



Stellungnahme
des Kommissariats der deutschen Bischöfe – Katholisches Büro in Berlin –
zur
Verordnung des Europäischen Parlaments und des Rates zur Festlegung
harmonisierter Vorschriften für Künstliche Intelligenz (Gesetz über
Künstliche Intelligenz) und zur Änderung bestimmter Rechtsakte

I. Grundsätzliche Bewertung

Das Kommissariat der deutschen Bischöfe – Katholisches Büro in Berlin – begrüßt die Vorlage des Verordnungsvorschlags der Kommission [im Folgenden: KOM] zur Festlegung Harmonisierter Vorschriften für Künstliche Intelligenz [im Folgenden: KI] und zur Änderung bestimmter Rechtsakte der Union [im Folgenden: KI-VO-E].

Die Beziehung zwischen Mensch und KI-System kann komplexe sowohl positive als auch negative Auswirkungen auf das menschliche Individuum, aber auch auf die Gesellschaft als Ganzes und die Umwelt haben. Auch die KOM geht in der Begründung ihres Verordnungsvorschlags¹ auf die individuelle und auch die gesellschaftliche Dimension von KI ein, wobei sie jedoch bei der Umsetzung des Entwurfs selbst der letztgenannten Dimension letztlich nicht hinreichend Beachtung schenkt.

Papst Franziskus wies bereits im Kontext der Veröffentlichung des „Rome Call for AI Ethics“ der Päpstlichen Akademie für das Leben im Jahr 2020 darauf hin, dass neue Technologien ein „Geschenk Gottes“² sind, das „gute Früchte“³ tragen kann, aber auch Risiken birgt. Der „Rome Call for AI Ethics“ ruft daher dazu auf, dass „KI-Systeme [...] so konzipiert, gestaltet und implementiert werden [müssen], dass sie dem Menschen und seiner Umwelt dienen und sie schützen.“⁴ KI soll dazu beitragen, „Lebensbedingungen zu schaffen (sowohl gesellschaftlich als auch persönlich), die es sowohl der Gemeinschaft als auch den einzelnen Mitgliedern ermöglichen, sich so weit wie möglich zu entfalten.“⁵

Die Kirche sieht es daher als wichtige Errungenschaft an, dass die KOM in dem KI-VO-E ebenfalls den Menschen selbst ins Zentrum stellen will: Die KI sei „ein Instrument, das als positive Kraft für die Gesellschaft im Dienst der Menschen steht und das letztlich zu einem größeren Wohlbefinden der Menschen beiträgt“⁶. Hiermit grenzt sich der Ansatz des KI-VO-E von dem „von großen Digitalkonzernen und kommerziellen Interessen

¹ COM(2021) 206 final (deutsche Fassung), S. 1.

² Redemanuskript v. Papst Franziskus zum „Rome Call for AI Ethics“, 2020.

³ Redemanuskript v. Papst Franziskus zum „Rome Call for AI Ethics“, 2020.

⁴ „Rome Call for AI Ethics“, 2020.

⁵ „Rome Call for AI Ethics“, 2020.

⁶ COM(2021) 206 final (deutsche Fassung), S. 1.

geprägten amerikanischen Modell⁷ und dem „eher kollektivistisch orientierten, staatlich geprägten chinesischen Ansatz“⁸ ab.⁹

Dem Gedanken folgend, dass das KI-System im Dienst des Menschen steht, nennt die KOM in der Vorschlagsbegründung als erstes Ziel des KI-VO-E den Schutz des Menschen in Form der Gewährleistung der bestehenden Grundrechte und der Werte der Union.¹⁰ Wir bedauern jedoch, dass dieses Ziel nicht konsequent umgesetzt wurde. Nicht alle KI-Systemen, die in besonders gefährdeten Bereichen, wie etwa demokratischen und rechtsstaatlichen Prozessen, dem Schutz der Umwelt oder der Verhinderung von Diskriminierung Anwendung finden können, werden erfasst. Auch werden weiterhin KI-Systeme nach dem KI-VO-E erlaubt, die nach hiesiger Auffassung untersagt werden sollten. Es gilt daher Lösungsansätze zu finden, die gewährleisten, dass die KI-Systeme „wirklich der gesamten ‚Menschheitsfamilie‘ [...] [dienen], wobei die jedem Mitglied innewohnende Würde und die natürliche Umwelt zu achten sind und die Bedürfnisse der Schwächsten berücksichtigt werden müssen“¹¹.

Vor diesem Hintergrund nimmt das Kommissariat der deutschen Bischöfe zu einigen ausgewählten Artikeln des KI-VO-E Stellung.

II. Begründung im Einzelnen

1. Regelungstechnik und Gesetzssystematik

Der KI-VO-E unterscheidet mit KI verbundene verbotene Praktiken (Art. 5 KI-VO-E), Hochrisiko-KI-Systeme (Art. 6 f. KI-VO-E) und KI-Systeme mit geringem Risiko (Art. 52 ff. KI-VO-E), die auf dem Binnenmarkt in Verkehr gebracht oder in Betrieb genommen werden. Art. 5 Abs. 1 KI-VO-E koppelt das Verbot an die Konsequenzen, die sich aus KI-Praktiken ergeben können. Art. 5 Abs. 2 KI-VO-E beschreibt mit KI verbundene Praktiken, die in der Regel mit schwerwiegenden Grundrechtseingriffen oder -verletzungen einhergehen, und verbietet sie. Art. 6 KI-VO-E fokussiert sich nach Erwägungsgrund 32 auf KI-Systeme, die „aufgrund ihrer Zweckbestimmung ein hohes Risiko bergen, die Gesundheit und Sicherheit oder die Grundrechte von Personen zu schädigen (...)“. Im Anhang III, auf den Art. 6 Abs. 2 KI-VO-E verweist, werden in bestimmten Bereichen verwendete KI als Hochrisiko-KI definiert. Die KOM kann hier nach Art. 7 Abs. 1 KI-VO-E unter den in Art. 7 Abs. 2 KI-VO-E bestimmten Voraussetzungen per delegiertem Rechtsakt KI-Systeme hinzufügen und somit als Hochrisiko-KI qualifizieren. KI-Systeme, die nach den Kriterien des Art. 6 Abs. 2 KI-VO-E i.V.m. Annex III als hochriskant zu qualifizieren sind, müssen den in Kapitel 2 normierten erhöhten Anforderungen für ein Inverkehrbringen und eine Inbetriebnahme in den Binnenmarkt genügen. Diesen Anforderungen unterliegen KI mit geringem Risiko nicht. Manche KI mit geringem Risiko hingegen müssen maximal nur geringe Transparenzpflichten erfüllen. Unabhängig davon steht es den Anbietern von KI-Systemen mit geringem Risiko frei, die Systeme nach Verhaltenskodizes zu gestalten, die

⁷ Die deutschen Bischöfe, Kommission für gesellschaftliche und soziale Fragen, Heft Nr. 52, Europa ist es Wert - Impulse der Bischöflichen Arbeitsgruppe Europa, 15. November 2021, S. 45.

⁸ Die deutschen Bischöfe, Kommission für gesellschaftliche und soziale Fragen, Heft Nr. 52, Europa ist es Wert - Impulse der Bischöflichen Arbeitsgruppe Europa, 15. November 2021, S. 46.

⁹ Vgl. ebenso vom Ansatz: Die deutschen Bischöfe, Arbeitshilfe, Heft Nr. 288, Medienbildung und Teilhabegerechtigkeit, Impulse der Publizistischen Kommission der Deutschen Bischofskonferenz zu den Herausforderungen der Digitalisierung, 29. September 2016, S. 8.

¹⁰ COM(2021) 206 final (deutsche Fassung), S. 3, 12.

¹¹ „Rome Call for AI Ethics“, 2020.

durch sie selbst oder durch Interessenvertretungen dieser Anbieter aufgestellt wurden (Art. 69 Abs. 1, 2 KI-VO-E).

Grundsätzlich orientieren sich damit die Anforderungen, die der KI-VO-E an ein KI-System stellt, an dem Risiko dieses KI-Systems, Gesundheit und Sicherheit der einzelnen Person oder die Grundrechte zu schädigen, wobei die Schwere des Schadens und die Wahrscheinlichkeit seines Auftretens beachtet werden müssen. Das Katholische Büro begrüßt diesen Ansatz, da auf diese Weise Grundrechte weitestmöglich geschont werden. Wir haben jedoch gesetzssystematische Bedenken. Diese beziehen sich darauf, dass die in Art. 6 Abs. 1 und Abs. 2 KI-VO-E i.V.m. Anhang III aufgezählten Hochrisiko-KI-Systeme – anders als die in Art. 5 KI-VO-E aufgeführten verbotenen KI-Praktiken – grundsätzlich erlaubt sind und nur unter einen Zulassungsvorbehalt nach den Art. 8 ff. KI-VO-E gestellt werden. Dabei erfassen beide Absätze des Art. 6 KI-VO-E abschließende Aufzählungen spezifisch in Annex II oder in Annex III beschriebener KI-Systeme.

