Die EU als Trendsetter weltweiter KI-Regulierung: Der Kommissionsentwurf für eine KI-Verordnung der EU

Andreas Ebert, Indra Spiecker genannt Döhmann*

Der neue Entwurf der Europäischen Kommission zur Regulierung von künstlicher Intelligenz ist, wie schon die DSGVO, ein weltweit einzigartiger Aufschlag, eine neuartige Technologie begleitend zu regulieren und Nebenwirkungen ihres Einsatzes zu vermeiden. Der folgende Beitrag stellt die Systematik und wichtigsten Regelungen des Entwurfs vor und bietet eine erste Analyse und Einordnung in einen weiteren Kontext.

I. Einleitung

Am 21.4.2021 veröffentlichte die Europäische Kommission einen Verordnungsentwurf zu harmonisierten Regelungen für die künstliche Intelligenz[1] [KIVO-E], der im Englischen[2] den verheißungsvollen Titel „Artificial Intelligence Act" trägt. Er soll einen rechtlichen Rahmen für vertrauenswürdige künstliche Intelligenz (KI) schaffen.[3] Der Entwurf geht auf die im April 2018 vorgelegte Europäische KI-Strategie[4] der Kommission zurück und greift den risikobasierten Ansatz aus dem Weißbuch für künstliche Intelligenz[5] auf; er verfolgt den Grundsatz der Technologieneutralität mit einigen Spezifizierungen für einzelne Anwendungsbereiche und inkludiert diverse dynamische Elemente, mit denen der weiteren Technikentwicklung Rechnung getragen werden kann. Der Entwurf ist geprägt durch den Versuch eines Ausgleichs zwischen dem begründeten Schutzanspruch gegenüber einer neuartigen Technologie, deren Auswirkungen noch nicht erfassbar sind, und der grundrechtlich gebotenen Ermöglichung von Innovation.

Der KIVO-E unterzieht KI einem Regulierungsregime. Damit muss künftig präventiv geprüft werden, ob KI-Anwendungen besonders hohe Risiken für bestimmte Rechtsgüter bewirken können und je nach Befund Schutzvorkehrungen getroffen werden. Zu diesem Zweck unterscheidet der KIVO-E zwischen vier verschiedenen Risikostufen: inakzeptables Risiko, hohes Risiko und geringes sowie minimales Risiko, wobei die letzten beiden Stufen im Entwurf nicht weiter unterschieden werden.[6] KI mit inakzeptablem Risiko verbietet der Entwurf bis auf eine Ausnahme vollständig; Systeme mit hohem Risiko unterliegen hohen Voraussetzungen. Die Umsetzung von Anforderungen des Entwurfs bei KI mit geringem und minimalem Risiko erfolgt hingegen auf freiwilliger Basis.

---

* Andreas Ebert ist Wissenschaftlicher Mitarbeiter an der Goethe-Universität Frankfurt a.M. im Projekt RoboTrust des Zentrums für Verantwortungsbewusste Digitalisierung (ZEVEDI) sowie im Kompetenzzentrum für Angewandte Sicherheitstechnologie (KASTEL) am Karlsruher Institut für Technologie; Prof. Dr. Indra Spiecker genannt Döhmann ist Inhaberin des Lehrstuhls für Öffentliches Recht, Informationsrecht, Umweltrecht und Verwaltungswissenschaften an der Goethe-Universität Frankfurt a.M.

[1] Proposal for a Regulation of the European Parliament and of the Council Laying Down Harmonised Rules on Artificial Intelligence and Amending Certain Union Legislative Acts, COM(2021) 206 final.

[2] Die verwendeten deutschen Begriffe sind Übersetzungen der Verfasser.

[3] Begründung KIVO-E, S. 1.

[4] Künstliche Intelligenz für Europa, COM(2018) 237 final.

[5] Weißbuch zur künstlichen Intelligenz, COM(2020) 65 final.

[6] Vgl. Begründung KIVO-E, S. 12 f.

Die Verordnung lässt sich in vier Teile einteilen. Der erste (Kapitel I) umfasst in einem allgemeinen Teil Ziele, Anwendungsbereich und Definitionen. Der zweite Abschnitt enthält Verbote, Voraussetzungen und Verfahren für bestimmte KI-Systeme (Kapitel II, III und IV). Im dritten Teil finden sich Regelungen, die geeignete Rahmenbedingungen für konkrete KI schaffen sollen (Kapitel V, VI, VII, VIII, IX, X). Zuletzt folgen Schlussbestimmungen (Kapitel XI, XII).

II. Gegenstand und Anwendungsbereich des Verordnungsentwurfs

Art. 1 KIVO-E listet eine Reihe ambitionierter Ziele: harmonisierte Regeln für Inverkehrbringen und Inbetriebnahme von KI-Systemen in der EU, das Verbot bestimmter KI-Praktiken und Voraussetzungen für bestimmte KI-Systeme und deren Anbieter. Daneben verfolgt der Entwurf die Schaffung harmonisierter Transparenzregeln für bestimmte KI-Anwendungen, namentlich solcher, die zur Interaktion mit natürlichen Personen bestimmt sind, Systeme zur Emotionserkennung und biometrischen Identifizierung sowie KI-Systeme, die Bilder, Videos oder Sprachaufnahmen generieren oder manipulieren (so genannte „Deep Fakes“). Zudem sind Regeln zur Marktüberwachung und -beobachtung vorgesehen.

