und Bürger bieten. In dieser Hinsicht sind die aktualisierten Kontrollvorschriften für die Ausfuhr von sensiblen Gütern und Dienstleistungen mit doppeltem Verwendungszweck und, was entscheidend ist, der immaterielle Wissenstransfer, einschließlich der KI-bezogenen Bereiche, ein Beispiel.

Es besteht oft Unsicherheit auf Seiten der F&E-Gemeinschaft, welcher rechtliche Rahmen für ihre Aktivitäten gilt, einschließlich des Risikos, für die Entwicklung oder Anwendung von KI-Anwendungen bestraft zu werden. Die aktuelle Gesetzgebung wurde hauptsächlich vor dem aktuellen „KI-Hype“ implementiert, daher müssen wir ihre Eignung prüfen, um die sehr unterschiedlichen Facetten von KI-gestützter Technologie zu berücksichtigen. Gegenwärtig sieht nur die DSGVO einige Regelungen vor, die sich mehr oder weniger spezifisch mit KI befassen. Bevor wir jedoch auf die Option einer allumfassenden KI-Regulierung zurückgreifen, müssen wir prüfen, ob die Kombination der EU-Grundrechtecharta mit der entsprechenden Rechtsprechung (einschließlich solcher Aspekte wie Rechte des Kindes oder Meinungsfreiheit) ausreichen würde, um z. B. maschinengesteuerte Gesichtserkennungstechnologien zu erfassen. Angesichts der Tatsache, dass

die Rechtsprechung den Interessenvertretern nur langsam Orientierungshilfen bietet, könnten daher, insbesondere im Bereich der Haftung, einige neue Gesetze gerechtfertigt sein.

### **Grenzen definieren**

Die Herausforderung für politische Interventionen besteht darin, auf zweifache Weise Rahmenbedingungen zu schaffen: Einerseits sollten die europäischen Regierungen in einer gemeinsamen und konzertierten Anstrengung Innovation und unternehmerische Aktivitäten anregen. Zum anderen sollten sie Bereiche definieren, in denen keine KI-basierten Technologien zum Einsatz kommen sollten. Das Vorsorgeprinzip ist seit 2005 in den EU-Verträgen verankert und sollte im Rahmen der Technologiefolgenabschätzung angewendet werden. Ein Weg wäre es, das künftige Verbraucherschutzrecht entlang der Regulierung der immer stärker in den Vordergrund tretenden Mensch-Maschine-Interaktionen zu gestalten. In diesem Zusammenhang sollten neben der Diffusion gefährlicher Produkte oder Dienstleistungen in erster Linie die „verteilten Verantwortlichkeiten“ angesprochen werden.

Eine beunruhigende Entwicklung, die es zu berücksichtigen gilt, ist, dass Unternehmen den Geltungsbereich des EU-Rechts verlassen und nach China oder in die USA gehen, um KI-Anwendungen zu entwickeln und die Produkte anschließend auf dem EU-Markt zu verkaufen. Um auf die Regulierungskompetenz der EU vertrauen zu können, muss daher ein stärkerer, wirksamerer und effizienterer EU-weiter Mechanismus entwickelt werden, um Doppelarbeit und Widersprüche zwischen Verbraucher-, Wettbewerbs-, Handels- und Menschenrechtsvorschriften zu beseitigen. Mehr Kohärenz und Harmonisierung würde den Interessen sowohl der Industrie als auch der Bürger besser dienen.

### **Die Ziele der Regulierung klarstellen**

Den derzeitigen Strategien fehlt oft ein bestimmtes Ziel, sondern sie versuchen, mehreren, teilweise konkurrierenden Zielen gleichzeitig gerecht zu werden. Daher liefern sie kein Instrument für konsistente oder konzertierte Maßnahmen in der gesamten EU. Darüber hinaus macht die Verbreitung von (nationalen) Strategien die Politik nicht handlungsorientierter. Mögliche Fallstricke der Regulierung lassen

sich am Beispiel der DSGVO veranschaulichen. Während diese eine Erfolgsgeschichte zu werden scheint, die das Potenzial hat, auch anderswo in der Welt verbreitet und übernommen zu werden, kennen Bürger oder Unternehmen die Vorteile der DSGVO nicht immer an, sondern sehen darin eher ein zusätzliches bürokratisches Hindernis. Die neue Europäische Kommission scheint in ihren Ankündigungen bereit zu sein, Fragen der Digitalisierung auf koordinierte und globale Weise zu behandeln.

## **Europäische Handlungsfähigkeit im Hinblick auf KI-Governance**

Die Abschlussdiskussion führte die Themen der drei vorangegangenen Sitzungen unter dem Gesichtspunkt der europäischen Handlungsfähigkeit in Bezug auf KI in der sich derzeit verändernden geopolitischen und wirtschaftlichen Weltordnung zusammen.

