10.20378/irb-58361

KI-Regulierung durch die EU – Eine Einschätzung der Regulierungsvorschläge der EU-Kommission vom April 2021 aus ökonomischer Perspektive

Michael Vogelsang

Hochschule Ruhr West, michael.vogelsang@hs-ruhrwest.de

Abstract:

Künstliche Intelligenz betrifft als Querschnittstechnologie alle betrieblichen Funktionen in unterschiedlichen Branchen. Die Beziehungen von Unternehmen, Bürgerinnen und Bürgern und Staat werden durch Anwendungen der Künstlichen Intelligenz beeinflusst oder gesteuert. Mit der Einführung von KI-Anwendungen gehen Vorteile, aber auch mögliche Risiken einher. Daher hat die Europäische Kommission im April 2021 einen Vorschlag für eine Regulierung von Künstlicher Intelligenz vorgestellt. In diesem Beitrag wird der Regulierungsvorschlag vor dem Hintergrund ökonomischer Theorien analysiert.

JEL Classification: C45, D40, L51

Keywords: Künstliche Intelligenz, Wirtschaftspolitik, Regulierung, Europäische Kommission.

1 Einleitung

Unternehmen und ihre Verbände, Nichtregierungsorganisationen, Wissenschaftlerinnen und Wissenschaftler aus der Informatik, Philosophie und Ökonomie und vielen anderen Bereichen haben in den letzten Jahren auf die Regulierungsvorschläge zur Künstlichen Intelligenz (KI) eingewirkt. Als Ergebnis hat die EU-Kommission am 21. April 2021 einen Regulierungsvorschlag zur Querschnittstechnologie KI vorgestellt. Ziel der Kommission ist eine vertrauenswürdige KI, die Sicherheitsanforderungen erfüllt und Menschenrechte schützt. Durch das regulierungsinduzierte Vertrauen erarbeitet sich Europa nach Hoffnung der EU-Kommission einen Wettbewerbsvorteil bei der Entwicklung von KI-Anwendungen.

Dieser Beitrag analysiert die möglichen Probleme, die mit den KI-Anwendungen einhergehen und bewertet die Regulierungsvorschläge der EU-Kommission anhand der drei Kriterien Klarheit, Zielsetzung und Effizienz vor dem Hintergrund ökonomischer Theorien.

2 KI-Entwicklung und Regulierungsbedarf

2.1 Aktueller Stand

Der Mathematiker Alan Turing schätzte 1950, dass zum Ende des Jahrhunderts Maschinen so weit wären, den von ihm konzipierten Turing Test zu bestehen. Wenn ein menschlicher Fragesteller nicht erkennen kann, ob die Antworten von einem Menschen oder einer Maschine stammen, sei das *Imitation Game* bestanden. Dies gelte dann als Indiz, dass Maschinen denken können (Turing, 2021).

In seiner Konzeption des *Imitation Game* ging Turing von einer schriftlichen Kommunikation aus. Google präsentierte im Jahr 2018, dass Maschinen ebenfalls in natürlicher Sprache mit Menschen kommunizieren können, ohne dass diesen auffällt, dass es sich bei dem Gesprächspartner um eine Maschine handelt.[58]

Für derartige NLP- (Natural Language Processing) Anwendungen werden neuronale Netze mit Beispielen für menschliche Sprache trainiert. Dieses Verfahren kann ebenso eingesetzt werden, um aus Programmierwünschen, die in menschlicher Sprache geäußert werden, den entsprechenden Code als Output zu erzeugen. OpenAI präsentierte[59] ein solches System für *Machine Programming* im Jahr 2021.

[58] https://www.youtube.com/watch?v=D5VN56jQMWM

[59] https://youtube.com/watch?v=SGUCcjHTmGY

Das Beispiel belegt die modulare Entwicklung von künstlich intelligenten Systemen: Vorhandene Algorithmen und bereits trainierte Systeme können für neue Anwendungen eingesetzt werden. Um *Machine Programming* nicht nur mit geschriebener, sondern auch mit gesprochener Sprache umzusetzen, reicht es aus, die Schnittstellen (APIs) zu vorhandenen Systemen zu schaffen.

Damit findet ein sich akzelerierender Wissens- und Methodenaufbau statt. Der Zustand der Singularität, in dem eine Software der menschlichen Intelligenz in allen Bereichen überlegen ist, ist abzusehen. Allerdings ist es von der Substituierbarkeit der Produktionsfaktoren abhängig, ob damit zusätzlich eine ökonomische Singularität im Sinne eines unbegrenzten Wachstums einhergehen wird (Nordhaus, 2021).

2.2 Algorithmic Decision Making

Durch die modulare Entwicklung sowie die gemeinsame Nutzung von Daten und Netzinfrastrukturen ergeben sich zunehmende Skalenerträge für KI-Anwendungen (Vogelsang, 2021). Dies verbessert die Prognosefähigkeit, d. h. die Genauigkeit der individualisierten Vorhersage je Merkmalsträger. Damit werden automatisierte Entscheidungssysteme (Algorithmic Decision Making - ADM) möglich, d. h. die Skalierung bei gleichzeitiger Individualisierung von Entscheidungen wird erreicht.

Die ökonomischen Voraussetzungen sind gegeben, damit ADM-Systeme zunehmend verbreitet werden. Dabei sind verschiedene Abstufungen möglich (Datenethikkommission, 2019):

- Algorithmenbasierte Entscheidungen
- Algorithmengetriebene Entscheidungen
- Algorithmendeterminierte Entscheidungen (vollständig automatisiert).

ADM-Systeme werden zuerst in Bereichen eingesetzt, bei denen der Routineanteil an den menschlichen Entscheidungen hoch und deren Relevanz niedrig ist. Mit zunehmender technologischer Entwicklung werden zudem komplexere und/oder weitreichendere Entscheidungen durch Algorithmen getroffen werden.

2.3 Entscheidungen: Vorteile und Gefahren

Selbst wenn Menschen rational entscheiden, unterliegen sie häufig einer begrenzten (bounded) Rationalität, weil beispielsweise Informationen unvollständig sind oder ein Druck für eine rasche Entscheidung besteht. Dies führt zur Anwendung von Heuristiken oder der Entscheidung für die erstbeste Alternative, die die Mindestbedingungen erfüllt (Gigerenzer, 2001).