1. Adressaten

Die wesentlichen Adressaten der Verordnung sind sowohl Anbieter („provider“), also Entwickler von KI-Systemen, die diese unter eigenem Namen oder eigener Marke in Verkehr bringen oder in Betrieb nehmen, als auch Nutzer („user“), solange letztere die Systeme für berufliche Aktivitäten verwenden (vgl. Art. 3 Nr. 2, 3 KIVO-E). Damit setzt der KIVO-E – anders als die DSGVO – bereits beim Entstehungsprozess an und nimmt ausdrücklich auch den Entwickler in die Verantwortung, allerdings nicht den einzelnen Programmierer in größeren Projekten.[7] Erfasst sind von beiden Definitionen auch öffentliche Stellen, auch wenn diese teilweise, z.B. für Bußgelder, durch die Mitgliedstaaten vom Anwendungsbereich ausgenommen werden können.

2. Räumlicher und sachlicher Anwendungsbereich

Vom räumlichen Anwendungsbereich erfasst sind nach Art. 2 I KIVO-E Anbieter, die ein KI-System in der EU in Verkehr bringen oder in Betrieb nehmen, Nutzer von KI-Systemen, die sich in der EU befinden, sowie Anbieter und Nutzer solcher Systeme in Drittländern, deren Ausgabewerte in der EU benutzt werden.

Typisch für eine europäische Verordnung enthält Art. 3 KIVO-E Legaldefinitionen. Die wohl wichtigste Definition des KI-Systems findet sich in Art. 3 Nr. 1 KIVO-E. KI-Systeme sind hiernach Software, die mittels einer oder mehrerer Techniken oder Konzepte aus Anhang I entwickelt werden und für eine gegebene Reihe an vom Menschen definierten Zielen Ausgabewerte generieren kann, die aus Inhalten, Vorhersagen, Empfehlungen oder Entscheidungen bestehen können, die die Umgebung beeinflussen, mit der sie interagieren. Anhang I listet konkrete Ansätze wie das überwachte und verstärkte maschinelle Lernen, aber auch allgemeinere Konzepte wie „statistische Methoden“. Mit dieser weiten Definition samt Regelbeispielen umfasst der Entwurf auch Software, die wohl von Informatikern nicht als KI

[7] Vgl. auch EG 53.

bezeichnet würde. Ausschlaggebend für die Einordnung scheint hier, wie die verbotenen KI-Einsätze zeigen, die Möglichkeit der späteren Verwendung zu sein. Ähnlich wie die DSGVO verfolgt der KIVO-E insoweit einen technologieneutralen Ansatz,[8] allerdings mit einigen Spezifizierungen für bestimmte Technologien. Verdeutlicht wird dies auch durch die Ermächtigung der Kommission, Anhang I erweiternde Durchführungsakte nach Art. 4, 73 KIVO-E zu erlassen.

III. Verbotene KI-Praktiken

Art. 5 I KIVO-E spricht ein grundsätzliches Verbot für bestimmte Praktiken aus, die durch Missbrauch zu manipulativer sozialer Kontrolle führen können.[9] Der KIVO-E stellt sich damit sichtbar in die Tradition des Technikrechts, mittels präventiver Maßnahmen die unsicheren Entwicklungen einer neuen Technologie begleitend einzuhegen, wie dies z.B. auch der DSGVO zu eigen ist.[10] Anders als die DSGVO benennt der KIVO-E eine Reihe von konkreten Verwendungen von KI, die als besonders gefährlich eingestuft und in der Folge mit erheblichen Beschränkungen, zum Teil Verboten, belegt werden.

1. Verhaltensmanipulation

In Art. 5 I lit. a) KIVO-E werden KI-Systeme verboten, wenn diese unterschwellige Techniken jenseits des menschlichen Bewusstseins (engl. „*subliminal techniques beyond a person's consciousness*“) verwenden, um das Verhalten einer Person wesentlich zu beeinflussen. Abs. I lit. b) untersagt ferner KI-Systeme, die besonders schutzbedürftige Gruppen aufgrund ihres Alters oder körperlicher oder psychischer Behinderungen ausnutzen, um das Verhalten einer jener Gruppen zugehörigen Person wesentlich zu beeinflussen. Die Vorschrift knüpft an „Dark Patterns“ an, also Gestaltungsmuster, die psychologische Erkenntnisse verwenden, um Nutzer zu bestimmten Verhaltensweisen zu bewegen (bspw. das Anklicken von „alle akzeptieren“ im Cookie-Banner auf einer Website durch besondere Hervorhebung oder graphische Gestaltung).[11] Interessanterweise ist – wie bei allen Verboten – allerdings nicht die Entwicklung, sondern nur der Einsatz, Inbetriebnahme und Inverkehrbringen untersagt.

In der ersichtlichen Ausrichtung, Autonomie, Freiheitlichkeit und Selbstbestimmung der Bürger zu wahren und Manipulation und unterschwellige Beeinflussung zu verhindern, ist die weite Formulierung der Vorschrift grundsätzlich positiv einzuschätzen. Auch der Ansatz, schutzbedürftige Gruppen besonders einzubeziehen und somit informationelle Machtasymmetrien durch digitale Techniken aktiv zu thematisieren, ist bemerkenswert. Offen bleibt jedoch, warum gerade diese Gruppen ausgewählt wurden und nicht allgemein an die Machtasymmetrie oder Diskriminierungsanfälligkeit angeknüpft wird. Dies würde verbraucherschutzrechtlichen Systematiken eher entsprechen.