### **Das Gleichgewicht zwischen dem europäischen Wertesystem und einem lebensfähigen Europa in der Welt herstellen**

Bei der Entwicklung von Politiken müssen wir den breiteren Kontext betrachten, um eine europäische Kapazität aufzubauen, die auf internationaler Ebene wettbewerbsfähig ist. Die Sorge ist, dass es einen Zielkonflikt gibt zwischen Wettbewerbsfähigkeit einerseits und der Rolle einer wertorientierten EU-Politik auf der Grundlage von Demokratie und Nichtdiskriminierung, dem Recht auf Datensicherheit und Transparenz andererseits. Doch die Verbindung muss kein Trade-off sein, sondern kann als „European branded AI“, d. h. als einzigartige „innovative EU-Dienstleistung“, vermarktet werden. Die „Vertrauenswürdigkeit“ in die europäische KI – das heißt, dass KI-Anwendungen durch Institutionen und Standards abgesichert sind – kann ein potenzielles Alleinstellungsmerkmal sein. Die EU könnte es sich zum Ziel setzen, auf diesem Gebiet eine Führungsrolle zu übernehmen und durch die Leitung internationaler und multilateraler Konferenzen zu diesem Thema Debatten über verantwortungsvolle und rechenschaftspflichtige KI-Forschung, Innovation und Anwendung zu fördern. Wichtig ist die Balance zwischen einem funktionierenden Regulierungsrahmen für KI und Anreizen für wachsende KI-Lösungen in bestimmten Bereichen. EU-Initiativen können sich auf die positiven Ergebnisse von KI-Dienstleistungen für Bürger und die Gemeinschaft konzentrieren.

## **Das Ökosystem im Blick haben und die wichtigsten Interessengruppen entlang der Wertschöpfungskette miteinbeziehen**

Wettbewerbsfähigkeit auf europäischer Ebene kann nur gemeinsam in einem europäischen Ökosystem erreicht werden, das die Infrastruktur für die Entwicklung und Anwendung von KI bietet. Dies bedeutet, dass nationale Digitalisierungs- und KI-Strategien in eine einzige europäische integriert werden müssen. Es wird notwendig sein, alle Teile der industriell-technologischen Basis der EU zu aktivieren, um Synergien zu nutzen und kostspielige Doppelarbeit zu vermeiden. Darüber hinaus sollte ein umfassender EU-Governance-Ansatz eine permanente Dialogplattform schaffen, die alle Beteiligten entlang der gesamten KI-Wertschöpfungskette einbezieht, um die derzeitige Entkopplung der Akteure zu verbessern. Dazu gehört die Zusammenführung von KI-Entwicklern, Interessenvertretern auf der Nutzerseite, Regulierungsbehörden, Unternehmen, Lieferanten, Bürgern und der organisierten Zivilgesellschaft.

## **Europaweite Infrastrukturen aufbauen, um KI-Champions zu ermöglichen**

Eine der Herausforderungen für die Zukunft wird es sein, einen besseren Datenaustausch zwischen den EU-Ländern zu erreichen. Die Menge und Qualität der in der EU verfügbaren Daten ist eine wertvolle Ressource für die Entwicklung von KI-Systemen, die derzeit jedoch nicht voll ausgeschöpft wird. Der Grund dafür, dass ein solcher Datenaustausch nicht leicht umzusetzen ist, liegt zum Teil darin, dass die Länder unterschiedliche Ansichten über die Notwendigkeit nationaler Einschränkungen haben. Eine Voraussetzung für Datenaustausch ist daher die Aufstellung von Regeln für die gemeinsame Nutzung und die Interoperabilität von Infrastrukturen. Darüber hinaus gehört der Aufbau von kofinanzierten und kooperierenden EU-weiten Infrastrukturen für Forschung – von der Grundlagenforschung bis zur angewandten Forschung – zu den unabdingbaren Rahmenbedingungen, um Ideen auf den Markt, in die öffentliche Verwaltung und in die Gesellschaft zu bringen. Nicht zuletzt muss der „Tech-Nationalismus“ einem einheitlichen europäischen Ansatz weichen, um die Vorteile von KI-basierter Technik zu nutzen. Dies ist auch der Schlüssel um auf dem globalen Markt aktiv zu werden. Europäische KI-Champions, die mit US-amerikanischen und chinesischen Initiativen konkurrieren, müssen auf gesamteuropäische Infrastrukturen und Ressourcen zurückgreifen können.

Ein herzlicher Dank gilt den Teilnehmern des Workshops für ihre Impulse und Diskussionsbeiträge.

**TITEL DER REIHE »#VERANTWORTUNGKI – KÜNSTLICHE INTELLIGENZ  
UND GESELLSCHAFTLICHE FOLGEN«**

**Heft 1/2020**

Isabella Hermann, Georgios Kolliarakis, Fruzsina Molnár-Gábor,  
Timo Rademacher, Frauke Rostalski

**VERTRAUENSWÜRDIGE KI?**

**VORAUSSCHAUENDE POLITIK!**





Der zunehmende Einsatz von sogenannter „Künstlicher Intelligenz“ (KI) verspricht viele Verbesserungen, beispielsweise durch Bilderkennung in der Medizindiagnose. Er birgt aber auch das Risiko, dass Menschen durch irrtümliche Vorhersagen von KI-Systemen zu Schaden kommen können. In solchen Fällen wird es immer schwieriger zu bestimmen, wer die Verantwortung trägt. Die Reihe #VerantwortungKI – Künstliche Intelligenz und gesellschaftliche Folgen bietet ein Forum für Beiträge über die ethischen, rechtlichen und gesellschaftspolitischen Chancen und Risiken des Einsatzes von KI mit einem besonderen Blick auf den Verantwortungsbegriff. Die Beitragsreihe wird von der interdisziplinären Arbeitsgruppe *Verantwortung: Maschinelles Lernen und Künstliche Intelligenz* betreut.