Dass die in dem KI-VO-E aufgezählten Hochrisiko-KI-Systeme gesetzssystematisch erlaubt sind und nur unter Zulassungsvorbehalt stehen, erscheint in zweierlei Hinsicht problematisch:

Zum einen besitzen einige der in Annex III genannten KI-Systeme tatsächlich ein so hohes Potenzial zur Schädigung von Grundrechten, dass es überzeugender erscheint, sie als „verboten“ im Sinne des Art. 5 KI-VO-E einzustufen als sie „lediglich“ als Hochrisiko-KI mit Zulassungsvorbehalt zu führen (hierzu ausführlich unter Punkt 4).

Zum anderen werden durch die abschließende Aufzählung der Hochrisiko-KI in Art. 6 KI-VO-E einige bereits existente oder zukünftig existierende KI-Systeme nicht von dem KI-VO-E umfasst, obwohl bei ihnen möglicherweise ebenfalls eine hochriskante Gefährdung gegeben ist. Zwar ist die KOM nach Art. 7 Abs. 1 KI-VO-E aufgrund eines delegierten Rechtsaktes zur Änderung der Liste in Annex III ermächtigt. Sie kann aber nur neue Hochrisiko-KI-Systeme den Bereichen des Annex III "hinzufügen"; die bereits genannten Bereiche kann sie nicht ex-post ändern. Die Verpflichtung zur Überprüfung, ob eine Änderung des Annex III erforderlich ist, besteht nach Art. 84 KI-VO-E darüber hinaus nur einmal jährlich. Infolgedessen werden neue oder neu erkannte Hochrisiko-KI nicht zeitnah in den Katalog des Annex III aufgenommen werden können. Während dieser Zeit sind diese KI-Systeme dann nach der Gesetzssystematik des KI-VO-E nicht nur erlaubt, sondern unterliegen nicht den Zulassungsvoraussetzungen der Art. 8 ff. KI-VO-E. Diese Rechtslage erscheint angesichts der mit Hochrisiko-KI verbundenen, potenziell tiefen Grundrechtseingriffe nicht zuletzt unter Berücksichtigung des Vorsorgeprinzips nicht akzeptabel.

Vor diesem Hintergrund wäre es aus kirchlicher Sicht sinnvoller gewesen, das Inverkehrbringen, die Inbetriebnahme und die Verwendung von Hochrisiko-KI unter ein Verbot mit Erlaubnisvorbehalt zu stellen. Ohne eine solche Änderung in der Gesetzssystematik des KI-VO-E hält das Katholische Büro es aber wenigstens für erforderlich, einen Auffangtatbestand in Art. 6 KI-VO-E – bspw. in Form eines Abs. 3 – einzufügen, mit dem noch nicht identifizierte oder zukünftig existente Hochrisiko-KI-Systeme erfasst und so den Anforderungen der Art. 8 ff. KI-VO-E unterworfen werden. Darüber hinaus sollte der KOM eine flexiblere und unkompliziertere Anpassungsmöglichkeit der Bereiche des Annex III nach Art. 7 Abs. 1 KI-VO-E ermöglicht und eine Bewertung und Überarbeitung des Annex III nach Art. 84 Abs. 1 KI-VO-E in einem Turnus von vielleicht sechs Monaten auferlegt werden.

Unter Einbeziehung dieser Änderungen könnte die KOM auch ihrem selbst in Erwägungsgrund 71 aufgeführten Ziel nachkommen, „einen (...) zukunftssicheren und gegenüber Störungen widerstandsfähigen Rechtsrahmen sicherzustellen“.

2. Anwendungsbereich, Art. 2 KI-VO-E

Nach Art. 2 Abs. 1 KI-VO-E erstreckt sich der Anwendungsbereich des Verordnungsentwurfs auf Anbieter, die KI-Systeme in der Union in Verkehr bringen oder in Betrieb nehmen (lit. a), auf die Nutzer von KI-Systemen, die sich in der EU befinden, (lit. b) und auf Anbieter und Nutzer, die in einem Drittland niedergelassen oder ansässig sind, wenn die Ergebnisse der KI-Systeme in der Union verwendet werden (lit. c).

Es ist zu begrüßen, dass insbesondere die Regelung des lit. c den Schutz der Bürger der Europäischen Union in den Blick nimmt. Nach Erwägungsgrund 11 sollen nämlich ausdrücklich auch Hochrisiko-KI-basierte Dienstleistungen vom Anwendungsbereich des KI-VO-E erfasst werden, die ein in der Union ansässiger Akteur an einen externen Akteur vergibt, die sich jedoch wieder auf die Bürger der Union auswirken. Der Anwendungsbereich des KI-VO-E hat somit eine extraterritoriale Komponente, wobei eine Auswirkung auf in der Union Ansässige immer erforderlich ist.

Nicht umfasst sind allerdings KI-Systeme, die in der EU entwickelt, aber in einem Drittland in Verkehr gebracht werden, ohne dass ihre Ergebnisse in der Union verwendet werden. Dies hat zur Folge, dass Systeme exportiert werden könnten, die für in der EU nach Art. 5 KI-VO-E verbotene Praktiken genutzt werden können oder die als Hochrisiko-KI nach Art. 6 KI-VO-E nicht die Zulassungsvoraussetzungen der Art. 8 ff. KI-VO-E erfüllen. Damit wird die KOM ihrem selbst in der Begründung unter 1.1 formulierten Ziel, „bei der Entwicklung einer sicheren vertrauenswürdigen und ethisch vertretbaren künstlichen Intelligenz weltweit eine Führungsrolle einzunehmen“¹², nicht gerecht. Das Katholische Büro regt daher an, soweit es die Kompetenz der EU zulässt, den Anwendungsbereich des Art. 2 Abs. 1 KI-VO-E so auszuweiten, dass er auch die oben dargestellte Konstellation umfasst.

Des Weiteren erscheint problematisch, dass die Eröffnung des Anwendungsbereichs des KI-VO-E nach Art. 2 Abs. 1 lit. b KI-VO-E über den in der EU befindlichen „Nutzer“ konstruiert wird. Dieser wird dann über Art. 3 Nr. 4 KI-VO-E als u.a. „natürliche oder juristische Person“ definiert, „die ein KI-System in eigener Verantwortung verwendet, es sei denn, das KI-System wird im Rahmen einer persönlichen und nicht beruflichen Tätigkeit verwendet“. Hieraus kann nur geschlossen werden, dass KI-Systeme, die spezifisch für Dienstleistungen oder sonstige Vorgänge konstruiert und eingesetzt werden, die natürliche Personen außerhalb ihrer beruflichen, insbesondere ihres privaten Verhaltens nutzen, aus dem Anwendungsbereich des KI-VO-E herausfallen. Dabei sind es gerade die persönlichen Tätigkeiten im privaten Kontext, in denen die Verwender von KI-Systemen aufgrund des ausgeprägten Wissens-, Macht- und Einflussgefälles besonders hohen Risiken ausgesetzt sind. Wir sprechen uns daher dafür aus, in der Definition des „Nutzers“ nach Art. 3 Nr. 4 KI-VO-E den einschränkenden Halbsatz „es sei denn, das KI-System...“ zu streichen.

Zudem statuiert Art. 2 Abs. 4 KI-VO-E eine Ausnahmeregelung für Behörden in Drittstaaten und internationalen Organisationen: Besteht eine Übereinkunft über die Zusammenarbeit im Bereich der Strafverfolgung und justiziellen Zusammenarbeit, findet die

¹² COM(2021) 206 final (deutsche Fassung), S. 1.

Verordnung keine Anwendung. Diese Ausnahme erzeugt eine bedenkliche Regelungslücke, da von KI-Systemen erzeugte Erkenntnisse, die von Behörden der EU nicht hätten erzeugt oder genutzt werden dürfen, über eine internationale Kooperation im Bereich der Strafverfolgung wieder einbezogen werden können. Gerade in Anbetracht der Grundrechtssensibilität im Bereich der Strafverfolgung wäre hier ein restriktiverer Ansatz wünschenswert.