Stark geschwächt werden beide Verbote ohnehin durch die Voraussetzung, eine wesentliche Verhaltensänderung hervorrufen zu müssen. Worauf es dafür ankommen soll, wird nicht

---

[8] Vgl. EG 6; Begründung KIVO-E, S. 12.
[9] EG 15.
[10] *Hornung/Spiecker gen. Döhmann* in *Simitis/Hornung/Spiecker gen. Döhmann*, Datenschutzrecht, 1. Aufl. 2019, DSGVO Einleitung Rn. 244 ff.
[11] Hierzu *Brignull*, darkpatterns.org (abgerufen am 14.5.2021).

näher definiert. So genanntes Nudging[12] wird dem Wortlaut nach zumindest in Teilen noch erfasst sein, zumal darüber zum Teil die Möglichkeit erheblicher Verhaltensänderungen behauptet wird[13]. Einschränkend wirkt auch, dass das KI-System ausdrücklich zu diesem Zweck programmiert worden sein muss.[14] Anbieter und Nutzer solcher KI-Systeme werden sowohl Wesentlichkeit als auch Absicht gegebenenfalls abstreiten und die Verantwortung für den Einsatz zwischen sich verschieben können. Schließlich ist zu beachten, dass die Norm nur das Verbot von Systemen vorsieht, die zu körperlichen und psychologischen Schäden führen oder wahrscheinlich führen könnten. Wirtschaftliche oder materielle Beeinträchtigungen sind nicht erfasst, obwohl diese eine weitere problematische Fallgruppe darstellen.

2. Social Scoring

Besondere Aufmerksamkeit hat das Thema Social Scoring durch Entwicklungen in China erhalten.[15] Wohl in Reaktion darauf verbietet Art. 5 I lit. c) KIVO-E KI-Systeme, die der Bewertung oder Einstufung der Vertrauenswürdigkeit von natürlichen Personen anhand ihres sozialen Verhaltens oder anderer bekannter oder prognostizierter Eigenschaften über einen bestimmten Zeitraum dienen. Daneben muss der entstehende „social score" zu mindestens einer nachteiligen oder ungünstigen Behandlung von natürlichen Personen(-gruppen) entweder in sozialen Kontexten führen, die nicht im Zusammenhang mit dem Kontext der ursprünglichen Datenerhebung stehen, oder diese Behandlung muss ungerechtfertigt oder unverhältnismäßig zum sozialen Verhalten oder dessen Schwere sein. Den Nachweis darüber zu erbringen, dürfte schwierig sein.

Kritikwürdig ist auch, dass sich das Verbot nur auf Behörden (wenn auch im funktionalen Sinn) bezieht. Die besondere Machtposition bei Einsatz solcher Systeme im privaten Bereich, etwa bei Vertragsschlüssen oder Zugang zu Diensten, bleibt ungeregelt. Das Europäische Parlament hatte sich bereits im Vorfeld für das Verbot von Social Scoring für Behörden und stärkere Rechenschaftspflichten für Private stark gemacht.[16] Der KIVO-E greift das Anliegen nur noch im Kontext von Bonitätsprüfungen als KI mit hohem Risiko auf und belässt damit eine große Lücke.[17]

3. Biometrische Echtzeit-Fernidentifizierungssysteme

Zu den im Einzelnen geregelten Einsatzmöglichkeiten von KI gehört auch die Benutzung sogenannter biometrischer Echtzeit-Fernidentifizierungssysteme nach Art. 3 Nr. 36, die Art. 5 I lit. d) KIVO-E nur im Prinzip verbietet, dann aber eine Reihe von Ausnahmen vorsieht und den Betrieb bei Wahrung begleitender Schutzmaßnahmen doch zulässt. Dabei kann es sich nicht nur um Gesichtserkennung oder Fingerabdrücke handeln, sondern auch um bewegte Merkmale wie z.B. das spezifische Gangbild.[18] Dies ist nicht grundlos bereits auf heftige Kritik

---

[12] Hierzu *Kemmerer/Möllers/Steinbeis/Wagner*, Choice Architecture in Democracies: Exploring the Legitimacy of Nudging, 2017; kritisch *Mitchell*, Northwestern University law review 99 (2004), 3 ff.
[13] Hierzu *Thaler/Sunstein*, Nudge: Improving Decisions About Health, Wealth and Happiness, 6.
[14] Vgl. EG 16.
[15] Hierzu *Maamar*, CR 2018, 820 (821 f.).
[16] *Europäisches Parlament*, Entschließung vom 20.1.2021, A9-0001/2021, Nr. 72.
[17] Vgl. Anhang III Nr. 5 b).
[18] Ein System des chinesischen Staats erkennt auch bei versuchter Verzerrung, bspw. Hinken, Individuen am Gangbild, s. *Kang*, Chinese 'gait recognition' tech IDs people by how they walk, AP vom 6.11.2018, https://apnews.com/article/bf75dd1c26c947b7826d270a16e2658a (abgerufen am 14.5.2021).

gestoßen, da Profilbildung und Nachverfolgbarkeit mittels solcher Systeme ungeahnt intensiv ausfallen können.

Im Gegensatz zu den anderen Tatbeständen handelt es sich hier um ein Verbot mit sehr kleinem Anwendungsbereich und bezieht sich nur auf die tatsächliche Benutzung. Zudem wird der Einsatz des Systems nur in öffentlich zugänglichen Bereichen und zu Zwecken der Strafverfolgung eingeschränkt und erlaubt über die qualifizierte Öffnungsklausel in Art. 5 IV KIVO-E weitreichende Beschränkungen im nationalen Recht. Die Mitgliedstaaten können jedoch auch höhere Voraussetzungen für den Einsatz schaffen.