Verglichen mit den menschlichen Möglichkeiten können vernetzte Maschinen grundsätzlich viel mehr Daten in kürzerer Zeit zur Entscheidungsfindung einsetzen. Ihre Rationalität ist in diesem Sinne unbegrenzter als die menschliche Rationalität.

Allerdings kann aus diesem Vorteil nicht geschlossen werden, dass automatisierte Entscheidungen grundsätzlich richtig oder diskriminierungsfrei seien. Die folgenden Risiken können im Zusammenhang mit softwarebasierten Entscheidungen auftreten:

- Der Code ist fehlerhaft programmiert.
- Overfitting: Auch wenn sich ein Künstliches Neuronales Netz (und weitere KI-Methoden) gut an Trainingsdaten anpasst, bedeutet dies nicht, dass das Modell auch gut generalisiert, d. h. auch auf unbekannte Daten angewendet gute Ergebnisse erzielt (Bejani und Ghatee, 2021).
- Angriffe und Manipulation durch Dritte (*Hacking, Terror*).
- Es kommt zu Verzerrungen bei den KI-basierten Entscheidungen (*AI Bias*). Die Verzerrungen können ihre Ursachen in den Daten, im Algorithmus oder im Zusammenspiel mit dem Nutzer haben (Cowgill & Tucker, 2019; Mehrabi et al., 2022). Zudem gibt es unterschiedliche Ansätze, um ein faires Ergebnis zu definieren.
- Insbesondere beim Einsatz von neuronalen Netzen ist die Bedeutung von einzelnen Gewichten kaum nachvollziehbar. Die Vorhersage bzw. Empfehlung oder Entscheidung wird intransparent (*Black Box*) (Martini, 2019).
- Menschen geben Empfehlungen von Maschinen einen Vertrauensvorschuss (Sundar & Kim, 2019). Das bedeutet auch, dass Menschen durch automatisierte Systeme leichter manipuliert werden können.
- Es kommt zu negativen Arbeitsmarkteffekten. KI wird Berufe verändern, d. h. die Tätigkeiten werden neu zusammengesetzt. Allerdings gibt es aus den empirischen Untersuchungen der letzten zehn Jahre keine Hinweise auf eine steigende Arbeitslosigkeit aufgrund eines vermehrten KI-Einsatzes (Lane, 2021).

Insgesamt kann der vermehrte Einsatz algorithmendeterminierter Entscheidungen dazu führen, dass die **menschliche Autonomie** in Gefahr gerät. Daher entsteht Regulierungsbedarf.

2.4 Regulierungsbedarf

Die Methoden, die unter dem Begriff Künstliche Intelligenz zusammengefasst werden, gelten als Querschnittstechnologie, die alle betrieblichen Funktionen sowie die Beziehungen zwischen Staat, Bürgern bzw. Konsumenten und Unternehmen beeinflussen.

Die Politikfelder, auf die Künstliche Intelligenz wirkt, sind vielfältig und beinhalten auch die Handels-, Arbeitsmarkt- und Sozialpolitik (Agrawal et al., 2019). Wird die Schutzfunktion des Staates in den Vordergrund gerückt, liegt es nahe, die Technologie selbst zu regulieren bzw. den Anbietern einer Regulierung zu unterwerfen. Rambachan et al. (2021) zeigen theoretisch, dass durch ein Audit der Algorithmen deren

Vorhersage verbessert wird und es einfacher wird, Diskriminierung zu entdecken. Martini sortiert die Bausteine einer Regulierung von Algorithmen aus juristischer Perspektive nach ihrem relativen Zeitraum:

Ex ante	**Begleitend**	**Ex Post**
Transparenzanforderungen	Dokumentationspflichten	Haftung
Qualitätsanforderungen (Daten)	Informationspflichten	Sanktionen
Risikoabschätzungen	Hoheitliche Einsichtsrechte	Rechtsschutz
Zulassungskontrolle	Ergebniskontrolle	Monitoring

Regulierte Selbstregulierung	
Zertifizierung / Auditierung	Selbstverpflichtungen

Tabelle 1: Bausteine einer Algorithmenregulierung (Quelle: Martini, 2019, S. 339)

Der Regulierungsvorschlag der EU (European Commission, 2021b) stellt eine Kombination dieser Maßnahmen dar und wird im folgenden Kapitel vorgestellt.

3 Überblick über die Regulierungsvorschläge

3.1 Chronologie

Im Jahr 2018 hat mit öffentlichen Konsultationen, Anhörungen im Europäischen Parlament, einem White Paper und Expertengruppen die Arbeit an dem Regulierungsvorschlag begonnen. Dieser wurde im April 2021 veröffentlicht (European Commission, 2021c). Anschließend gab es ein weiteres Konsultationsverfahren („*have your say*“), bei dem 304 Rückmeldungen[60] von Verbänden, Unternehmen oder Einzelpersonen eingingen. Zum Jahreswechsel 2021/22 befand sich der Regulierungsentwurf in dem Abstimmungsverfahren zwischen den Europäischen Institutionen und den Mitgliedsländern.

3.2 Zusammenfassung der Inhalte

Die folgende Darstellung fasst die wichtigsten Punkte des Regulierungsentwurfs (European Commission, 2021b) zusammen. Die Artikelangaben beziehen sich auf dieses Dokument, sofern es nicht explizit anders angegeben wird:

Ziele und Geltungsbereich: Die vorrangige Absicht des vorliegenden Regulierungsvorschlags ist es, die Risiken, die mit KI einhergehen, einzudämmen. Das Ziel sei

[60] https://ec.europa.eu/info/law/better-regulation/have-your-say/initiatives/12527-Artificial-intelligence-ethical-and-legal-requirements_de

eine vertrauenswürdige KI („development of an ecosystem of trust by proposing a legal framework for trustworthy AI", European Commission, 2021b, S. 1). Im Anhang I wird geregelt, was darunter zu verstehen ist: Machine Learning Ansätze werden dort ebenso genannt wie „*statistical approaches*".

