Deutscher Bundestag
Ausschuss für Digitales

Ausschussdrucksache
20(23)83

26.09.2022

CISPA
HELMHOLTZ CENTER FOR INFORMATION SECURITY

Saarbrücken, 23. September 2022

Stellungnahme anlässlich der öffentlichen Anhörung am 26. September 2022 zur „EU-Verordnung zu Künstlicher Intelligenz unter Einbeziehung von Wettbewerbsfähigkeit im Bereich Künstliche Intelligenz und Blockchain-Technologie“

Prof. Dr. Dr. h. c. Michael Backes ist Gründungsdirektor und CEO des CISPA Helmholtz-Zentrum für Informationssicherheit. Das CISPA Helmholtz-Zentrum für Informationssicherheit ist eine Großforschungseinrichtung des Bundes innerhalb der Helmholtz-Gemeinschaft. CISPA Wissenschaftler:innen erforschen die Informationssicherheit in all ihren Facetten. Sie betreiben modernste Grundlagenforschung sowie innovative anwendungsorientierte Forschung und arbeiten an drängenden Herausforderungen der Cybersicherheit, der Künstlichen Intelligenz und des Datenschutzes. CISPA-Forschungsergebnisse finden Einzug in industrielle Anwendungen und Produkte, die weltweit verfügbar sind. Damit stärkt das CISPA die Konkurrenzfähigkeit Deutschlands und Europas. Es fördert außerdem Talente und ist eine Kaderschmiede für hervorragend ausgebildete Fach- und Führungskräfte für die Wirtschaft. So trägt das CISPA sein Know-how auch in die Zukunft.

1) **Bei der fallweisen und sektoralen Erfassung von Risiken und Bedrohungen und dem Wunsch von z.B. Unternehmensverbänden bezüglich Sandboxen und einer Kombination von ex ante Risiko-Selbstbewertung und ex-Post Durchsetzung bei KI mit als hoch eingestuftem Risiko stellt sich insbesondere mit Blick auf die Empfehlung der Kommission zum Einsatz von KI in öffentlichen Diensten als sogenannte „Test Umgebung" die Frage, inwiefern und wodurch sichergestellt werden kann, dass**
    - **trotz Kombination mehrerer Anwendungen keine Schäden gerade bei Einsatz in kritischen Infrastrukturen entstehen,**
    - **der Grad möglicher Gefährdungslagen auch unter Einbeziehung von sektorübergreifenden Schnittstellen erhoben und**
    - **allgemeine Haftungsfragen geklärt, mögliche Sicherheitslücken gefunden, gemeldet und schnell behoben werden können?**

Die Kombination mehrerer KI-Anwendungen ist ein komplexer Prozess, da die Komposition und Interaktion verschiedener KI-Systeme bisher kaum erforscht sind. Auch innerhalb von Testumgebungen sollte die Kombination mehrerer KI-Anwendungen deshalb nur in sorgsam geprüften Fällen umgesetzt werden. Kritische Infrastrukturen eignen sich generell nur sehr eingeschränkt als Testumgebung, auch hier sollte ein Einzelfallprüfung durchgeführt werden.

Allgemeine Haftungsfragen werden in der KI-VO teilweise in Artikel 16 (Pflichten der Anbieter von Hochrisiko-KI-Systemen) und 24 (Pflichten der Produkthersteller) geregelt, hier besteht aber noch Konkretisierungsbedarf. Insbesondere die Haftung bei Fehlfunktionen bei komplexen KI-Systemen sollte spezifischer geregelt werden.

Falls in Testumgebungen Sicherheitslücken gefunden werden, sollten diese gemeldet und schnell behoben werden. Die Vorgaben „security-by-design/default" sollten auch für KI-Systeme gelten, um das Ziel einer „Trustworthy AI" zu erreichen.