3. Verbotene KI-Systeme, Art. 5 KI-VO-E

Erwägungsgrund 15 des KI-VO-E weist darauf hin, dass KI-Systeme „*auch missbraucht werden und neue und wirkungsvolle Instrumente für manipulative, ausbeuterische und soziale Kontrollpraktiken bieten*“ können und solche Praktiken „*verboten werden [sollten], weil sie im Widerspruch zu den Werten der Union stehen, nämlich der Achtung der Menschenwürde, Freiheit, Gleichheit, Demokratie und Rechtsstaatlichkeit sowie der Grundrechte der Union*“. Art. 5 Abs. 1 KI-VO-E nennt nun vier verbotene Praktiken: die „*unterschwellige Beeinflussung außerhalb des Bewusstseins einer Person*“ (lit. a), das Ausnutzen einer „*Schwäche oder Schutzbedürftigkeit einer Gruppe von Personen*“ (lit. b), die Verwendung von KI-Systemen durch „*Behörden oder in deren Auftrag zur Bewertung oder Klassifizierung der Vertrauenswürdigkeit natürlicher Personen*“ [im Folgenden: ‚Social Scoring‘] (lit. c) und die Verwendung „*biometrischer Echtzeit-Fernidentifizierungssysteme in öffentlich zugänglichen Räumen zu Strafverfolgungszwecken*“ (lit. d). Das Katholische Büro begrüßt, dass der KI-VO-E die in lit. a bis lit. d genannten Praktiken im Bereich der KI als „verboten“ qualifiziert. Alle vier in Art. 5 Abs. 1 KI-VO-E mit der Verwendung von KI verbundenen Praktiken sind jedoch zu eng konstruiert. Darüber hinaus sollten sie um die Verwendung einiger der in Art. 6 Abs. 2 i.V.m. Annex III KI-VO-E genannten Hochrisiko-KI-Systeme ergänzt werden, die als verboten einzustufen sein sollten (hierzu ausführlich unter Punkt 4).

a) Art. 5 Abs. 1 lit. a KI-VO-E

Techniken der „*unterschwelligen Beeinflussung außerhalb des Bewusstseins einer Person*“ sind nach Art. 5 Abs. 1 lit. a KI-VO-E nur verboten, wenn sie „*eingesetzt werden, um das Verhalten einer Person in einer Weise wesentlich*“ zu beeinflussen, die dieser Person oder einer anderen Person einen „*physischen und psychischen Schaden*“ zufügt oder zufügen kann.

Zunächst sind die Merkmale der verbotenen Manipulation selbst nicht eindeutig. Verboten sein sollen „*Techniken der unterschwelligen Beeinflussung außerhalb des Bewusstseins einer Person*“. Unter der „*unterschwelligen*“ Beeinflussung eines Menschen werden aber in der Psychologie diesem Menschen dargebotene „*subliminale Reize*“ verstanden, die der Person selbst nicht bewusst sind.¹³ Dass die Beeinflussung nach Art. 5 Abs. 1 lit. a KI-VO-E zusätzlich zu „*unterschwellig*“ auch „*außerhalb des Bewusstseins einer Person*“ erfolgen soll, erscheint insofern gedoppelt und daher überflüssig. Das Katholische Büro regt daher an, dieses Tatbestandsmerkmal entweder zu präzisieren oder zu streichen.

Des Weiteren ist es problematisch, dass nach Art. 5 Abs. 1 lit. a KI-VO-E offenbar eine Manipulation nur dann als „*wesentlich*“ angesehen wird, wenn sie einer Person einen „*physischen oder psychischen Schaden*“ zufügt. Denn es sind durchaus „*manipulative, ausbeuterische und soziale Kontrollpraktiken*“ im Sinne des o.g. Erwägungsgrunds 16 des

¹³<https://www.spektrum.de/lexikon/psychologie/unterschwellige-wahrnehmung/16041>.

KI-VO-E denkbar, die keinen solchen Schaden verursachen und dennoch zu einer Sozialkontrolle von Menschen führen und/oder Grundrechte verletzen. Daher liegt es auch nahe, sich auf den Standpunkt zu stellen, dass bereits das manipulierte Verhalten selbst einen hinreichend zu missbilligenden Erfolg der Manipulation darstellt, um ihn über Art. 5 Abs. 1 lit. a KI-VO-E zu verbieten, selbst wenn es nicht zu einem „*physischen oder psychischen*“, also letztlich gesundheitlichen Schaden kommt. Darüber hinaus wird nicht weiter definiert, welcher Maßstab für einen psychischen Schaden verwendet werden soll.

Schließlich berücksichtigt Art. 5 Abs. 1 lit. a KI-VO-E zwar gesundheitliche, aber keine wirtschaftlichen oder finanziellen Schäden. Dies ist aus kirchlicher Sicht wenig sinnvoll, da der KI-VO-E nach seiner jetzigen Konstruktion gemäß Art. 2 Abs. 1 lit. b KI-VO-E und Abs. 3 Nr. 4 KI-VO-E gerade nur auf „*Nutzer*“ angewendet werden soll, die mit KI-Praktiken in einer beruflichen (und nicht persönlichen) Tätigkeit konfrontiert sind. Gerade im beruflichen Kontext können aber unterschwellig manipulierende KI-Systeme auch erhebliche wirtschaftliche und finanzielle Schäden verursachen. Warum solche Schäden anders als die gesundheitlichen Schäden nicht ausreichen sollen, um eine KI-Praktik als verboten zu qualifizieren, ist nicht nachzuvollziehen. Dies gilt umso mehr, als dass wirtschaftliche und finanzielle Schäden durchaus zu gesundheitlichen Schäden führen können. Daher befürworten wir, auch den wirtschaftlichen respektive finanziellen Schaden als Alternative zu den „*physischen oder psychischen*“ Schäden in die Norm aufzunehmen.

b) Art. 5 Abs. 1 lit. b KI-VO-E

Gem. Art. 5 Abs. 1 lit. b KI-VO-E ist ein KI-System verboten, „*das eine Schwäche oder Schutzbedürftigkeit einer Gruppe von Personen*“ (Alter, körperliche Behinderung...) „*ausnutzt, um das Verhalten einer dieser Gruppe angehörenden Person in einer Weise wesentlich zu beeinflussen, die dieser Person oder einer anderen Person einen physischen oder psychischen Schaden zufügt oder zufügen kann*“.

Aus kirchlicher Sicht begrüßen wir grundsätzlich, dass der KI-VO-E die Risiken von KI-Praktiken auch unter dem Gesichtspunkt der besonderen Schutzbedürftigkeit von bestimmten Personengruppen betrachtet.

Allerdings stellen sich auch hier aufgrund des teilweise identischen Wortlauts die bereits unter **a)** dargestellten Problematiken (*unterschwellige Beeinflussung, wesentliche Beeinflussung, psychischer und wirtschaftlicher Schaden*).

c) Art. 5 Abs. 1 lit. c KI-VO-E

Art. 5 Abs. 1 lit. c KI-VO-E beinhaltet ein Verbot des Inverkehrbringens, der Inbetriebnahme oder der Verwendung von „*KI-Systemen durch Behörden oder in deren Auftrag zur Bewertung oder Klassifizierung der Vertrauenswürdigkeit natürlicher Personen über einen bestimmten Zeitraum auf der Grundlage ihres sozialen Verhaltens*“, wobei die „*soziale Bewertung*“ zu einer Schlechterstellung oder Benachteiligung bestimmter natürlicher Personen oder Personengruppen führen muss, sei es aufgrund von sozialen Zusammenhängen, „*die in keinem Zusammenhang zu den Umständen*“ der Datenerzeugung oder -erfassung stehen (i), oder „*in einer Weise, die sich im Hinblick auf ihr soziales Verhalten oder dessen Tragweite ungerechtfertigt oder unverhältnismäßig erweist*“ (ii).

Wir weisen darauf hin, dass es grundsätzlich aus „*sozialethischer Sicht [nicht akzeptabel] ist, Menschen allein auf der Grundlage ihrer Daten und unter Nutzung algorithmischer*

Berechnungen zu bewerten oder zu kategorisieren".¹⁴ Darüber hinaus ist es nicht nachvollziehbar, warum die Verwendung der hier beschriebenen KI-Systeme des ‚Social Scoring‘ nur durch öffentliche Behörden einem Verbot unterliegen soll. Denn ihre Nutzung durch Private, insbesondere Unternehmen, kann negative Ergebnisse zur Folge haben, die mit denen in i) und ii) vergleichbar sind. Daher spricht sich das Katholische Büro dafür aus, ein solches Verbot in den KI-VO-E aufzunehmen und Art. 5 Abs. 1 lit. c KI-VO-E auf KI-gestützte Praktiken des ‚Social Scoring‘ durch Private zu erweitern. Ein solches umfassendes Verbot des ‚Social Scoring‘ durch Private ist erforderlich, da es sich bisher nicht aus anderen Rechtsvorschriften – wie der Datenschutzgrundverordnung (DS-GVO) oder der dem Parlament derzeit zur Abstimmung vorliegenden Entwurf der KOM zum Digital Services Act (DSA) – ergibt. Wir regen daher an, die Worte *„durch Behörden oder in deren Auftrag“* aus Art. 5 Abs. 1 lit. c KI-VO-E zu streichen.