Reguliert werden die „provider", was in der deutschsprachigen Version mit „Anbieter" übersetzt wird. Nach Artikel 3 gehören zu den „providern" natürliche oder juristische Personen, die KI-Systeme entwickeln oder auf den Markt bringen.

Zudem enthält der Regulierungsentwurf mit den Reallaboren auch Elemente für eine Innovationsförderung (s. u.).

Risikoklassen: Drei Risikoklassen werden eingeführt: Nicht akzeptables Risiko, hohes Risiko und alle anderen (Art. 5-7 und Anhang III).

Nicht akzeptables Risiko: Systeme mit einem nicht akzeptablen Risiko werden verboten. Dazu gehört das Social Scoring durch staatliche Behörden ebenso wie Systeme, die Personen manipulieren, um sich oder anderen Schaden zuzufügen (Art. 5).

Hochrisiko-Anwendungen: Dies bezieht sich auf Anwendungen, die die Gesundheit, Sicherheit oder die Lebensläufe Einzelner beeinflussen. Beispiele: KI-Anwendungen für Wasser- oder Kraftwerke und Entscheidungen über Kredite oder Jobs. Innerhalb der Hochrisiko-Anwendungen wird zwischen KI-Systemen für Produkte, für die heute schon eine Prüfung auf Konformität durch Dritte notwendig ist (z. B. Spielzeug), und anderen Anwendungen unterschieden (Art. 6 und Anhang III).

Informationspflichten: Automatisierte *Chat-Bots* und *Deepfakes* müssen gekennzeichnet werden, auch wenn diese nicht als hochriskant eingestuft werden (Art. 52).

Vorschriften für Unternehmen: Die Vorschriften sind abhängig von der Risikoklasse, d. h. Anbieter von hochriskanten Anwendungen müssen die folgenden Pflichten erfüllen:

- Aufbau eines Risikomanagementsystems (Art. 9)
- Datenanforderungen (Art. 10)
- Technische Dokumentation und Aufzeichnungspflichten (Art. 11-12)
- Informations- und Transparenzpflichten gegenüber den Nutzern (Art. 13)
- „Human oversight" (Art. 14)
- Einführung eines Qualitätsmanagementsystems (Art. 17)
- Konformitätserklärung (Art. 19). Die Konformitätsbewertungsverfahren werden in vielen Fällen unternehmensintern durchgeführt. Sobald die Sicherheit von Produkten oder biometrische Systeme zur Identifikation betroffen sind, sollen Dritte (Zertifizierer) damit beauftragt werden.
- Den Marktüberwachungsbehörden Zugriff auf Daten, Schnittstellen und ggfs. Quellcode ermöglichen (Art. 64).

Behörden: Die Mitgliedsstaaten schaffen nationale Aufsichtsbehörden, die auch als Marktüberwachungsbehörden fungieren (Art. 59). Zur Beratung und zur Koordination zwischen den Ländern wird ein Europäischer Ausschuss für Künstliche Intelligenz gegründet (Art. 56). Zudem wird eine Datenbank für Hochrisiko-Systeme von der EU eingerichtet (Art. 60).

Strafen: Je nach Vergehen sind bis 30 Millionen Euro oder bis zu 6% des weltweiten Jahresumsatzes möglich (Art. 71).

Innovationen: Mit Reallaboren („*Regulatory Sandbox*“*)* soll, unter behördlicher Aufsicht, die Entwicklung von KI-Systemen erleichtert werden (Art. 52).

3.3 Regulierungsgrundsätze

Der Regulierungsvorschlag umfasst 85 Artikel zuzüglich Anhänge und schafft einen Regulierungsrahmen für KI als Querschnittstechnologie. Der Entwurf der EU-Kommission lässt sich durch die drei folgenden Regulierungsgrundsätze charakterisieren:

- Die Einteilung in Risikoklassen.
- Die überwiegende Eigenverantwortung der Unternehmen für Dokumentation, Qualitätsmanagement, Information und Kontrolle in Kombination mit ausgeprägten behördlichen Rechten für ein ex Post-Monitoring. Ausgenommen sind die Produkte, für die eine Konformitätserklärung durch Dritte notwendig ist.
- Der Aufbau von neuen öffentlichen Institutionen in Verbindung mit Strafandrohungen bei Verstößen, um die Regulierung zu koordinieren und durchzusetzen.

Im folgenden Kapitel steht die Einordnung der Lösungsansätze für die zuvor genannten KI-Probleme im Fokus. Im Anschluss erfolgt eine zusammenfassende Bewertung.

4 Ökonomische Einordnung der Lösungsansätze

4.1 Überblick

Dieses Kapitel greift die in Kapitel 2.3 aufgeführten Risiken auf und verbindet sie mit einer ökonomischen Theorie.

KI-Risiken	**Ökonomische Perspektive**
Programmierfehler	Risikomanagement (Kap. 4.2)
Mangelnde Generalisierbarkeit	
Angriffe durch Dritte	
Verzerrungen (*AI Bias*)	Asymmetrische Information / Prinzipal-Agent-Theorie (Kap. 4.3)
Mangelnde Transparenz (*Black Box*)	
Globaler Wettlauf Forschung & Entwicklung Marktstruktur	Innovations- und Wettbewerbspolitik (Kap. 4.4)
Gefährdung der menschlichen Autonomie	Entscheidungstheorie und Ethik (Kap. 4.5)

Tabelle 2: KI-Probleme und ökonomische Theorie

In den folgenden Kapiteln erfolgt eine kurze Bewertung des Regulierungsvorschlags aus der Perspektive der jeweiligen Theorie.

4.2 Risikomanagement

Im Standardansatz des Risikomanagements ergibt sich die Risikomatrix aus den Kombinationen von Schadenshöhe und Eintrittswahrscheinlichkeit. Handlungsbedarf besteht insbesondere bei hoher Eintrittswahrscheinlichkeit und hoher Schadenshöhe. Inwieweit Einzelrisiken, für die keine explizite Verteilungsfunktion bekannt ist, zu einem Gesamtrisiko aggregiert werden können, ist umstritten (Bao et al., 2019).