2) **Im Bereich der Blockchain-Technologie fordern führende Kryptoexpert*innen in einem Brief an den US-Kongress Regulierung von Kryptoassets sowie eine Überprüfung in Bezug auf den Mehrwert beim Einsatz von Blockchain-Technologien. Im Bereich der KI-Anwendungen wird von umfassenden Herausforderungen bei der Regulierung von Algorithmen zur Gestaltung sozio-ökonomischer und ökologischer Implikationen ausgegangen. Kann bei beiden Technologien gewährleistet werden, dass eine angemessene Anwendung nach Kosten-Nutzen-Abwägung erfolgt, z.B. im Bereich des ökologischen Ressourcenverbrauchs, oder dem Generieren eines echten Mehrwerts gegenüber klassischen IT-Anwendungen und wenn ja, anhand welcher messbarer Kriterien und werden gesellschaftliche/soziale Folgewirkungen ausreichend berücksichtigt?**

Zunächst möchte ich anmerken: Blockchain-Technologien und KI-Anwendungen sind zwei völlig verschiedene Themenkomplexe, die nicht einfach miteinander verglichen werden können und deren Kosten-Nutzen-Abwägung unterschiedlich sind. In meinem Statement möchte ich nur kurz auf Blockchains eingehen, der Fokus dieser Stellungnahme liegt auf KI-Systemen. Ein Teil der technischen Kritik im Brief an den US-Kongress bezüglich Blockchain-Technologie ist aus technischer Sicht inkorrekt (insbesondere die Behauptung, dass umkehrbare Transaktionen oder der Schutz der Privatsphäre nicht möglich sind). Dennoch sind die grundsätzlichen Bedenken valide und eine Regulierung zum Schutz von Nutzern und Investoren ist generell auch in diesem Bereich angebracht. Es kann nicht grundsätzlich davon ausgegangen werden, dass die Blockchain-Technologie einen Mehrwert oder ein optimales Kosten-Nutzen-Verhältnis bringt. Dieser Aspekt ist abhängig von der Anwendung und bei vielen Anwendungen noch offen. Ein positiver Effekt hat sich aber bereits in der

Forschung gezeigt, wo durch die Prominenz der Blockchain-Technologie in den letzten Jahren große Fortschritte und technische Durchbrüche bei Technologien ermöglicht wurden, die im Kontext von Blockchains benutzt werden, aber auch in anderen Bereichen Anwendung finden können. So gab es z.B. viele Verbesserungen im Bereich von sog. Zero-Knowledge Proofs, die auch in anderen Bereichen zum Schutz der Privatsphäre eingesetzt werden können. Verschiedene Aspekte von KI-Anwendungen werden bei den Antworten auf andere Fragen, insbesondere Frage 12, behandelt.

3) Inwieweit wird sich die KI-Verordnung auf die Wettbewerbsfähigkeit Europas im internationalen Vergleich auswirken?

Zur Auswirkung des Entwurfs der KI-VO auf die Wettbewerbsfähigkeit der europäischen Wissenschaft und Forschung möchte ich auf die Antwort der Frage 14 verweisen.

4) Kann die KI-Verordnung in der Entwurfsfassung Diskriminierung zum Beispiel gegenüber Frauen oder PoC verhindern? Wo muss gegebenenfalls nachgesteuert werden?

Der Entwurf der KI-VO und auch der Kompromissvorschlag der tschechischen Ratspräsidentschaft adressieren Diskriminierungsrisiken für bestimmte Gruppen nicht explizit in den Normtexten. Die Nichtdiskriminierung als EU-Grundrecht nach Art. 21 der EU-Grundrechtecharta ist Schutzgegenstand des gesamten Entwurfs, der den Schutz aller Grundrechte von KI-Entscheidungen Betroffener zum Ziel hat.

Eine herausgehobene Stellung des Schutzes vor Diskriminierung und dem Risiko, historische Diskriminierungsmuster fort zu schreiben, findet sich aber in den Erwägungsgründen zur Risikobewertung von biometrischen Identifikationsverfahren, Social Scoring, KI-Systemen zum Recruiting und dem Einsatz von KI für den Zugang zu öffentlichen Diensten und Leistungen.