Nach Art. 5 I lit. d) KIVO-E ist die Benutzung zulässig, wenn sie für einen der genannten Zwecke strikt erforderlich ist, z.B. die gezielte Suche nach bestimmten potentiellen Verbrechensopfern wie vermisste Kinder, die Verhinderung eines Terrorangriffs oder einer Bedrohung für Leben und physische Sicherheit natürlicher Personen sowie bei Verdächtigen und Tätern einer Tat aus Art. 2 II des Rahmenbeschlusses über den Europäischen Haftbefehl.[19] Das Einfallstor für weitere weitreichende Überwachung ist damit aufgestoßen, auch wenn weitere Voraussetzungen nach Art. 5 II KIVO-E anknüpfen an Wahrscheinlichkeit und Umfang des Schadenseintritts, die Konsequenzen für Rechte und Freiheiten der Betroffenen und nicht näher bestimmte Schutzmaßnahmen und Bedingungen hinsichtlich Zeit, Raum und betroffener Personen einschließlich eines vorherigen Beschlusses eines Gerichts oder einer Behörde.

Es ist erkennbar, dass die Kommission sich bemüht, Zweifel am staatlichen Einsatz von biometrischen Identifizierungssystemen zu zerstreuen. Auch die Stellung unter den verbotenen Praktiken lässt politische Hintergründe vermuten, zumal tatsächlich gar kein Verbot vorliegt. Die Anforderungen im Entwurf sind jedoch bei Weitem nicht ausreichend, um die Bedenken hinsichtlich der Risiken zu entkräften; nicht nur für die Betroffenen, sondern für die Gesellschaft insgesamt. Die erlaubten Zwecke zum Einsatz eines solchen Systems sind zu unbestimmt. Ein Verdacht reicht bereits aus. Die zu treffenden Schutzmaßnahmen sind kaum konkretisiert; die gerichtliche Anordnung läuft bekanntermaßen oftmals leer.

IV. KI-Systeme mit hohem Risiko

Der Schwerpunkt der Verordnung liegt auf der Regulierung von KI-Systemen mit hohem Risiko, die nicht verboten, aber Anforderungen für Entwicklung und Einsatz unterstellt werden.

1. Einstufung als KI-System mit hohem Risiko

Ob ein KI-System einem hohen Risiko unterliegt, richtet sich nach Art. 6 I, II KIVO-E. Der Entwurf unterscheidet zwei sektorbezogene Ansätze, um das hohe Risiko festzustellen. Zum einen bezieht sich Art. 6 I KIVO-E auf KI-Systeme, die als Produkt oder Sicherheitskomponente eines Produkts bereits unter Unionsvorschriften in Anhang II fallen und zusätzlich einer so genannten Konformitätsbewertung durch Dritte nach der jeweiligen Unionsvorschrift unterzogen werden müssen. Genannt sind bspw. die Maschinenrichtlinie [20] oder die

[19] Rahmenbeschluss des Rates v. 13.6.2002 über den Europäischen Haftbefehl und die Übergabeverfahren zwischen den Mitgliedstaaten, 2002/584/JI.

[20] RL 2006/42/EG des Europäischen Parlaments und des Rates vom 17.5.2006 über Maschinen und zur Änderung der RL 95/16/EG.

Spielzeugsicherheitsrichtlinie [21] . Diesbezüglich wird offenbar das Manipulations- und Beeinflussungspotential als besonders problematisch gesehen.

Zum anderen verweist Art. 6 II KIVO-E auf Anhang III, der diverse besonders problematische Einsatzgebiete listet, etwa biometrische Identifikation; Zugang und Bewertung in Schul- und Berufsausbildung; Bewerbungen und Bewertungen in Arbeitsverhältnissen; Zugang zu Leistungen wie Sozialleistungen, Krediten und Notfalldiensten; Einsatz für Strafverfolgung, Migrationssteuerung sowie bei Gerichten. Problematisch ist, dass der Entwurf durch die Hintertüre der Einstufung als hohes Risiko eine Reihe dieser freiheitsrechtlich äußerst problematischen Anwendungen in die Legalität hievt, etwa den Einsatz von Lügendetektoren durch Strafverfolgungs-, Grenz- und Migrationsbehörden.

Der technologiegetriebenen Dynamik im KI-Bereich begegnet der KIVO-E dadurch, dass die Kommission auch Anhang III nach Art. 7, 73 KIVO-E adäquat ergänzen kann.

2. Anforderungen an KI-Systeme mit hohem Risiko

Ist ein KI-System nach Art. 6 KIVO-E mit hohem Risiko behaftet, muss es die Vorgaben der Art. 8-15 KIVO-E einhalten (Art. 8 I KIVO-E). Dabei handelt es sich im Wesentlichen um Verfahrensanforderungen, mittels derer das Risiko minimiert werden soll.

Dies umfasst zunächst ein zu dokumentierendes Risikomanagementsystem (Art. 9 I KIVO-E) mit konkreten Maßnahmen zur Risikobewältigung (Art. 9 II lit. d KIVO-E), die dem Stand der Technik entsprechen und dazu führen, dass das Rest- und Gesamtrisiko des KI-Systems als akzeptabel einzustufen ist (Art. 9 III S. 2, IV UAbs. 1 S. 1 KIVO-E). Erkennbar verlangt dies von den Nutzern eines KI-Systems eine beständige Kontrolle und Nachjustierung.