Im Kontext von KI-Anwendungen hängt die Eintrittswahrscheinlichkeit von Schadensereignissen u. a. von der Expertise des Entwicklers, dem Umfang der Qualitätssicherung, dem Umfang und der Qualität der Daten sowie der Nutzungsart ab. Live-Systeme, die während der Nutzung das Modell weiter trainieren, sind anfälliger für Fehlentwicklungen als abgeschottete Systeme, die in der Vergangenheit trainiert wurden und nur auf der Basis fester Gewichte eingesetzt werden.

Die Risikohöhe wird von dem konkreten Einsatzgebiet der KI-Anwendung bestimmt. Fehler in Infrastruktur- und Versorgungsunternehmen können schwerwiegendere Schadensfälle auslösen als eine KI-Anwendung im Bereich der unternehmerischen oder privaten Dienstleistungen.

Im Regulierungsvorschlag der EU (European Commission, 2021b) werden Risikoklassen gebildet, aber abweichend von dem Standardabsatz des Risikomanagements nicht nach Risikohöhe und Eintrittswahrscheinlichkeit differenziert. Dies führt zu einer Einstufung als „high risk“ für den Einsatz von KI in einem Wasserwerk ebenso wie für die Vorausauswahl im Bewerbungsprozess. Da sich die Risikoprofile bezüglich Eintrittswahrscheinlichkeit und Schadenshöhe aber unterscheiden, sollten sie auch unterschiedlichen Maßnahmen zur Risikovorbeugung unterliegen. Dies gilt ebenso für die Differenzierung zwischen Live- und statischen Systemen.

4.3 Prinzipal-Agent-Theorie

In seiner berühmten Veröffentlichung „*Markets for Lemons*“ schreibt George Akerlof: „*the difficulty of distinguishing good quality from bad is inherent in the business world*“ (Akerlof, 1970, S. 500). Dieses Zitat bezieht sich auf die zentrale Rolle von Produkteigenschaften. Nach dem Beispiel von Akerlof sei die Qualität von Gebrauchtwagen nicht unmittelbar von außen zu erkennen, da asymmetrische Informationen vorliegen. Dies ist ein Standardbeispiel aus der Prinzipal-Agent-Theorie.

Krafft et al. (2020) übertragen die Prinzipal-Agent-Theorie auf *Algorithmic Decision Making*. Bei den Agenten, die die Entwickler oder Betreiber von ADM-Systemen sowie die ADM-Systeme selbst sein können, bleiben Absichten, Systemeigenschaften, Wissen, Informationen oder die Entscheidungsprozesse selbst (*Black Box*) im Dunkeln. Die Erwartungen der Prinzipale, die sie als Nutzer an die Systeme haben, werden bei asymmetrischen Informationen möglicherweise enttäuscht oder sie werden Opfer von diskriminierenden Entscheidungen (Krafft et al., 2020).

Typische Maßnahmen bei asymmetrischer Information sind Signalling und Screening (Riley, 2001), d. h. eine Offenlegung der verborgenen Informationen. Im Zusammenhang mit ADM-Systemen weisen Cowgill & Tucker (2019) auf das Problem hin, dass eine vollständige Transparenz über die Entscheidungsparameter, die Entscheidungsqualität negativ beeinflussen könnten. Dies könnte beispielsweise der Fall sein, wenn Bewerber vorab die Parameter eines Vorauswahltools für die Stellenbesetzung kennen.

Wischmeyer sieht vor allem den Staat in der Verantwortung, um die *Black Box* „zu öffnen“. Er solle sich an Expertennetzwerken beteiligen, spezialisierte Agenturen gründen, Standardisierungsprozesse sowie den Rahmen für Zertifizierung und Auditing setzen (Wischmeyer, 2020).

Krafft et al. (2020) unterscheiden dagegen die Anwendungen danach, ob es auf die Ergebnisse, die Ziele oder die Prozesse ankommt oder dem Markt die Verantwortung für die Evaluation des ADM-Systems überlassen werden kann. Der letzte Fall wäre beispielsweise bei Modeempfehlungen gegeben. Dabei verursachen Empfehlungen,

die nicht mit den Vorlieben der Kundinnen oder Kunden kongruent sind, keine Schäden an Leib oder Seele und die verärgerten Nutzer können, Wettbewerb vorausgesetzt, einen anderen Anbieter auswählen. Die anderen Stufen werden danach unterschieden, welchen individuellen Schaden die Fehlentscheidungen anrichten. Bei medizinischen Diagnosen sollte ein umfassender Audit des ADM-Systems erstellt werden. Bei der Vorauswahl von Bewerbungen würde es reichen, nur die Eigenschaften des Systems zu testen, ohne die „*Black Box vollständig auszuleuchten*“ (Krafft et al., 2020, S. 13).

In dem Regulierungsvorschlag der EU werden die Vorschläge aus der ökonomischen Theorie teilweise ausgegriffen. Ein *Screening* wird für Hochrisiko-Anwendungen vorgeschrieben. Dies geschieht durch die Datenbank für Hochrisiko-Anwendungen und die Konformitätserklärungen. Ein *Signalling* muss für automatisierte Chat-Bots und Deepfakes stattfinden.

Nach dem Regulierungsvorschlag unterliegen KI-Systeme für Infrastruktur und für Anwendungen, die mit Entscheidungen über Kreditvergaben oder Stellenbesetzungen persönliche Lebenswege beeinflussen, denselben Regulierungsvorschriften. Dabei wird außer Acht gelassen, dass bei Infrastrukturanwendungen der Prinzipal im Sinne der Prinzipal-Agent-Theorie die Gesellschaft darstellt, während es sich bei der zweiten Gruppe von Anwendungen um Nutzerinnen und Nutzer handelt. Anstatt diese durch Testfunktionen, Datenspenden, Berichtsportalen o. ä. aktiv mit als Prinzipale einzubeziehen, verbleiben die Dokumentations-, Qualitätsmanagement-, Kontrollpflichten sowie Monitoringrechte bei Unternehmen und Behörden. Zudem ist es aus Sicht des Risikomangements nicht sachgerecht, beide Arten von Anwendungen zu einer Risikogruppe zusammenzufassen (s. o.).