Des Weiteren weist das Katholische Büro darauf hin, dass der Einsatz der in Art. 5 Abs. 1 lit. c KI-VO-E erfassten KI-Systeme im Wortlaut dieses Artikels nur mit Blick auf einen sehr begrenzten Zweck verboten wird, nämlich *„zur Bewertung oder Klassifizierung der Vertrauenswürdigkeit“*. Der zu diesem Artikel gehörende Erwägungsgrund 17 bezieht sich allerdings auf KI-Systeme, die *„das soziale Verhalten natürlicher Personen für allgemeine Zwecke“* bewerten. Da auch solche allgemeinzweckorientierte ‚Social Scoring‘-KI die *„Gesellschaftsförderlichkeit“* oder gar *„Gesellschaftsschädlichkeit“* einer Person auswerten und, wie Erwägungsgrund 17 feststellt, potenziell die Menschenwürde, das Recht auf Nichtdiskriminierung sowie die Werte der Gleichheit und Gerechtigkeit verletzen, sollten sie direkt vom Wortlaut des Art. 5 Abs. 1 lit. c KI-VO-E erfasst werden. Eine Engführung auf ‚Social Scoring‘-KI *„zur Bewertung oder Klassifizierung von Vertrauenswürdigkeit“* wird den mit diesen KI-Systemen verbundenen Risiken nicht gerecht. Das Katholische Büro spricht sich daher dafür aus, im Wortlaut des Art. 5 Abs. 1 lit. c KI-VO-E die Beschreibung der KI-Systeme als *„zur Bewertung von oder Klassifizierung der Vertrauenswürdigkeit natürlicher Personen“* durch *„zur Bewertung oder Klassifizierung des sozialen Verhaltens natürlicher Personen“* zu ersetzen.

In Parallelität zu Art. 5 Abs. 1 lit. a KI-VO-E ist es aus kirchlicher Sicht problematisch, in Art. 5 Abs. 1 lit. c KI-VO-E ‚Social Scoring‘ nur dann zu verbieten, wenn es zu einer *„Schlechterstellung oder Benachteiligung bestimmter natürlicher Personen [...] in sozialen Zusammenhängen“* führt. Es wird dadurch verkannt, dass eine mögliche Einflussnahme auf die Freiheitsrechte einer Person schon zuvor droht. Denn bereits das Bewusstsein der ständigen Bewertung des eigenen Verhaltens in „Gut“ und „Böse“ durch ein KI-System kann deutliche Auswirkung z.B. auf das soziale Zusammenleben innerhalb der Gesellschaft, aber auch auf die Frage haben, ob man Versammlungen etc. im öffentlichen Raum durchführt.

d) Art. 5 Abs. 1 lit. d KI-VO-E

Art. 5 Abs. 1 lit. d KI-VO-E untersagt die *„biometrische Echtzeit-Fernidentifizierung in öffentlich zugänglichen Räumen zu Strafverfolgungszwecken“*. Es bestehen aber drei Ausnahmetatbestände, nach welchen eine solche Identifizierung unter gewissen Bedingungen zulässig ist, wie etwa die *„gezielte Suche nach bestimmten potentiellen Opfern“* (i), das *„Abwenden einer konkreten, erheblichen und unmittelbaren Gefahr“* (ii) oder *„zum Erkennen, Aufspüren oder Verfolgen eines Täters oder Verdächtigen einer Straftat im*

¹⁴ Die deutschen Bischöfe, Kommission für gesellschaftliche und soziale Fragen, Heft Nr. 52, Europa ist es Wert - Impulse der Bischöflichen Arbeitsgruppe Europa, 15. November 2021, S. 50.

Sinne des Art. 2 Abs. 2 des Rahmenbeschlusses 2002/584/JI des Rates“ (iii), wenn die Tat mit einer Freiheitsstrafe oder einer freiheitsentziehenden Maßregel der Sicherung im Höchstmaß von mindestens drei Jahren bedroht ist.

Das Katholische Büro hat in seinem Konsultationsbeitrag zum Weißbuch¹⁵ darauf hingewiesen, dass die gesetzgeberischen Kompetenzen der EU in den Bereichen Strafverfolgung und Gefahrenabwehr nur sehr schwach ausgeprägt sind. Es spricht vielmehr einig und nicht zuletzt das Subsidiaritätsprinzip dafür, dass es den mitgliedstaatlichen Gesetzgebern überlassen bleiben sollte, in diesen Bereichen auch bei Involvierung von KI-Systemen eigene, ihrer mitgliedstaatlichen Rechtsordnung angepasste Regelungen bzgl. der Verwendung von Fernidentifizierungssystemen zu treffen. Die in Art. 5 Abs. 4 S. 2 KI-VO-E integrierte Ermächtigung für nationale Gesetzgeber, detaillierte Vorschriften für etwa die Beantragung der nach Art. 5 Abs. 3 KI-VO-E erforderlichen Genehmigung durch Justizbehörden für die Anwendung biometrischer Echtzeit-Fernidentifizierung zu Strafverfolgungszwecken erlassen zu können, hilft dieser Mangel an Kompetenzen nicht ab.

Sofern man aber im Rahmen des KI-VO-E Regelungen zur KI-gestützten biometrischen Echtzeit-Fernidentifikation treffen will, ist es aus hiesiger Sicht problematisch, dass dieser Vorschlag KI-Systeme mit biometrischer Echtzeit-Fernidentifizierung nur unter sehr engen Bedingungen verbietet. Denn nach der Konstruktion des Art. 5 Abs. 1 lit. d KI-VO-E ist die biometrische Echtzeit-Fernidentifizierung zu allen anderen Zwecken als der Strafverfolgung, in allen nicht öffentlichen zugänglichen Räumen und in Online-Räumen ebenso wie jede biometrische Fernidentifizierung, die nicht in Echtzeit erfolgt, erlaubt.

Wir regen an, die Breite der erlaubten Anwendungsmöglichkeit für biometrische Echtzeit-Fernidentifizierung zu überdenken und Art. 5 Abs. 1 lit. d KI-VO-E angesichts ihrer hohen Grundrechtssensibilität mit Blick auf folgende Gesichtspunkte enger zu fassen:

Zunächst überzeugt die Abgrenzung der gem. Art. 5 Abs. 1 lit. d KI-VO-E verbotenen biometrischen Echtzeit-Fernidentifizierungssystemen zu den erlaubten Fernidentifizierungssystemen anhand des Kriteriums der Zeit nicht. Nach Erwägungsgrund 8 unterscheiden sich nämlich Echtzeit-Fernidentifizierungssysteme von der nachträglichen biometrischen Fernidentifizierung dadurch, dass die Erfassung der biometrischen Daten, der Abgleich und die Identifizierung *„unverzüglich, zeitnah oder auf jeden Fall ohne erhebliche Verzögerung“* erfolgen soll. Dabei soll eine Umgehung der Verordnung über den Begriff *„Echtzeit-Nutzung“* nicht erfolgen, indem *„kleinere Verzögerungen vorgesehen werden“*. Im Bereich der Datenverarbeitung nehmen aber Prozesse nur wenige Augenblicke in Anspruch, weswegen nicht erkennbar ist, wo die Grenze zwischen *„nachträglich“* und *„zeitnah“* liegen soll. Schließlich ist auch schwierig begründbar, warum sich die Intensität des Eingriffs in Grundrechte durch eine biometrische Echtzeit-Fernidentifizierung von der durch eine biometrische Fernidentifizierung unterscheidet. Eine Überwachung der Bevölkerung und damit eine erhebliche Einschränkung der persönlichen Freiheit tritt bei beiden Varianten ein.

Des Weiteren sollte die Begrenzung auf *„öffentliche Räume“* in Art. 5 Abs. 1 lit. d KI-VO-E nach Ansicht des katholischen Büros aufgegeben und das Verbot auf *„Online-Räume“* erweitert werden. Erwägungsgrund 9 erläutert zu Art. 5 Abs. 1 lit. d KI-VO-E, dass unter dem Begriff *„öffentlicher Raum“* bisher lediglich öffentlich physisch zugängliche und explizit keine Online-Räume verstanden werden sollen. Allerdings können auch Online-

¹⁵ COM(2011) 144 final (deutsche Fassung).

Räume zu „*öffentlichen Räumen*“ i.S.d. Art. 5 Abs. 1 lit. d KI-VO-E werden, wenn durch sie Öffentlichkeit entsteht, d.h. eine gesellschaftlich relevante Interaktion zwischen Menschen oder anders öffentliche Kommunikation stattfindet.¹⁶ In den letzten Jahren wurden immer mehr Orte der öffentlichen Begegnung wie etwa Vorträge und Theateraufführungen in den Online-Raum verlegt. Ohne die Einbeziehung von „Online-Räumen“ in Art. 5 Abs. 1 lit. d KI-VO-E käme es zu der paradoxen Situation, dass bei einer Veranstaltung in einem öffentlichen Raum in Präsenz eine Überwachung mittels biometrischer Echtzeit-Fernidentifizierung zu Strafverfolgungszwecken nicht erfolgen dürfte, wohingegen dies rechtlich möglich wäre, wenn exakt diese Veranstaltung online durchgeführt werden würde.