Datensätze zum Training von KI unterliegen ua angemessenen Datenverwaltungspraktiken, bspw. hinsichtlich Datenaufbereitung, -bewertung und -mängeln (Art. 10 II KIVO-E). Daneben müssen sie relevant, repräsentativ, fehlerfrei und vollständig sein (Art. 10 III S. 1 KIVO-E).

Ausnahmsweise erlaubt Art. 10 V KIVO-E die Verarbeitung besonderer Kategorien personenbezogener Daten nach Art. 9 I DSGVO, soweit es für die Feststellung und Korrektur von Bias unbedingt erforderlich ist und angemessene Schutzmaßnahmen für Grundrechte und -freiheiten der Betroffenen getroffen werden. Die Norm macht damit von der Generalklausel[22] in Art. 9 II lit. g DSGVO Gebrauch: Die Feststellung von Verzerrungen ist zumindest als ein öffentliches Interesse einzustufen; ob dieses jedoch auch erheblich ist, ist fraglich, zumal diese Erkenntnisse beim Anbieter verbleiben und für weitere Anwendungen genutzt werden können. Daneben stellt sich immer wieder die Frage, ob es aus technischer Perspektive möglich ist, diese Qualität zu garantieren. Erkennbar wird aus der Vorschrift, dass im Übrigen die DSGVO unverändert zu beachten ist, wenn es um die Erhebung, Speicherung und Zusammenführung personenbezogener Daten geht.

Daneben müssen die Systeme so transparent ausgestaltet sein, dass die Nutzer die Ausgabewerte interpretieren und verwenden können (Art. 13 I 1 KIVO-E) sowie die Möglichkeit einer effektiven Aufsicht und Ergebnisverifikation durch natürliche Personen

[21] RL 2009/48/EG des Europäischen Parlaments und des Rates vom 18.6.2009 über die Sicherheit von Spielzeug.
[22] *Petri* in: *Simitis/Hornung/Spiecker gen. Döhmann*, DSGVO Art. 9 Rn. 68.

besteht (Art. 14 KIVO-E). Hierzu müssen Bedienungshinweise beigefügt werden, welche die Informationen aus dem Katalog in Art. 13 III KIVO-E wie Kontaktinformationen und mögliche Fehlerquellen enthalten. Transparenz soll auch eine unionsweite Datenbank für KI-Systeme mit hohem Risiko schaffen (Art. 60 KIVO-E). Ferner sind Anforderungen an Genauigkeit, Robustheit und Cybersicherheit zu beachten: Die KI-Systeme müssen widerstandsfähig gegenüber Fehlern gestaltet, vor Zugriff durch unbefugte Dritte und Datenmanipulation geschützt und mit Lösungen wie Fail-Safe- oder Backup-Plänen abgesichert werden (Art. 15 III, IV KIVO-E). Dies steigert Anforderungen für die Anbieter in nicht unbeträchtlichem Ausmaß, stellt aber gleichzeitig sicher, dass die Anforderungen des KIVO-E von Anfang an mitgedacht werden müssen.

3. Verpflichtungen für Anbieter und Nutzer

Die Überwachung dieser Voraussetzungen obliegt gemäß Art. 16 II lit. a KIVO-E grundsätzlich dem Anbieter des Systems.[23] Als wohl wichtigste Anforderung muss er hierzu gem. Art. 19 I 1 KIVO-E das KI-System mit hohem Risiko dem Konformitätsbewertungsverfahren nach Art. 43 KIVO-E unterziehen. Der KIVO-E sieht je nach Art von KI-System verschiedene Verfahren zur Konformitätsbewertung vor: die interne Kontrolle durch den Anbieter selbst nach Anhang VI oder die Kontrolle durch benannte Stellen nach Anhang VII.

Für den Großteil der KI-Systeme mit hohem Risiko sieht der KIVO-E das Verfahren der internen Kontrolle vor. Dieses Instrument der Selbstregulierung durch Selbsteinschätzung und Bewertung von Qualitätsmanagement und technischer Dokumentation ist für Anbieter wenig eingriffsintensiv. Lediglich bei KI-Systemen zur biometrischen Identifikation und Kategorisierung müssen Anbieter diese Einschätzung durch benannte Stellen durchführen lassen und auch dann nur, wenn keine harmonisierten Standards oder gemeinsame Spezifikationen der EU bei der Entwicklung Anwendung finden. Diese Stellen werden durch die zuständigen Behörden nach den Art. 30 ff. KIVO-E benannt und müssen die Anforderungen nach Art. 33 KIVO-E erfüllen. Nach außen soll die Konformitätsbewertung demonstriert werden durch eine CE-Kennzeichnung und eine Konformitätserklärung (Art. 19 I, 48, 49 KIVO-E). Hier zeichnen sich erhebliche Vollzugsdefizite ab, denn Selbstregulierung ist immer nur so gut wie die begleitenden Kontrollen und diese sind angesichts der Komplexität von KI-Systemen kaum griffig ausgestaltet.

Der Anbieter ist darüber hinaus dazu verpflichtet, ein Qualitätsmanagementsystem nach Art. 17 KIVO-E zu implementieren und selbstständig Abhilfemaßnahmen zu treffen, sollte das KI-System nicht mehr den Anforderungen entsprechen (Art. 21 KIVO-E). Auch hiermit wird ein dynamisches Element konsequenter Nachverfolgung eingefügt, das allerdings gleichfalls vor allem auf Selbstregulierung setzt.