4.4 Innovations- und Wettbewerbspolitik

Innovationen und Marktstruktur stehen in einem interdependenten Verhältnis. Innovationen können über Patente zu zeitweisen Monopolrenten führen. Andererseits entstehen bei intensivem Wettbewerb und ähnlicher Produktionstechnologie Innovationsanreize, um der Tendenz zu Nullgewinnen zu entkommen („*escape-competition-effect*“). Sobald die Technologie in einer Branche unausgewogen ist und es zum Beispiel einen Technologieführer gibt, wird stärkerer Wettbewerb auf der Produktebene zu geringeren Innovationsanreizen bei den Nachzüglern führen (*„Schumpeter-Effekt“*). Aghion et al. (2005) setzen den „*escape-competition*“- und den *„Schumpeter-Effekt“* so in eine Beziehung zueinander, dass sich zwischen Innovationen und Wettbewerbsintensität die Form eines nach unten geöffneten „U‘s“ ergibt.

Der Markt der Künstlichen Intelligenz besteht aus mehreren Ebenen:

- KI-Programmbibliotheken wie Tensorflow und Pytorch werden als Open-Source-Software von den Technologieführern Google und Facebook (Meta) entwickelt und zur Verfügung gestellt (Vogelsang, 2021).
- Rechnerkapazität kann über Cloud-Dienste gebucht werden. Häufig eingesetzte Dienste sind beispielsweise AWS (Amazon) und Azure (Microsoft), die zudem KI-Anwendungen anbieten.
- Die konkreten Software-Entwicklungen, Datenaufbereitungen und Implementierungen werden oft von mittelständischen Unternehmen angeboten, die häufig ebenfalls die KI-Programmbibliotheken und Cloud-Dienste der Marktführer einsetzen.

Damit gibt es unterschiedliche Teilmärkte, die durch Wissens-Spillover-Effekte miteinander verbunden sind. Ergänzend kommt es in den einzelnen Teilmärkten zu learning-by-doing Effekten, die sich auf den Code der Algorithmen ebenso wie auf vortrainierte Modelle beziehen. Die Markteintrittsbarrieren auf der Ebene der konkreten Software-Entwicklung sind also gering, auf den anderen beiden Teilmärkten dagegen eher als hoch einzuschätzen. Der Wettbewerb auf der Ebene der konkreten Software-Entwicklung sorgt zudem für eine Vielfalt in Bezug auf die eingesetzten Methoden.

Mit dem EU-Regulierungsvorschlag kommen erhöhte Compliance-Kosten auf die Anbieter und Entwickler von KI-Anwendungen zu. Compliance-Kosten verringern den Anreiz für kleine Unternehmen, in einen Markt neu einzutreten, wie in dem Zusammenhang mit Datenschutzregeln bereits gezeigt wurde (Campbell et al., 2015; Johnson & Shriver, 2020).

In einer Begleitstudie, die im Auftrag der EU-Kommission zum Regulierungsvorschlag verfasst wurde, werden die Auswirkungen auf die Marktstruktur offen angesprochen: „*A substantial volume of costs would stem from setting up internal QMS for companies, particularly SMEs. As the proposed regulation may reach all industries, many firms in lightly regulated industries might need to invest in a substantial one-off cost for market entry, effectively setting up an entry barrier and dampening market competition*“ (European Commission, 2021c).

Damit verringert sich ebenfalls der „competition-escape“ Effekt. Zudem verändert sich der Fokus in den Unternehmen, wenn die Erfüllung von Regulierungsauflagen an Bedeutung zunimmt. Die Monitoring-Rechte der Behörden sind weitgehend wie zum Beispiel der Zugriff auf die Trainings-, Validierungs- und Testdaten und ggfs. den Quellcode. Das Innovationstempo wird ggfs. verlangsamt und der Abstand zu den Branchenführern vergrößert sich, so dass auch auf dieser Seite weniger Innovationsanreize entstehen („Schumpeter-Effekt“). Auf der anderen Seite wird versucht, mit KI-Reallaboren („*Regulatory Sandboxes*“*)* die Innovationstätigkeit zu unterstüt-

zen (s. o.). In den Reallaboren dürfen unter behördlicher Aufsicht und unter bestimmten Bedingungen KI-Systeme auch mit personenbezogenen Daten entwickelt werden, die für andere Zwecke erhoben wurden. Später müssen die verwendeten Daten wieder gelöscht werden. Ob die Unternehmen die KI-Reallabore nutzen oder versuchen werden, die behördliche Aufsicht durch die Verlagerung von Entwicklungsaktivitäten in andere Länder zu umgehen, ist dabei eine offene Frage.

Im Ergebnis dieser Effekte wird sich die nach unten offene U-Kurve im Sinne von Aghion et al. (2005) durch den EU-Regulierungsvorschlag nach unten verschieben. Das bedeutet, zu jeder Marktstruktur wird es geringere Innovationsanreize geben.

Die EU-Kommission hält dem entgegen, dass Vertrauen („trustworthyness“) in KI die Nachfrage nach KI-Systemen befördern würde. Zudem könnten sich europäische Entwickler auf vertrauenswürdige KI-Systeme spezialisieren, was ihnen global einen Wettbewerbsvorteil einbringen würde (European Commission, 2021a). Zudem besitzt der internationale KI-Wettbewerb auch eine militärische Komponente (Horowitz, 2018).

Peukert et al. (2022) haben in Bezug auf die Datenschutzgrundverordnung festgestellt, dass sich auch Anbieter außerhalb der EU an die Regeln angepasst hätten. Sie beschreiben dies als extraterritoriale Reichweite der Regulierung. Aber sie stellen auch fest, dass Google nach der Einführung der Datenschutzgrundverordnung seinen Marktanteil im Bereich der Webtechnologien ausgebaut hatte.

Ein ähnliches Regulierungsergebnis, dass „trustworthy AI“ zu einem wichtigen Trend bei Unternehmensstrategie und -marketing wird und gleichzeitig der Markt der Entwickler konzentrierter wird, ist auch bei der Umsetzung des vorliegenden EU-Regulierungsvorschlags zur KI nicht auszuschließen.