Noch unbefriedigender ist, dass Art. 5 Abs. 1 lit. d KI-VO-E die biometrische Echtzeit-Fernidentifizierung lediglich zu „*Strafverfolgungszwecken*“ verbietet und diese somit zu allen anderen Zwecken erlaubt. Die Überwachung im öffentlichen Raum kann unabhängig vom Zweck zu einer weitreichenden Kontrolle des Bewegungsverhaltens und der sozialen Interaktion von Personen führen, die sich im Bewusstsein dieser Überwachung auch von der Ausübung ihrer Grundrechte abgehalten oder zumindest in dieser gestört sehen können.

Im Übrigen sind die Ausnahmeregelungen (i)-(iii) von dem ohnehin schon sehr engen Verbot der biometrischen Echtzeit-Fernidentifizierung auch noch sehr weitgehend. Der Verweis in (iii) auf Art. 2 Abs. 2 des Rahmenbeschlusses zum Europäischen Haftbefehl umfasst beispielsweise auch Betrugsdelikte und weitere Vermögensdelikte. Im Vergleich zu den unter (i) und (ii) genannten Tatbeständen, die primär dem Schutz von Leib und Leben dienen, wird hier die Schwelle zur Anwendung der biometrischen Echtzeit-Fernidentifizierung eher niedrig angesetzt. Das erscheint angesichts der in der Abwägung stehenden Rechtsgüter – der Beeinträchtigung der Freiheit des Individuums und dem Verlust eines möglicherweise auch nur geringen Vermögenswertes – durchaus problematisch. Zur Veranschaulichung wäre es nach dieser Regelung möglich, Tatverdächtige eines nach deutschem Strafrecht einfachen Betrugs (§ 263 Abs. 1 StGB), der lediglich mit einer Geldstrafe oder einer Haftstrafe bis zu fünf Jahren bedroht ist, außerhalb von öffentlichen Räumen zu Strafverfolgungszwecken durch biometrische Echtzeit-Fernidentifizierung identifizieren zu lassen.

Schließlich sollte beachtet werden, dass die Verwendung von KI-Systemen bei der biometrischen Echtzeit-Fernidentifizierung zu Strafverfolgungszwecken mit einer Reihe praktischer Probleme behaftet ist. Erfahrungswerte bei der Anwendung biometrischer Echtzeit-Fernidentifizierungssysteme mit KI-Komponente zeigen, dass marginalisierte Gruppen unverhältnismäßig betroffen werden. Es wurde aber offengelegt, dass die falsch positive Rate von Echtzeit-Fernidentifizierungssystemen vermehrt Frauen und farbige Personen betrifft, sodass diese bei einer Strafverfolgung mittels biometrischer Echtzeit-Fernidentifizierung öfter als der Durchschnitt kontrolliert und festgehalten werden.¹⁷ Diese Instrumente sind also in gewisser Weise voreingenommen und besitzen

¹⁶BBST-Online_Publikation, Nr. 07/2015, S. 45, abrufbar unter:

https://www.bbsr.bund.de/BBSR/DE/veroeffentlichungen/bbsr-online/2015/DL_ON072015.pdf;jsessionid=75A7F72081131888B9FD2E7DAC870BE5.live11314?__blob=publicationFile&v=1.

¹⁷ Gesichtserkennungstechnologien: grundrechtsrelevante Erwägungen im Rahmen der Strafverfolgung, FRA European Union Agency for fundamental rights, Gesichtserkennungstechnologien: grundrechtsrelevante Erwägungen im Rahmen der Strafverfolgung (europa.eu), S. 4.

den Anschein technischer Neutralität.¹⁸ Darüber hinaus hat man in einem gemeinsamen Pilotprojekt „Sicherheitsbahnhof Berlin Südkreuz“ bei der Anwendung intelligenter Videoanalysetechniken mithilfe von Gesichtserkennungssystemen festgestellt, dass weiterhin ein so substanzieller, falsch positiver Output des Systems besteht, dass eine weitere Überprüfung durch Personal erforderlich wird und die gewünschte Ersparung von Ressourcen nicht eintritt.¹⁹

4. Hochrisiko-KI-Systeme, Art. 6 ff. KI-VO-E

Art. 6 KI-VO-E nennt Klassifizierungsvorschriften für Hochrisiko-KI-Systeme. Ein KI-System gilt nach Art. 6 Abs. 1 KI-VO-E als Hochrisiko-KI-System, wenn sie als Sicherheitskomponente eines Produkts verwendet wird oder es selbst das Produkt ist, das unter die in Anhang II genannten Rechtsvorschriften zur Harmonisierung fällt (lit. a) oder das KI-System selbst oder als Sicherheitskomponente als Teil eines Produkts einer Konformitätsbewertung bzgl. Inbetriebnahme und Inverkehrbringen durch Dritte gem. den in Anhang II genannten Vorschriften unterzogen werden muss (lit. b).

Art. 6 Abs. 2 KI-VO-E qualifiziert dann spezifisch KI-Systeme, die in Annex III genannt werden, als hochriskant. Annex III unterscheidet acht Bereiche, in die diese KI-Systeme eingruppiert werden: den Bereich der *„Biometrischen Identifizierung und Kategorisierung natürlicher Personen“* (Nr. 1), der *„Verwaltung und Betrieb kritischer Infrastrukturen“* (Nr. 2), der *„Allgemeinen und beruflichen Bildung“* (Nr. 3), der *„Beschäftigung, Personalmanagement und Zugang zur Selbstständigkeit“* (Nr. 4), der *„Zugänglichkeit und Inanspruchnahme grundlegender privater und öffentlicher Dienste und Leistungen“* (Nr. 5), der *„Strafverfolgung“* (Nr. 6), der *„Migration, Asyl und Grenzkontrolle“* (Nr. 7) und der *„Rechtspflege und demokratischen Prozesse“* (Nr. 8).

Art. 6 KI-VO-E liegt dabei die sich aus Erwägungsgrund 32 ergebende Idee zugrunde, dass solche KI-Systemen als hochriskant eingestuft werden sollen, die *„aufgrund ihrer Zweckbestimmung ein hohes Risiko bergen, die Gesundheit und Sicherheit oder die Grundrechte von Personen zu schädigen, wobei sowohl die Schwere des möglichen Schadens als auch die Wahrscheinlichkeit seines Auftretens zu berücksichtigen sind, und sofern sie in einer Reihe von Bereichen verwendet werden, die in der Verordnung ausdrücklich festgelegt sind“*.

Zwar begrüßt das Katholische Büro dem Grundsatz nach, dass die KOM KI-Systeme nach der Höhe der in ihrer Zweckbestimmung zum Ausdruck kommenden Risiken der Verursachung eines Schadens für Gesundheit, Sicherheit oder Grundrechte von Personen als hochriskant oder eben nicht hochriskant einstuft. Dabei fehlt jedoch eine Berücksichtigung der Risiken, die manche KI-Systeme für die Gesellschaft als Ganzes, das Gemeinwesen, die Demokratie und insgesamt die Werte der Union des Art. 2 EUV darstellen können. Auch der „Rome Call for AI Ethics“ fordert insoweit, dass KI-Systeme nicht nur auf das Wohl des einzelnen Menschen, sondern auch *„auf das Wohl der Menschheit und der Umwelt [und des] gemeinsamen Hauses [...]“* ausgerichtet werden.²⁰

Vor allem aber erscheint die konkret in Art. 6 KI-VO-E gewählte bereichsorientierte Konstruktion zur Bestimmung von Hochrisiko-KI-Systemen unterkomplex. Denn, hierauf hat

¹⁸ Crawford, K. (2019), „Regulate facial-recognition technology“, Nature 572 (2019), 29. August 2019, S. 565.

¹⁹ Vgl. Enquete-Kommission Künstliche Intelligenz, Teilbericht, Projektgruppe KI und Staat, S. 87.

²⁰ „Rome Call for AI Ethics“, 2020.

das Katholische Büro bereits in seinem Konsultationsbeitrag zum Weißbuch KI der KOM hingewiesen, es sind ethisch sensible, mit Schadensrisiken für die Gesundheit, Sicherheit oder Grundrechte von Personen oder das Gemeinwesen einhergehende Anwendungsbereiche von KI-Systemen bereichsübergreifend zu finden. Daher haben wir uns in unserem Konsultationsbeitrag gegen einen bereichsspezifischen Ansatz und für eine kontextabhängige Grundrechts-Folgenabschätzung jedes KI-Systems – ähnlich wie die Datenschutz-Folgenabschätzung in Art. 35 DS-GVO – ausgesprochen, um dessen Gefährdungspotential zu bestimmen.