Nutzer sind zur Verwendung des Systems nach den Bedienungshinweisen verpflichtet. Dies umfasst die Umsetzung der Aufsichtspflichten nach Vorgabe des Anbieters (Art. 29 II KIVO-E). Zudem müssen sie, soweit sie die Kontrolle über Eingabewerte haben, sicherstellen, dass die benutzten Daten für den Verwendungszweck relevant sind (Art. 29 III KIVO-E). Insbesondere müssen sie die Informationen aus den Bedienungshinweisen nach Art. 13 KIVO-E zur

[23] Vgl. auch Art. 18, 20 KIVO-E.

Durchführung von Datenschutzfolgeabschätzungen gemäß Art. 35 DSGVO verwenden (Art. 29 VI KIVO-E).

Anbieter und Nutzer treffen ferner Melde- und Anzeigepflichten. Sollte das KI-System zu einem Risiko im Sinne von Art. 65 I KIVO-E führen, müssen Anbieter dies den zuständigen Behörden anzeigen (Art. 22 KIVO-E). Zudem müssen sie schwere Vorfälle und Fehlfunktionen melden, insoweit diese zu einer Verletzung unionsrechtlich geschützter Grundrechte führen, Art. 62 I UAbs. 1 KIVO-E. In den gleichen Fällen bestehen diese Pflichten auch für den Nutzer gegenüber dem Anbieter gemäß Art. 29 IV UAbs. 1 KIVO-E. Schwere Vorfälle sind nach Art. 3 Nr. 44 KIVO-E Vorfälle, die zum Tod oder schweren Schäden für Rechtsgüter wie Gesundheit, Eigentum, Umwelt oder zu Störungen im Betrieb von kritischer Infrastruktur führen können. Hier ist zu kritisieren, dass die grundlegende Problematik der informationellen Machtasymmetrie nicht mehr aufgegriffen wird, sondern auf klassische Rechtsgüter abgestellt wird. Zudem besteht keine Anzeigepflicht gegenüber den von Verarbeitungen durch KI-Systeme Betroffenen.

V. Transparenzvoraussetzungen für bestimmte KI-Systeme

Ferner sieht der Entwurf in Art. 52 KIVO-E Informations- und Kennzeichnungspflichten für bestimmte KI-Systeme vor. Diese Pflichten sind ergänzend oder auch separat von den Anforderungen an KI-Systeme mit hohem Risiko (vgl. Art. 52 IV KIVO-E) anwendbar. Da die Betroffenen solcher Systeme zumeist weder über die Fähigkeiten zur Überprüfung verfügen noch ihnen Möglichkeiten effektiver Kontrolle oder effektiven Rechtsschutzes – Beweislastumkehr etc. fehlen nämlich – zu Gebote stehen, laufen die Vorschriften ins Leere und täuschen eine menschliche Einflussnahmemöglichkeit vor, die gar nicht besteht.

1. Mensch-Maschine-Interaktion

KI-Systeme, die zur Interaktion mit natürlichen Personen bestimmt sind, müssen die jeweiligen Personen informieren, dass sie mit einem KI-System interagieren (Art. 52 I KIVO-E). Wie die Information erfolgen muss, steht nicht fest. Während dies bei einem Staubsaugroboter in den meisten Fällen offensichtlich sein wird, wird ein Chatbot nun deutlich gekennzeichnet sein müssen. Gemäß S. 2 sind KI-Systeme, die aufgrund gesetzlicher Grundlage zur Strafverfolgung oder -prävention verwendet werden, weitestgehend von der Informationspflicht ausgenommen. Dies ist angesichts der großen Gefährdungslage in diesem Bereich fragwürdig.

2. Emotionserkennung und biometrische Kategorisierung

Daneben müssen Nutzer von Emotionserkennungssystemen und Systemen zur biometrischen Kategorisierung die von diesen Systemen betroffenen Personen über den Betrieb des Systems informieren (Art. 52 II KIVO-E). Das Emotionserkennungssystem umfasst interessanterweise nach der Legaldefinition in Art. 3 Nr. 34 KIVO-E auch das Erkennen von Absichten aufgrund biometrischer Daten. Diese Norm ist insofern ein Novum, als dass Emotionen bislang nicht durch ausdrückliche Regelungen geschützt werden. Der Begriff der Absicht könnte auch allgemeine Verhaltensvorhersagen durch KI-Systeme umfassen. Da es sich bei beiden Varianten um die Verarbeitung personenbezogener Daten handelt, wird die Informationspflicht regelmäßig zu jenen aus den Art. 13 und 14 DSGVO hinzutreten.

3. Deep Fakes

Schließlich müssen Nutzer von so genannten „Deep Fakes“, also Bilder, Ton- oder Videoinhalte, die durch ein KI-System so verändert wurden, dass sie existierenden Personen, Objekten, Orten oder Ereignissen erkennbar ähneln, offenlegen, dass es sich um künstlich generierten oder manipulierten Inhalt handelt (Art. 52 III UAbs. 1 KIVO-E). Ausnahmen bestehen ua zur Ausübung der Meinungs-, Kunst- und Forschungsfreiheit (Art. 52 III UAbs. 2 KIVO-E). Deep Fakes stellen enorme Herausforderungen für die Gesellschaft dar, wenn sie manipulativ falsche Informationen verbreiten oder Teilnehmer des demokratischen Diskurses diskreditieren.[24] Mit dem Entwurf trägt die Kommission diesem Umstand in gewisser Weise Rechnung; die Regulierung bleibt jedoch zu unspezifisch. Kalifornien etwa verbietet die Generierung und Verbreitung von Deep Fakes von Politikern 60 Tage vor einer Wahl.[25]

VI. Vorschriften zur Schaffung von angemessenen Rahmenbedingungen für KI

Der KIVO-E umfasst verschiedene Maßnahmen zur Innovationsförderung und zur Durchsetzung der Anforderungen.