4.5 Entscheidungstheorie und Ethik

Wiederholte menschliche Entscheidungen folgen einem bestimmten Schema:

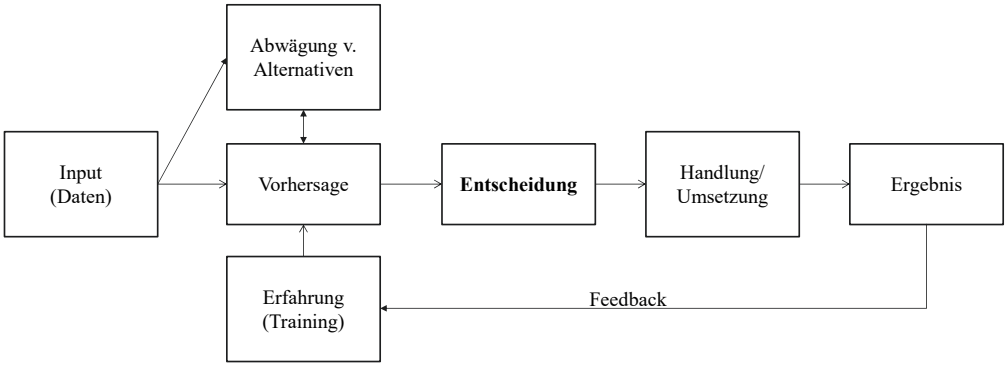

Abbildung 1: Anatomie einer Entscheidung

Darstellung auf Basis von Agrawal et al. (2018, S. 75), übersetzt und ergänzt

Grundsätzlich kann eine KI alle Entscheidungen berechnen, die nicht einer begrenzten (bounded) Rationalität (s. o.) unterliegen. Das Training eines künstlichen neuronalen Netzes ähnelt dem in der Abbildung gezeigten Schema einer menschlichen Entscheidung: Handlung/Umsetzung werden durch eine Verlustfunktion ersetzt, die im Ergebnis minimiert werden soll.

Der Vorteil einer KI-basierten Entscheidung liegt in ihrer strengen Rationalität, d. h. der Entscheidungsablauf per se ist, im Gegensatz zum Menschen, frei von psychologischen Einflüssen. Die angesprochenen Probleme (AI Bias, s. o.) liegen auf der Ebene der Daten und der Formulierung des Ziels, d. h. konkret der Verlustfunktion, die für die Abwägung von Alternativen benötigt wird.

Die Dokumentations- und Kontrollpflichten, die der EU-Regulierungsvorschlag für hochriskante Anwendungen vorsieht (s. o.), werden böswillige Programmierer (s. Prinzipal-Agent-Problematik) oder kriminelle Dritte allerdings nicht daran hindern, KI-Systeme für ihre eigenen Zwecke zu korrumpieren. Zudem unterscheidet der Regulierungsvorschlag nicht zwischen entkoppelten Systemen, bei denen der Trainingsprozess bereits beendet wurde und nur das Modell mit feststehenden Gewichten weiterverwendet wird, und lernenden Systemen, die kontinuierlich auf neue Daten zugreifen. Ebenso wird nicht unterschieden, ob die KI-Anwendungen selbständig Entscheidungen treffen oder zur Vorbereitung von Entscheidungen durch Menschen eingesetzt werden sollen. Die genannten Ansätze unterscheiden sich grundlegend in ihren Risikoprofilen, welche im Regulierungsvorschlag nicht abgebildet werden.

Der Fall, dass eine Maschine autonom Entscheidungen für Menschen übernimmt, ist aus ethischer Perspektive besonders relevant. Als die wichtigsten ethischen Prinzipien im Zusammenhang mit KI gelten (European Commission, 2019; Floridi & Cowls, 2021):

- Wahrung der menschlichen Autonomie
- Schadensvermeidung („prevention of harm“)
- Fairness
- Erklärbarkeit / Nachvollziehbarkeit („explicability“)

Die Wahrung der menschlichen Autonomie soll im Regulierungsvorschlag durch die Forderung nach „human oversight“ sichergestellt werden. Human oversight bezieht sich dabei auf den Provider, d. h. das KI zur Verfügung stellende Unternehmen muss jederzeit in der Lage sein, die Entscheidungen zu überschreiben oder das KI-System zu stoppen (European Commission, 2021b, Art. 14).

Eine menschliche Kontrolle durch die betroffenen Prinzipale (s. o.) ist in diesem Regulierungsvorschlag nicht vorgesehen. Allerdings billigt schon die Datenschutzgrundverordnung in Art. 22 den Betroffenen bei automatisierten Entscheidungen im

Einzelfall „mindestens das Recht auf Erwirkung des Eingreifens einer Person seitens des Verantwortlichen, auf Darlegung des eigenen Standpunkts und auf Anfechtung der Entscheidung“ zu (Europäische Union, 2018). Es bleibt aber die Frage offen, „ob der menschliche Entscheider nicht regelmäßig ohnehin dem vom KI-System ausgegebenen Ergebnis folgen wird“ (Ballestrem et al., 2020, S. 46). Zudem zweifeln Kaminski und Urban (2021) die Durchsetzbarkeit dieser Regel an[61].

Ohnehin spielen bei vielen KI Regulierungs- und Ethikvorschlägen die Betroffenen („wider stakeholders“) nur eine untergeordnete Rolle (Ayling und Champman, 2021). Noch weitreichender ist der Vorwurf, dass die ethische Debatte vielfach nur die Rolle eines Feigenblattes übernimmt.[62]

5 Bewertung

Methoden der Künstlichen Intelligenz verbinden Skalierung und Individualisierung. Ohne risikoadjustierten Preisen kann dies zu einer Übernutzung der Technologie über das wohlfahrtsoptimale Maß hinaus oder, umgangssprachlich formuliert, zu einer Sorglosigkeit im Umgang führen. Daher ist der Eingriff eines Regulierers aus mikroökonomischer Theorie angezeigt.

Mit dem vorgelegten Regulierungsvorschlag (European Commission, 2021b) wird der Versuch unternommen, die Querschnittstechnologie KI zu regulieren. Eine Bewertung aus Sicht der ökonomischen Theorie erfolgt anhand der drei Kriterien der Klarheit, Ziele und Effizienz der Maßnahmen.