Seine Bedenken gegen die Verwendung eines bereichsorientierten Ansatzes der Bestimmung von Hochrisiko-KI-Systemen sieht das Katholische Büro leider auch in der Gestaltung des Art. 6 KI-VO-E, die die KOM vorgeschlagen hat, bestätigt. Denn zum Ersten enthält Annex III, auf den Art. 6 Abs. 2 KI-VO-E verweist, KI-Systeme mit Anwendungsbereichen, die als inakzeptable Praktiken nach Art. 5 KI-VO-E verboten werden sollten (1). Zum Zweiten sind die in Annex III aufgeführten Bereiche unvollständig (2). Zum Dritten sollten bestimmte KI-Systeme in den Bereichen des Annex III ergänzt werden (3).

(1) Wir sprechen uns dafür aus, zumindest zwei mit KI verbundene Praktiken direkt in Art. 5 KI-VO-E zu verbieten:

Zum einen sollten nicht nur KI-Systeme, die die individuellen Grundrechte erheblich beeinträchtigen, sondern auch solche, die das Potential haben, erhebliche Schäden für die Gesellschaft als Ganzes, das Gemeinwesen, die Demokratie und insgesamt die Werte der Union zu verursachen, nicht nur als hochriskant eingestuft, sondern verboten werden. Hierzu sind beispielsweise KI-Systeme zu zählen, die spezifisch dafür programmiert werden, den öffentlichen Diskurs und die demokratische Meinungs- und Willensbildung einer Gesellschaft, etwa durch die Verbreitung von sog. Fake News, zu manipulieren, gesellschaftlich zu polarisieren und soziale Verwerfungen zu verursachen. Dies hat die KOM auch erkannt und fordert daher in Erwägungsgrund 15, dass solche manipulativen Praktiken, da sie im „*Widerspruch zu den Werten der Union stehen, nämlich [...] Demokratie*“, verboten werden sollen. In Art. 5 KI-VO-E findet sich jedoch nicht das Verbot solcher Praktiken; sie sollten aber dort aufgenommen werden. Ansonsten drohen KI-Systeme, die bei demokratischen Prozessen angewendet werden, sich zur „Achillesferse“ unserer Rechtsordnung zu entwickeln.

Zum anderen ist es ebenso erforderlich, dass Entscheidungen über Leben und Tod eines Menschen nicht einem KI-System überlassen werden. Der Bereich 5 des Annex III in lit. c überantwortet die „*Entsendung oder Priorisierung des Einsatzes von Not- und Rettungsdiensten*“ aber einem KI-System und stuft dieses nur als Hochrisiko-KI und nicht als verbotene Praktik im Bereich der KI ein. Der Einsatz von KI anstelle von Personen, die im Zweifelsfall z.B. über Leben und Tod wie etwa bei der Priorisierung im Einsatz von Not- und Rettungsdiensten entscheiden müssen, ist hoch problematisch. Maschinen können selbst keine ethischen Entscheidungen treffen; vielmehr erfolgt die Entscheidung aufgrund programmierter Abwägungsprozesse.²¹ In einem von Menschen entwickelten Algorithmus wären Leben und Tod lediglich Faktoren einer Formel und es käme hierdurch zu einer Abwägung von Menschenleben, bei welcher nach dem größtmöglichen vorher definierten Nutzen die Entscheidung über

²¹ Die deutschen Bischöfe, Kommission für gesellschaftliche und soziale Fragen, Heft Nr. 52, Europa ist es Wert - Impulse der Bischöflichen Arbeitsgruppe Europa, 15. November 2021, S. 48.

das Überleben eines Menschen durch das KI-System getroffen werden würde. Damit wäre der Mensch nicht mehr das Ziel, sondern nur noch Objekt der Entscheidung und damit in seiner unantastbaren Menschenwürde verletzt. Nur ein Mensch ist in der Lage, einen anderen Menschen nicht nur als Faktor, sondern in seinem sozialen Achtungsanspruch als Mensch und seiner Subjektstellung in Entscheidungsprozessen zu würdigen. Eine Übertragung solcher Abwägungsprozesse wie in Annex III Nr. 5 lit. c bei dem Einsatz von Not- und Rettungsdiensten, sollte daher untersagt und in den verbotenen Praktiken in Art. 5 KI-VO-E aufgeführt werden.

- (2) Wir regen an, den Annex III um einen weiteren Bereich Nr. 10 zu erweitern, nämlich um KI-Systeme, die großflächige Effekte auf die Umwelt haben. Auch Papst Franziskus weist in seiner Ansprache zum "Rome Call for AI Ethics" darauf hin, dass Technologien wie KI es gestatten, Eingriffe mit planetarischer Tragweite vorzunehmen.²² Zwar hat die KOM die Bedeutung der Umwelt partiell erkannt, indem sie in Erwägungsgrund 28 fordert, dass das „Grundrecht auf ein hohes Umweltschutzniveau, das in der Charta verankert ist und mit der Unionspolitik umgesetzt wird, bei der Bewertung der Schwere des Schadens“ zum Tragen kommen muss. Allerdings beschränkt sie sich bei der Berücksichtigung des Umweltaspekts allein auf die Tatsache, dass die Umwelt bei der Bewertung des Schadens miteinbezogen werden sollte. Es können aber auch KI-Systeme selbst, beispielsweise solche, die für Geo-Engineering eingesetzt werden, aufgrund der großskaligen Auswirkung und der möglichen langfristigen Dauer von Umweltschäden einen erheblichen Schaden für eine Vielzahl von Menschen verursachen.

In Anbetracht dessen stimmt die Kirche auch mit dem Vorschlag des Europäischen Rates überein, dass KI-Systeme „zur Kontrolle von Emissionen und Umweltverschmutzung“²³ ebenfalls als hochriskant eingestuft werden.

- (3) Schließlich sollten zumindest in dem 8. Bereich des Annex III weitere KI-Systeme ergänzt werden.

Im 8. Bereich des Annex III unter dem Titel „Rechtspflege und demokratische Prozesse“ werden nur KI-Systeme aufgeführt, die die Rechtspflege betreffen; KI-Systeme, die in demokratischen Prozessen Anwendung finden, fehlen. Im Umkehrschluss ist der Einsatz jeglicher Art von KI-Systemen im Rahmen von demokratischen Prozessen ohne Zulassungsvoraussetzungen erlaubt. Erwägungsgrund 40 hebt aber hervor, dass „KI-Systeme, die für die Rechtspflege und demokratische Prozesse bestimmt sind, [...] angesichts ihrer möglichen erheblichen Auswirkungen auf die Demokratie [und] die Rechtsstaatlichkeit [...] als hochriskant eingestuft werden“ sollen. Daher sollten KI-Systeme, die in demokratischen Prozessen eingesetzt werden und noch nicht, wie oben dargelegt, zu den die demokratische Willensbildung manipulierenden Praktiken nach Art. 5 KI-VO-E zu zählen sind, im Bereich 8 des Annex III noch ergänzt und damit als Hochrisiko eingestuft werden.

Tatsächlich führt der 8. Bereich des Annex III bisher nur KI-Systeme auf, „die bestimmungsgemäß Justizbehörden bei der Ermittlung, bei der Auslegung von Sacherhalten

²² vgl. „Rome Call for AI Ethics“, 2020.

²³ EU-Ratspräsidentschaft schlägt wesentliche Änderungen am KI-Gesetzentwurf vor – EURACTIV.de.

und Rechtsvorschrift und bei der Anwendung des Rechts auf konkrete Sachverhalte unterstützen sollen“ auf KI-Systeme, die diese Aufgaben voll automatisch ausführen und beispielsweise zukünftig als RichterIn oder Richter fungieren könnten, werden von dieser Formulierung nicht umfasst. Das hätte zur Folge, dass sie den Zulassungsbedingungen der Art. 8 ff. KI-VO-E nicht unterfallen würden. Dies erscheint aber widersinnig, wenn bereits weniger autonome KI-Systeme als hochriskant eingestuft werden.

5. Risikomanagementsysteme, Art. 9 KI-VO-E

Art. 9 Abs. 1 KI-VO-E fordert die Einrichtung, Anwendung, Dokumentation und Aufrechterhaltung von Risikomanagementsystemen für Hochrisiko-KI-Systeme. Abs. 2 dieser Norm umschreibt die Schritte, die das Risikomanagementsystem umfasst. Spezifisch Art. 9 Abs. 8 KI-VO-E normiert, dass bei jeder Art des Risikomanagementsystems vor allem zu beachten ist, ob das Hochrisiko-KI-System *„wahrscheinlich für Kinder zugänglich ist oder Auswirkungen auf Kinder hat“*.