1. Regulatory Sandboxes

Art. 53, 54 KIVO-E sehen die Einführung von so genannten „Regulatory Sandboxes“ (Experimentierfeldern) in Form von kontrollierten Umgebungen für die Entwicklung und Testung von KI-Systemen vor. Die Kommission legt die genaueren Modalitäten und Voraussetzungen nach Art. 53 VI KIVO-E mittels Durchführungsakten fest. Die Einrichtung wiederum erfolgt durch Behörden der Mitgliedstaaten und den Europäischen Datenschutzbeauftragten. Die Teilnehmer der Sandboxes bleiben haftbar für Schäden, die Dritte durch den Einsatz innerhalb des Experimentierfeldes erleiden (Art. 53 IV KIVO-E). Bekannt ist dieses Konzept bereits aus dem Bereich innovativer Finanzdienstleistungen.[26]

Sehr problematisch gestaltet sich Art. 54 KIVO-E. Darin ist enthalten eine Zweckerweiterung für die Weiterverarbeitung von zuvor rechtmäßig erhobenen personenbezogenen Daten zu Zwecken der Entwicklung und Testung von KI-Systemen. Faktisch handelt es sich hier um eine gesetzliche Regelung zur Zweckvereinbarkeit gemäß Art. 6 IV DSGVO.[27] Die Anforderungen sind hoch: So muss unter anderem ein erhebliches öffentliches Interesse vorliegen, die Daten zur Einhaltung der Verordnung notwendig und nicht etwa durch anonymisierte oder synthetische Daten ersetzbar sein und alle personenbezogenen Daten nach Abschluss des Verfahrens gelöscht werden. Vor dem Hintergrund, dass personenbezogene Daten häufig auch im Nachhinein aus trainierten KI-Systemen rekonstruierbar sind,[28] ist diese Erlaubnis jedoch sehr fragwürdig.

---

[24] Hierzu *Chesney/Citron*, California Law Review 107 (2019), 1753 (1776 ff.).
[25] *Kalbhenn*, MMR-Aktuell 2019, 421493.
[26] Hierzu *Wendt*, CF 2020, 366.
[27] Zu den Voraussetzungen s. *Roßnagel*, in: Simitis/Hornung/Spiecker gen. Döhmann, DSGVO, Art. 6 Abs. 4 Rn. 22 ff.
[28] Vgl. *Veale/Binns/Edwards*, Phil. Trans. R. Soc. A. 376: 201800083; *Winter/Battis/Halvani*, ZD 2019, 489 (492 f.) m.w.N.

2. Governance und Kontrollinstitutionen

Gemäß Art. 59 I KIVO-E schaffen oder designieren die Mitgliedstaaten jene Behörden, die die Anwendung und Umsetzung der Verordnung überwachen; ua eine nationale Aufsichtsbehörde (Art. 59 II 1 KIVO-E).

Vergleichbar zum Europäischen Datenschutzausschuss (EDSA) sieht Art. 56 I KIVO-E die Einrichtung eines Europäischen KI-Ausschusses (EKIA) aus den zuständigen Aufsichtsbehörden aus den Mitgliedstaaten und dem Europäische Datenschutzbeauftragten (Art. 57 I KIVO-E) vor. Der EKIA hat im Wesentlichen eine Beratungsfunktion. Hierzu veröffentlicht der EKIA, wie der EDSA, Empfehlungen und Stellungnahmen und sammelt Erfahrungen aus den Mitgliedstaaten (Art. 58 KIVO-E). Anders als für den EDSA ist aber kein verbindliches Kohärenzverfahren vorgesehen, sodass absehbar konfligierende und widersprüchliche Entscheidungen der nationalen Behörden und strategisches Verhalten der KI-Industrie zu beobachten sein werden. Im schlimmsten Fall kann sich dies zu einem Unterbietungswettbewerb entwickeln.

Für KI-Systeme ohne hohes Risiko sollen Anbieter zur Erstellung von Verhaltenskodizes („codes of conduct“) hinsichtlich der Einhaltung von Voraussetzungen des KIVO-E und der Erreichung von Zielen wie Umweltverträglichkeit und Barrierefreiheit auf freiwilliger Basis ermutigt werden (Art. 69 I KIVO-E).

Eine Besonderheit bilden Vorschriften zur so genannten Marktüberwachung; gemäß Art. 63 I KIVO-E ist die europäische Marktüberwachungsverordnung[29] entsprechend anwendbar. Die Marktüberwachungsbehörde, üblicherweise identisch mit der Aufsichtsbehörde, darf in diesem Zusammenhang unter anderem auf alle Trainings-, Validierungs- und Testdatensätze und unter Begründung auf den Quellcode des KI-Systems zugreifen (Art. 64 I, II KIVO-E) und Testungen durchführen (Art. 64 V KIVO-E).