Klarheit: Eine sich schnell verändernde Querschnittstechnologie bringt Abgrenzungsprobleme mit sich, welches bei der Spezifizierung der Methoden, die unter die

[61] “Failing to clarify a substantive basis for contestation potentially allows self-interested decision-makers to defang the right, making it useless in practice.“ (Kaminski und Urban, 2021, S. 2032)

[62] “Deviations from the various codes of ethics have no consequences. And in cases where ethics is integrated into institutions, it mainly serves as a marketing strategy. Furthermore, empirical experiments show that reading ethics guidelines has no significant influence on the decision-making of software developers. In practice, AI ethics is often considered as extraneous, as surplus or some kind of “add-on” to technical concerns, as unbinding framework that is imposed from institutions “outside” of the technical community.” (Hagendorff, 2020, S. 113)

“Regulation of any kind is strenuously opposed in the Valley” (Russell, 2019, S. 252)

„Nehmt der Industrie die Ethik weg!” Überschrift eines Beitrags von Thomas Metzinger, Professor für theoretische Philosophie an der Universität Mainz, in dem er “ethics washing” thematisiert und als Mitglied über die Vorgänge in der oben zitierten Expertenkommission der EU-Kommission zu “Trustworthy AI” berichtet. www.tagesspiegel.de vom 8. April 2019.

Regulierung fallen sollen, deutlich wird. Zudem bleibt unklar, ob das entwickelnde, das anwendende oder das zur Verfügung stellende Unternehmen (oder alle) für die Einhaltung der Regulierungsvorschriften verantwortlich ist.

Zielsetzung: Im Vordergrund steht die Eindämmung von Risiken, aber mit den Reallaboren enthält der Regulierungsvorschlag auch ein Element der Innovationsförderung.

Effizienz der Maßnahmen: Die Kombination aus eigenverantwortlich von den Unternehmen einzuhaltende Pflichten mit behördlichen ex-post Kontrollrechten und Strafen ist ein effizientes Regulierungsdesign. Die zusätzlichen Informationspflichten für Anwendungen, wie Chat Bots oder Deepfakes, entsprechen den Empfehlungen aus der Prinzipal-Agent-Theorie.

Die in Kapitel 4 genannte Kritik bezieht sich auf die Zusammenfassung von Risiken mit unterschiedlichen Charakteristiken (Kap. 4.2), die Nicht-Berücksichtigung von Prinzipalen (Kap. 4.3 und 4.5) und die negative Beeinflussung der Wettbewerbsintensität (Kap. 4.3). Tendenziell haben größere Unternehmen in Zukunft nicht nur die Vorteile, dass sich Anwendungen über eine größere Kundenzahl besser skalieren lassen und der größere Datenschatz die Qualität der KI-Modelle verbessert, sondern auch, dass für sie die Regulierungskosten leichter zu tragen sind. Die Durchschnittskosten erhöhen sich bei ihnen durch zusätzliche Compliance-Maßnahmen weniger stark als bei kleinen Unternehmen.

Das übergeordnete und langfristige Ziel der KI-Regulierung ist es, die Autonomie der Menschen zu bewahren. „Human oversight“ wird in dem vorliegenden Regulierungsvorschlag nur gefordert, um auf Seiten der Betreiber KI-Anwendungen zu überwachen. Es fehlt für Anwendungen die Festlegung eines „Human Veto“ auf Seiten der Nutzer, d. h. das einklagbare Recht der Nutzerinnen und Nutzer, eine automatisiert getroffene Entscheidung durch einen Menschen überprüfen zu lassen. Dabei sind diese die Prinzipale, deren Wohlfahrt durch asymmetrische Informationen negativ beeinflusst werden könnten. Die menschliche Autonomie wird nicht bereits dadurch gewahrt, dass Dokumentations-, Informations- und Kontrollmaßnahmen eingeführt werden.

6 Zusammenfassung

Künstliche Intelligenz verbessert die Produktivität und Qualität, senkt die Kosten und erhöht die Flexibilität von Unternehmen. Neben den Vorteilen aus unternehmerischer Sicht können die Methoden der KI mit neuen Problemen und Gefahren einhergehen. Dies können Programmierfehler ebenso wie Verzerrungen bei automatisierten Entscheidungen betreffen.

Um mögliche Risiken abzuwenden, hat die EU-Kommission im April 2021 einen Regulierungsvorschlag für die Querschnittstechnologie KI vorgestellt. In diesem Beitrag werden die Prinzipal-Agent-Theorie, das Risikomanagement, die Innovations- und Wettbewerbspolitik sowie die Entscheidungstheorie und Ethik auf KI bezogen. Die sich daraus ergebenden Empfehlungen werden mit dem Regulierungsvorschlag verglichen und aus Sicht der ökonomischen Theorie bewertet.

Dabei lässt sich die Sinnhaftigkeit einzelner Bausteine des Regulierungsvorschlags aus der Theorie bestätigen. Kritisch sind insbesondere die grobe Einteilung in Risikoklassen, die Wettbewerbswirkungen und die Vernachlässigung der Nutzerinnen und Nutzer als Prinzipale zu werten. Das oberste Ziel, langfristig die menschliche Autonomie zu wahren, wird mit dem vorliegenden Regulierungsvorschlag vermutlich nicht erreicht werden.

7 Literaturverzeichnis

Aghion, P., Bloom, N., Blundell, R., Griffith, R. & Howitt, P. (2005): Competition and Innovation: An inverted-U relationship. In: Quarterly Journal of Economics. Vol. 120 (2), S. 701–728.

Agrawal, A., Gans, J. & Goldfarb, A. (2018): Prediction Machines: The Simple Economics of Artificial Intelligence, Harvard Business Review Press, Boston, Massachusetts.

Agrawal, A., Gans, J. & Goldfarb, A. (2019): Economic Policy for Artificial Intelligence. In: Innovation Policy and the Economy. Vol. 19, S. 139–159.

Akerlof, G. A. (1970): The market for "lemons": Quality uncertainty and the market mechanism. In: Quarterly Journal of Economics. Vol. 84 (3), S. 488–500.

Ayling, J. & Champman, A. (2021): Putting AI Ethics to work: are the Tools fit for Purpose?. In: AI and Ethics. Abgerufen am 10.04.2022 von https://doi.org/10.1007/s43681-021-00084-x.

Ballestrem, J. Graf, Bär, U., Gausling, T., Hack, S. & von Oelffen, S. (2020): Künstliche Intelligenz: Rechtsgrundlagen und Strategien in der Praxis, Springer Gabler, Wiesbaden.