Das Katholische Büro befürwortet die Verpflichtung zur Einrichtung eines Risikomanagementsystems für Hochrisiko-KI-Systeme. Insbesondere begrüßt es, dass bei der Umsetzung des Risikomanagements nach Abs. 8 der Norm zu berücksichtigen ist, *„ob das Hochrisiko-KI-System wahrscheinlich für Kinder zugänglich ist oder Auswirkungen auf Kinder hat“*. Denn auch wenn die Anwendung von KI neue Möglichkeiten und Vorteile bietet, kann sie nachteilige Auswirkungen auf Kinder, aber auch auf andere vulnerable Gruppen der Gesellschaft, bspw. Bildungsbenachteiligte oder Menschen mit geistiger Behinderung haben. Alle diese Gruppen besitzen ein erhöhtes Risiko, durch KI-Systeme ausgenutzt oder fehlgeleitet zu werden.

Der europäische Gesetzgeber sollte bei der Umsetzung des Risikomanagementsystems für ein Hochrisiko-KI-System vorschreiben, dass nicht nur deren Zugänglichkeit für Kinder, sondern für alle vulnerablen Personengruppen insgesamt und deren Auswirkungen auf sie, berücksichtigt werden. Die besondere Schutzbedürftigkeit vulnerabler Gruppen wird auch in Art. 5 Abs. 1 lit. b KI-VO-E sowie Art. 7 Abs. 2 lit. f KI-VO-E beachtet. In Anlehnung an den letztgenannten Artikel sprechen wir uns dafür aus, in Art. 9 Abs. 8 KI-VO-E nicht nur *„Kinder“*, sondern *„Kinder und aufgrund eines Ungleichgewichts in Bezug auf Machtposition, Wissen, wirtschaftliche oder soziale Umstände schutzbedürftige Personen“* miteinzubeziehen.

6. Menschliche Aufsicht, Art. 14 KI-VO-E

Art. 14 KI-VO-E bestimmt, dass Hochrisiko-KI-Systeme während des Betriebs von natürlichen Personen wirksam beaufsichtigt werden müssen (menschliche Aufsicht). Damit folgt die KOM dem in ihrem Weißbuch vorgeschlagenen Ansatz für eine menschliche Aufsicht, die Überwachung während des Betriebes eines KI-Systems in Echtzeit zu ermöglichen.²⁴ Hierbei setzt die KOM in Art. 14 Abs. 4 KI-VO-E gleichzeitig auf drei Lenkungs- und Kontrollmechanismen. Sie fordert zum einen die Fähigkeit des Menschen, das System vollständig zu verstehen und damit in jeden Entscheidungszyklus eingreifen zu können (*Human in the loop* - lit. a), die Fähigkeit, mit einer *„Stopptaste“* das System unterbrechen zu können (*Human on the loop* - lit. e) und die Möglichkeit, das Ergebnis des KI-Systems nicht zu verwenden und/oder das Ergebnis anderweitig außeracht zu lassen (*Human on command* - lit. d). Die menschliche Aufsicht kann dabei durch zwei

²⁴ WEISSBUCH Zur Künstlichen Intelligenz – ein europäisches Konzept für Exzellenz und Vertrauen, S. 25.

Vorkehrungen nach Art. 14 Abs. 3 KI-VO-E alternativ gewährleistet werden: Entweder wird die menschliche Aufsicht „*vor dem Inverkehrbringen oder der Inbetriebnahme vom Anbieter bestimmt und, sofern technisch machbar, in das Hochrisiko-KI-System eingebaut*“ (lit. a) oder sie ist dazu geeignet, „*vom Nutzer umgesetzt zu werden*“ (lit. b).

Das Katholische Büro begrüßt das Erfordernis der menschlichen Aufsicht in Art. 14 KI-VO-E ausdrücklich. Dabei ist wichtig, dass auch im Fall einer menschlichen Aufsicht nach Art. 14 Abs. 3 lit. a KI-VO-E („*in das System eingebaut*“) die nötige Distanz für eine kritische Beurteilung der Ergebnisse eines KI-Systems gewährleistet ist, die Art. 14 Abs. 4 lit. b KI-VO-E fordert. Nach dieser Norm muss zu Recht für beide Varianten der menschlichen Aufsicht des Art. 14 Abs. 3 KI-VO-E ermöglicht werden, dass „*sich [beaufsichtigende Personen einer] möglichen Neigung zu einem automatischen oder übermäßigen Vertrauen in das Hochrisiko-KI-System hervorgebrachte Ergebnis („Automatisierungsbias“) bewusst [...] bleiben*“.

7. Pflichten für Anbieter und Nutzer und andere Beteiligte, Art. 16-29 KI-VO-E

Die Kommission schlägt in ihrem KI-VO-E vor, im Kapitel 3 schwerpunktmäßig dem Anbieter umfassende Pflichten aufzuerlegen (vgl. Art. 16-29 KI-VO-E), wohingegen Händler, Einführer und Nutzer (deutlich) weniger Pflichten zugewiesen bekommen. Dem Nutzer obliegt primär die Pflicht, der Gebrauchsanweisung eines KI-Systems Folge zu leisten, Art. 29 Abs. 1 KI-VO-E. Hinzu kommen noch ein paar wenige in den Abs. 2-6 desselben Artikels geregelte Pflichten, wie etwa die Aufbewahrung automatisch von KI-Systemen erzeugter Protokolle (Abs. 5) oder das Durchführen einer Datenschutz-Folgenabschätzung nach der DS-GVO (Abs. 6).

Sollte unserem Vorschlag gefolgt werden, in die Legaldefinition des Nutzers nach Art. 3 Nr. 4 KI-VO-E auch den nicht beruflich handelnden Nutzer aufzunehmen (siehe oben), bedarf es der Modifikation von Art. 29 KI-VO-E: Um einer Überregulierung vorzubeugen und den Einzelnen auch im Privaten zu ermutigen, die technologischen Ressourcen anzunehmen, wäre zu überlegen, ob jedenfalls dem privaten Nutzer die Befolgung der dem System beigefügten Gebrauchsanweisung i.S.d. Art. 29 Abs. 1 1. Hs. KI-VO-E aufzuerlegen wäre. Von weiteren Pflichten sollte dagegen abgesehen werden.

8. Konformitätsbewertung, Art. 43 KI-VO-E und Harmonisiert Normen, Art. 40 KI-VO-E

Mit dem 5. Kapitel schlägt die KOM ein Konformitätsbewertungsverfahren vor. Art. 43 KI-VO-E unterscheidet dann bei der Festlegung der Anforderungen einer Konformitätsbewertung zwischen KI-Systemen, die dem Bereich 1 des Annex III unterfallen, und solchen, die den Bereichen 2- 8 des Annex III zugehörig sind.

Die erstgenannte Gruppe betrifft die Konformitätsbewertung ausschließlich für biometrische Fernidentifizierungssysteme, bei welchen entweder die Übereinstimmung mit den grundlegenden Anforderungen des Kapitel 2 (Art. 43 Abs. 1, 2 KI-VO-E) oder mit den harmonisierten Normen nach Art. 3 Nr. 27 KI-VO-E (Art. 40, 43 Abs. 1 KI-VO-E) bzw. den gemeinsamen Spezifikationen (Art. 41, 43 Abs. 2 KI-VO-E) bestehen muss. Ist eine Übereinstimmung mit den Anforderungen in Art. 40 KI-VO-E oder Art. 41 KI-VO-E gegeben, ist das Konformitätsbewertungsverfahren nach Anhang VI ausreichend. Nur, wenn KI-Systemen nach Anhang III Nr. 1 eine solche Kongruenz nicht aufweisen, muss das strengere Verfahren des Anhangs VII befolgt werden. KI-Systeme, die die Bereiche des Annex III Nr. 2-8 betreffen, müssen dahingegen stets nur das Konformitätsverfahren nach Annex VI beachten. Der wesentliche Unterschied der Verfahren nach Anhang VI und VII ist,

dass nach Anhang VI eine externe Überprüfung der Einhaltung der in Kapitel 2 genannten Anforderungen nicht erfolgt und eine durchgeführte Selbstbewertung durch den Anbieter ausreichend ist. Die Einbindung einer notifizierten Stelle ist ausdrücklich ausgenommen.

Das Katholische Büro begrüßt, dass alle von dem KI-VO-E erfassten KI-Systeme einer Konformitätsbewertung unterliegen sollen, die nach Anhang VI entweder eine interne oder nach Anhang VII eine externe Kontrolle fordert.

Allerdings geht diese aus drei Gründen nicht weit genug.