Bei einem riskanten KI-System kann die Marktüberwachungsbehörde die Einhaltung der Anforderungen des KIVO-E überprüfen (Art. 65 II UAbs. 1 S. 1 KIVO-E). Ein Risiko in diesem Sinne stellt ein KI-System dar, das Gesundheit, Sicherheit oder Grundrechte stärker beeinträchtigen kann, als es im Verhältnis zu seiner Zweckbestimmung oder seiner nach vernünftigem Ermessen vorhersehbaren Verwendung als vernünftig und vertretbar gilt (vgl. Art. 65 I KIVO-E). Gegebenenfalls kann die Behörde den Betreiber zu Maßnahmen verpflichten, um die Konformität mit dem KIVO-E herzustellen oder das System vom Markt auszuschließen oder zu entfernen (Art. 65 II UAbs. 2 KIVO-E).

Bei den Bußgeldvorschriften finden sich Parallelen zur DSGVO: Ein Verstoß gegen den KIVO-E kann mit einem Bußgeld bis zu 20.000.000 Euro oder 4% des weltweiten Jahresumsatzes eines Unternehmens geahndet werden (Art. 71 IV KIVO-E). Bei Verstößen gegen die verbotenen KI-Praktiken in Art. 5 KIVO-E oder die Vorschriften zur Datenqualität in Art. 10 KIVO-E beträgt der Rahmen sogar bis zu 30.000.000 Euro oder 6 % des weltweiten Jahresumsatzes (Art. 71 III KIVO-E). Damit schafft die Kommission eindrucksvolle Anreize zur Einhaltung der Vorschriften. Gemäß Art. 71 VII KIVO-E können die Mitgliedstaaten jedoch entscheiden, inwieweit

[29] VO (EU) 2019/1020 des Europäischen Parlamentes und des Rates vom 20. Juni 2019 über Marktüberwachung und die Konformität von Produkten sowie zur Änderung der Richtlinie 2004/42/EG und der Verordnungen (EG) Nr. 765/2008 und (EU) Nr. 305/2011.

öffentlichen Stellen Bußgelder auferlegt werden können. Die Erfahrungen mit der DSGVO[30] lassen befürchten, dass der Verordnung mit Blick auf staatlichen Einsatz von KI der Wind aus den Segeln genommen werden wird – obwohl der chinesische Citizen Score eindrücklich zeigt, wie wichtig eine Kontrolle staatlicher KI ist.

VII. Fazit

Dass Informationen und Informationstechnologie in den darauf basierenden Entscheidungen oftmals nicht erkennbar und für den Einzelnen auch nicht kontrollierbar oder konterkarierbar sind, wird im Prinzip aufgegriffen. Viele Ansätze des KIVO-E sind sinnvoll, teilweise fehlt den Anforderungen jedoch die Schärfe. Unklare Definitionen führen in Rechtsunsicherheit und einen Wettbewerb, in dem sich nicht unbedingt Qualität, Nachhaltigkeit und Verantwortungsbewusstsein durchsetzen.

Insofern ist der sektorbezogene Ansatz des Entwurfs durchaus ein probates Mittel, um Rechtssicherheit zu erreichen. Die Liste der verbotenen KI-Anwendungen ist ein erster Schritt, entbehrt aber einer leitenden Systematik. Die Rücksicht auf wirtschaftliche Interessen geht wiederum oft zu weit. Es ist geradezu in der Definition des KI-Systems mit hohem Risiko angelegt, dass Rechtsverletzungen zu hohen und irreversiblen Benachteiligungen und negativen Konsequenzen der davon Betroffenen führen. Daher ist nicht überzeugend, dass die Anbieter solcher Systeme in den allermeisten Fällen die zentrale Konformitätsbewertung ohne externe Kontrolle durchführen dürfen: Der Bock wird einmal mehr zum Gärtner gemacht.

Irritierend ist außerdem, dass es keine Rechte für diejenigen gibt, die von KI beurteilt und gesteuert werden, obwohl der KIVO-E behauptet, das Individuum schützen und das Vertrauen der Menschen wecken zu wollen. Auch die Rechtsdurchsetzung wird nicht gestärkt. Es fehlen Beweislastumkehrungen, pauschalierte Schadenssummen und Kausalitätserleichterungen. Ohnehin ist die Vollzugsseite wenig weitsichtig geregelt; aus den Fehlern des Datenschutzrechts hat man wenig gelernt. Abstimmungsmodalitäten zwischen den Aufsichtsbehörden und effektive Vollstreckung angesichts der ohnehin schwierigen technischen Rahmenbedingungen fehlen.

Zudem sind einige Wertungen des Entwurfs wenig überzeugend. Große Teile der Zivilgesellschaft fordern etwa das Verbot der biometrischen Fernidentifizierung wegen deren weitreichender Chilling Effects für die Ausübung von Freiheitsrechten. Der KIVO-E öffnet den Staaten jedoch das Tor für deren Einsatz. Daneben sind viele der KI-Systeme mit hohem Risiko auch in hohem Maße umstritten. Emotionserkennungssysteme von Privaten fallen nach aktuellem Stand überhaupt nicht unter die Kategorie des hohen Risikos. Hier bleibt der Entwurf deutlich hinter den Erwartungen an einen europäischen Dritten Weg der Freiheitlichkeit zurück. Letztlich ist zu hoffen, dass im Trilogverfahren noch im Sinne eines echten Technikrechts mit Weitsicht präziser reguliert wird und auch weitere Mechanismen aus der DSGVO übernommen werden, etwa Verbandsklagerechte, immaterieller Schadensersatz und effektive Vollzugsmechanismen.

[30] Deutschland hat von der entsprechenden Klausel in der DSGVO nicht Gebrauch gemacht, *Boehm* in Simitis/Hornung/Spiecker gen. Döhmann, DSGVO Art. 83 Rn. 55.