Bao, C., Wan, J., Wu, D. & Li, J. (2019): Aggregating Risk Matrices under a Normative Framework. In: Journal of Risk Research. Vol. 24 (8), S. 1–17.

Bejani, M. M. & Ghatee, M. (2021): A systematic review on Overfitting Control in shallow and deep Neural Networks. In: Artificial Intelligence Review, S. 1–48.

Campbell, J., Goldfarb, A. & Tucker, C. (2015): Privacy Regulation and Market Structure. In: Journal of Economics & Management Strategy. Vol. 24 (1), S. 47–73.

Cowgill, B. & Tucker, C. (2019): Economics, Fairness and Algorithmic Bias [Online], NBER Konferenzbeitrag. Abgerufen von https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3361280.

Datenethikkommission (2019): Gutachten der Datenethikkommission [Online], Berlin, Datenethikkommission der Bundesregierung. Abgerufen am 10.04.2022 von https://www.bmi.bund.de/SharedDocs/downloads/DE/publikationen/themen/it-digitalpolitik/gutachten-datenethikkommission.pdf.

Europäische Union (2018): Datenschutz-Grundverordnung DSGVO. Abgerufen am 10.04.2022 von https://dejure.org/gesetze/DSGVO/22.html.

European Commission (2019): Ethics Guidelines for Trustworthy AI. Abgerufen am 10.04.2022 von https://data.europa.eu/doi/10.2759/177365.

European Commission (2021a): Communication: Fostering a European approach to Artificial Intelligence. Abgerufen am 10.04.2022 von https://eur-lex.europa.eu/legal-content/EN/ALL/?uri=COM:2021:205:FIN.

European Commission (2021b): Proposal for a Regulation on the European Parliament and of the Council laying down harmonised rules on Artificial Intelligence (Artificial Intelligence Act). Abgerufen am 10.04.2022 von https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:52021PC0206.

European Commission (2021c): Study to Support an Impact Assessment of Regulatory Requirements for Artificial Intelligence in Europe - Final Report. Abgerufen am 10.04.2022 von https://op.europa.eu/.

Floridi, L. & Cowls, J. (2021): A Unified Framework of Five Principles for AI in Society, in Floridi, L. (Hrsg.), Ethics, Governance, and Policies in Artificial Intelligence, Cham, Springer International Publishing, S. 5–17.

Gigerenzer, G. (2001): The adaptive toolbox, in Bounded rationality: The adaptive toolbox; Report of the 84th Dahlem Workshop on Bounded Rationality: the Adaptive Toolbox, Berlin, 14-19.03.1999, Cambridge, Mass., MIT Press.

Hagendorff, T. (2020): The Ethics of AI Ethics: An Evaluation of Guidelines. In: Minds and Machines. Vol. 30, S. 99–120.

Horowitz, M. C. (2018): Artificial Intelligence, International Competition, and the Balance of Power. In: Texas National Security Review. Vol 1 (3), S. 37–57.

Johnson, G. & Shriver, S. (2020): Privacy & Market Concentration: Intended & Unintended Consequences of the GDPR, Marketing Sciences Institute. Abgerufen am 10.04.2022 von https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3477686.

Kaminski, M. E. & Urban, J. M. (2021): The Right to Contest AI. In: Columbia Law Review. Vol. 121 (7), S. 1957–2047.

Krafft, T. D., Zweig, K. A. & König, P. D. (2020): How to Regulate Algorithmic Decision-Making. In: Regulation & Governance (2022) (16), S. 119–136.

Lane, M. (2021): The impact of Artificial Intelligence on the labour market: What do we know so far?, OECD Publishing, Paris.

Martini, M. (2019): Blackbox Algorithmus – Grundfragen einer Regulierung Künstlicher Intelligenz, Springer, Berlin Heidelberg.

Mehrabi, N., Morstatter, F., Saxena, N., Lerman, K. & Galstyan, A. (2022): A Survey on Bias and Fairness in Machine Learning. Abgerufen am 10.04.2022 von https://arxiv.org/pdf/1908.09635.pdf.

Nordhaus, W. D. (2021): Are we approaching an Economic Singularity?: Information Technology and the Future of Economic Growth. In: American Economic Journal/Macroeconomics. Vol. 13 (1), S. 229–332.

Peukert, C.; Bechtold, St.; Batikas, M.; Kretschmer, T. (2022): Regulatory Spillovers and Data Governance: Evidence from the GDPR. In: Marketing Science. Abgerufen am 10.04.2022 von https://pubsonline.informs.org/doi/pdf/10.1287/mksc.2021.1339.

Rambachan, A., Kleinberg, J., Mullainathan, S. & Ludwig, J. (2021): An Econom-ic Approach to Regulating Algorithms, NBER Working Paper Series, No. 27111.

Riley, J. G. (2001): Silver Signals: Twenty-five years of Screening and Signaling. In: Journal of Economic Literature. Vol. 49, S. 432–478.

Russell, S. (2019): Human Compatible: Artificial Intelligence and the Problem of Control, Viking.

Sundar, S. S. & Kim, J. (2019): Machine Heuristic: When We Trust Computers More than Humans with Our Personal Information. Abgerufen am 10.04.2022 von https://doi.org/10.1145/3290605.3300768.

Turing, A. M. (1950 - Neuauflage 2021): Computing Machinery and Intelligence / Können Maschinen denken? (Englisch/Deutsch): Great Papers Philosophie, Reclam Verlag, Ditzingen.

Vogelsang, M. (2021): Datenskaleneffekte und Künstliche Intelligenz: Ein ökonomischer Blick auf die KI-Bibliotheken Tensorflow von Google und Pytorch von Facebook. In Budzinski, O., Haucap, J., Stöhr, A. & Wentzel, D. (Hrsg.), Zur Ökonomik von Sport, Entertainment und Medien, Berlin, De Gruyter Oldenbourg, 2021, S. 295–316.

Wischmeyer, T. (2020): Artificial Intelligence and Transparency: Opening the Black Box. In Wischmeyer, T. & Rademacher, T. (Hrsg.), Regulating artificial intelligence, Cham, Switzerland, Springer, S. 75–100.