Erstens bedarf es zur Sicherstellung, dass die KI-Systeme, wie es die KOM in Erwägungsgrund 62 vorsieht, ein „*hohes Maß an Vertrauenswürdigkeit*“ besitzen, einer ex-ante Evaluierung mittels einer Konformitätsbewertung durch unabhängige Dritte. Allerdings legt – wie soeben dargelegt – Art. 43 Abs. 1 KI-VO-E nur für den Bereich der biometrischen Fernidentifizierungstechniken eine Konformitätsbewertung durch eine extern zu benennende Stelle fest. Wir sehen kritisch, dass im Übrigen eine „interne Kontrolle“ ausreichend sein soll. Vor allem eine externe Stelle nach einem „Zwei-Augen-Prinzip“ gewährleistet hinreichend Rechtssicherheit und stärkt zudem das Vertrauen in die Sicherheit von KI-Systemen, das erforderlich ist, um neben den Anbietern auch andere Akteure zu der Verwendung von KI-Systemen mit hohem Risiko zu motivieren.

Zweitens sieht Art. 40 KI-VO-E eine sog. „Konformitätsvermutung“ nach dem 2008 erneuerten New Legislative Framework Verfahren vor. Bei diesem Verfahren kann bei der Anwendung von harmonisierten Normen von einer Übereinstimmung mit den entsprechenden Anforderungen der jeweiligen EU-Richtlinie ausgegangen werden. Grundsätzlich ist dieses Prozedere sinnvoll, da es bereits existierende Normierungen anwendet und eine Dopplung etwa von Verfahren zur Konformitätsbewertung vermieden werden kann. Daher sollte die Normung wie in Erwägungsgrund 61 beschrieben, „*eine Schlüsselrolle dabei spielen, (...) die Einhaltung der Verordnung zu gewährleisten*“. In dem konkreten Fall des KI-VO-E liegt nach Art. 40 KI-VO-E eine Konformität mit den Vorschriften des Kapitels 2 vor, wenn die KI-Systeme die Anforderungen der auf der Grundlage des Gesetzes entwickelten und harmonisierten Normen – legaldefiniert in Art. 3 Nr. 27 KI-VO-E – erfüllt. Allerdings wird von manchen Autoren kritisiert, dass aktuell gültige harmonisierte Normen teilweise keine eindeutige Anleitung bieten, wie ein KI-System rechtssicher, fair und frei von Diskriminierung integriert werden kann.²⁵ Die spezifische Normanpassung dieser Vorschriften im Hinblick auf KI-Systeme sollte daher dringend und zügig erfolgen, damit insbesondere für biometrische Fernidentifizierungssysteme die vorgesehenen Kontrollmöglichkeiten ausreichend sind. Darüber hinaus regt das Katholische Büro an, grundsätzlich das Erfordernis einer externen Kontrolle für alle Hochrisiko-KI-Systeme zu normieren.

9. Transparenzpflichten für bestimmte KI-Systeme, Art. 52 KI-VO-E

Für „bestimmte“ KI-Systeme, unabhängig davon, ob sie Geringrisiko- oder Hochrisiko-KI-Systeme sind, fordert der Verordnungsentwurf die Einhaltung von „*Transparenzpflichten*“, Art. 52 KI-VO-E. Dies gilt für KI-Systeme, die für die Interaktion mit natürlichen Menschen bestimmt sind (Abs. 1), ein Emotionserkennungssystem oder ein System zur biometrischen Kategorisierung nutzen (Abs. 2), oder für Nutzer eines Systems, die ein

²⁵ vgl. auch Exam AI, N. Becker, P. Junginger, L. Martinez, D. Krupka, KI in der Arbeitswelt, Übersicht einschlägiger Normen und Standards, 2021, S. 14.

Bild-, Ton- oder Videoinhalt erzeugen oder manipulieren. Ausgenommen sind jeweils KI-Systeme, die zur Verwendung zur Aufdeckung, Verhütung, Ermittlung und Verfolgung von Straftaten gesetzlich zugelassen sind. Das Katholische Büro begrüßt die vorgesehenen „*Transparenzpflichten*“. Insgesamt gehen diese jedoch nicht weit genug.

In dem „Rome Call for AI Ethics“ wird als Mindestvoraussetzung für ethische KI genannt, dass jedem Menschen bewusst sein muss, wenn er oder sie mit einer Maschine interagiert.²⁶ Diese dem Nutzer zu transferierende Information ist die Grundvoraussetzung für die autonome Entscheidung des Einzelnen, ob er überhaupt ein Produkt mit integrierter KI oder eine Software nutzen möchte.

Bei den in Art. 52 Abs. 1 KI-VO-E bereits vorgesehenen Transparenzpflichten erscheint bedenklich, dass sie nur für solche KI-Systeme gelten, die von vornherein für die „*Interaktion mit natürlichen Personen*“ bestimmt sind. Einerseits besteht bei KI-Systemen die Möglichkeit, dass sie den ursprünglichen Anwendungsbereich verlassen und entgegen der ursprünglichen Bestimmung in Interaktion mit einer natürlichen Person treten. Andererseits kommt es wesentlich darauf an, ob der Entwickler selber als Verwendungszweck die Interaktion mit Menschen ansieht und als solche bestimmt. Damit besteht in dieser Fallkonstellation ein erhebliches Umgehungspotential der „*Transparenzpflichten*“ und lässt also viel Spielraum für die Interaktion mit KI-Systemen, die nicht als solche erkennbar sind. Aus kirchlicher Sicht sollte der Anwendungsbereich daher nicht durch den Bestimmungszweck beschränkt sein.

Darüber hinaus wäre die Erstellung einer EU-Datenbank wünschenswert, in der nicht nur – wie derzeit in Art. 51 KI-VO-E vorgesehen – Hochrisiko-KI-Systeme registriert werden, sondern zur Förderung der Transparenz und als wichtige Maßnahme für das Schaffen von Vertrauen in KI-Systeme alle derzeit in der EU auf den Markt befindliche KI-Systeme aufgeführt sind. Zur weiteren Steigerung der Transparenz könnte eine generelle Kennzeichnungspflicht nicht nur für Hochrisiko-KI-Systeme, sondern für alle KI-Systeme festgeschrieben werden.

10. Bereits in Verkehr gebrachte oder in Betrieb genommene KI-Systeme, Art. 83 KI-VO-E

In Art. 83 Abs. 2 sieht der KI-VO-E vor, dass alle Hochrisiko-KI-Systeme – ausgenommen solche KI-Systeme, die als Komponente eines IT-Großsystems dem Art. 83 Abs. 1 KI-VO-E unterfallen –, die vor dem Datum des Inkrafttretens der Verordnung „*in Verkehr gebracht oder in Betrieb genommen wurden, nur dann, wenn diese Systeme danach in ihrer Konzeption oder Zweckbestimmung wesentlich geändert wurden*“, dem Anwendungsbereich der Verordnung unterfallen. Bereits heute sind aber bereits eine Reihe von KI-Systemen im Einsatz, deren Zweckbestimmung aller Voraussicht nach nicht mehr wesentlich geändert werden wird. Solche Systeme wären dann nicht von dem Anwendungsbereich des KI-VO-E erfasst, sodass ihr Einsatz für Praktiken nach Art. 5 KI-VO-E nicht verboten wäre und sie sich auch im Übrigen ohne eine KI-spezifische Regulierung auf dem Markt befinden. Als Beispiel für ein KI-System, das bereits jetzt erprobt wird und daher nach derzeitigem Stand nicht den Anforderungen des KI-VO-E gerecht werden müsste, ist ein Pilotprojekt aus Mannheim: Im öffentlichen Raum erfolgt hier eine intelligente, auf Algorithmen basierende Videoüberwachung, die zum Ziel hat, die Straßenkriminalität im öffentlichen Raum zu minimieren.²⁷ Dieses System verwendet eine Erkennung von

²⁶ Redemanuskript v. Papst Franziskus zum „Rome Call for AI Ethics“, 2020.

²⁷ <https://www.iosb.fraunhofer.de/de/projekte-produkte/intelligente-videoueberwachung.html>.

Verhaltensmustern und würde, auch wenn etwa eine weitere Komponente wie die biometrische Fernidentifizierung neben der Verhaltenserkennung in das KI-System integriert würde, aus dem Anwendungsbereich des KI-VO-E fallen. Denn aufgrund der weiterhin bestehenden identischen Zweckbestimmung, nämlich der Bekämpfung der Straßenkriminalität, würde Art. 83 Abs. 2 KI-VO-E nicht greifen. Das Katholische Büro spricht sich daher dafür aus, die Eingrenzung des Anwendungsbereichs auf bereits in Betrieb genommene oder in Verkehr gebrachte KI-Systeme, die ihren Zweck wesentlich ändern, aufzuheben. Dies könnte dadurch erfolgen, dass der letzte Halbsatz von Art. 83 Abs. 2 KI-VO-E („*nur dann, wenn diese Systeme danach in ihrer Konzeption oder Zweckbestimmung wesentlich geändert wurden*“) gestrichen wird.

Berlin, den 9. Mai 2022

Kommissariat der deutschen Bischöfe
– Katholisches Büro in Berlin –
Hannoversche Straße 5,
10115 Berlin,
Tel.: 030 – 28878-0,
Mail: post@kath-buero.de