Zielführend erscheint es, den Schutz vor Diskriminierung als inhärenten Teil des Qualitätsmanagements von KI-Systemen in den Entwurf zu integrieren. Ansätze dazu sind bereits im Kompromissvorschlag der tschechischen Ratspräsidentschaft zu finden. So sieht Art. 10 (Data and Data Governance) Abs. 2 lit. f vor, dass für Hochrisikosysteme die Trainingsdaten auf Diskriminierungsrisiken hin untersucht werden sollen. Annex VI zur Technischen Dokumentation führt diesen Ansatz weiter aus und sieht vor, dass die verpflichtende technische Dokumentation für Hochrisikosysteme detaillierte Informationen zu erwartbaren unerwünschten Ergebnissen, Risiken für Grundrechtseinschränkungen und explizit auch Diskriminierungsrisiken enthalten soll.

Begleitend wäre eine Integration von Anforderungen an Nichtdiskriminierung in Best-Practices, technische Standards und Zertifizierungen zentral. Darüber sehe ich es als zwingend notwendig an, dass in den Bereichen Fairness und Erklärbarkeit von KI-Systemen noch mehr geforscht wird, um die Grundursachen von Diskriminierung durch solche Systeme besser zu verstehen und entsprechende Gegenmaßnahmen zu entwickeln, die diese Problematik auf einer tiefen Ebene bereits teilweise oder ggf. sogar vollständig abfangen (siehe auch Frage 12).

5) Sind die in der DSGVO und im Verordnungs-Entwurf verankerten Regelungen zu Informations- und Beschwerderechten für Betroffene von KI-Entscheidungen ausreichend? Und wie könnten Betroffene für jene Rechte sensibilisiert werden?

Die in Art. 52 des Entwurfs vorgesehenen Transparenzpflichten für bestimmte KI-Systeme sind ein wichtiger Ansatz, allerdings gehen sie nicht weit genug. Bedenkenswert wäre für bestimmte Einsatzgebiete mit hohem Risiko ein Recht der Betroffenen auf opt-out von Entscheidungen eines KI-Systems oder alternativ ein Beschwerderecht, das zur Überprüfung der Entscheidung durch einen Menschen führt. In der DSGVO werden Aspekte wie das Recht auf Auskunft, das Recht auf Berichtigung und Löschung, das Recht auf Verarbeitungseinschränkung der Daten, das Recht auf Widerspruch der Datenverarbeitung oder das Recht auf Datenübertragbarkeit verankert. Es sollte geprüft werden, ob diese Betroffenenrechte auch gegenüber KI-Systemen angemessen durchgesetzt werden können oder ob flankierende Regelungen in der KI-VO sinnvoll wären.

Aus Forschungssicht sind KI-Systeme häufig noch Black Boxes, also Entscheidungen können nicht immer im Detail nachvollzogen werden (siehe auch Frage 12). Hier gibt es einen sehr großen Bedarf an weiterer Forschung, um Erklärbarkeit und Transparenz von KI-Systemen zu verbessern und eine forensische Analyse von Entscheidungen zu ermöglichen.

6) Wie verlässlich ist eine Konformitätsbewertung von Hochrisiko-Anwendungen, die durch die Anbieter selbst durchgeführt wird? Brauchen wir gerade in sensiblen Bereichen eine externe Prüfung?

In diesem Bereich erfolgt ein produktsicherheitsrechtlicher Ansatz der Konformitätsbewertung anhand technischer Standards. Die Verlässlichkeit wird davon abhängen, ob, wann und unter welchen Bedingungen derartige Standards entwickelt werden können. Eine Prüfung durch Anbieter führt zu einem einfacheren Prozess, erlaubt aber ggf. den laxen Umgang mit einer entsprechenden Prüfungspflicht. In Risiko- und Hochrisikobereichen sollte eine externe Prüfung erfolgen, um eine sorgfältige und nachvollziehbare Prüfung sicherzustellen. Dies sollte aber so ausgestaltet werden, dass Innovationen nicht gehemmt und die Kosten sowie der Aufwand überschaubar bleiben. Eine Überregulierung sollte vermieden werden, um die Chancen und Potentiale auch solcher Anwendungen effektiv nutzen zu können.

Eine Pflicht der KI-Betreiber, die Aufsichtsbehörde über Vorfälle mit Gefährdung der Grundrechte zu unterrichten, ist sinnvoll. Hier sollte ein ähnliches Prinzip wie die Anzeigepflicht im Datenschutzrecht umgesetzt werden. Aber hier sind auch ähnliche Hemmnisse in der Praxis zu erwarten, eine entsprechende Regulierung mit positiven Anreizen für eine Selbstanzeige erscheint sinnvoll.

7) Sehen Sie wesentliche begriffliche Unklarheiten in der KI-VO und, falls ja,

- **welche regulatorischen Komplikationen ergeben sich möglicherweise daraus,**
- **und wie ließen sich derartige Komplikationen vermeiden oder beheben?**

Ich verweise hier auf meine Antwort zu Frage 13.

8) Laut einer aktuellen Umfrage von Bitkom betrachten 49 Prozent der befragten Unternehmen rechtliche Unsicherheiten als Hemmnis für die Einführung von KI-Anwendungen. Wird die KI-VO ihrer Meinung nach zu einer Verbesserung der Situation führen, oder könnte sie sie ggfs. sogar verschärfen – insbesondere für KMUs?

Die KI-VO ist ein sehr wichtiger Schritt für Deutschland und Europa, um Rechtssicherheit im Bereich der KI-Systeme zu schaffen und ihren Einsatz zu regulieren. Ähnlich wie schon bei der Datenschutzgrundverordnung und ähnlichen Verordnungen wird die EU hier ein globaler Wegbereiter sein und ein internationales Vorbild schaffen, wie KI-Systeme das Konzept der Wertgebundenheit und Vertrauenswürdigkeit von Technologie durch Regulierung umsetzen können („Brüssel-Effekt"). Diese Rechtssicherheit ist richtig und wichtig, um Unternehmen, Forschungseinrichtungen und auch der Zivilgesellschaft klare Regeln vorzugeben.

Dabei ist wichtig, zu beachten, dass die Forschungs- und Entwicklungstätigkeit derer, die an von Grund auf vertrauenswürdigen, sicheren, transparenten und fairen Systemen arbeiten, nicht behindert werden darf. Nur so werden europäische Forscherinnen und Forscher nicht strukturell benachteiligt, und nur so kann die EU im internationalen Wettbewerb mit China und den USA bestehen. Deshalb ist es von zentraler Wichtigkeit, dass die geplante KI-Verordnung forschungs- und innovationsoffen oder besser sogar forschungs- und innovationsfördernd gestaltet wird. Eine generelle Privilegierung für KI-Forschung, wie sie im Kompromissvorschlag der tschechischen Ratspräsidentschaft vorgesehen ist, halte ich für zentral.

Die vorgesehenen Erleichterungen für KMU und Start-ups im vorliegenden Entwurf sind derzeit eventuell nicht ausreichend, um das Ziel der Innovationsförderung zu erreichen. Die Privilegierungen erschöpfen sich weitgehend darin, den Betreibern nach Art. 55 Abs. 1 a) KI-VO-E leichteren Zugang zu den „regulatory sandboxes" zu gewähren und in der Schaffung besonderer Kommunikationsmaßnahmenkanäle, um Leitlinien, Hilfen und Best Practices für die Anwendung der KI-VO-E für diese Unternehmen zu geben. Dies wurde durch den Kompromissvorschlag der tschechischen Ratspräsidentschaft nochmals erweitert. Eine weitergehende Erleichterung sollte angestrebt werden, um Innovationen nicht zu hemmen. Einen Schritt in eine sinnvolle Richtung geht der Kompromissvorschlag der tschechischen Ratspräsidentschaft mit dem Entwurf des Art. 55a. Dieser eröffnet die Möglichkeit, Kleinstunternehmen teilweise von kostspieligen Verpflichtungen wie einem umfangreichen Qualitätssicherungssystem zu befreien, wenn dies das Schutzniveau nicht negativ beeinträchtigt. Hier sollte in Zusammenarbeit mit KMU-Stakeholdern über weitere sinnvolle Privilegierungen diskutiert werden, solange dies nicht zu höheren Risiken führt.

Bedeutsamer dürfte demgegenüber die Verpflichtung der Aufsichtsbehörden sein, bei der Festsetzung von Gebühren für die Konformitätsbewertungsverfahren die Größe des betroffenen Unternehmens zu berücksichtigen., Art. 55 Abs. 2 KI-VO-E.

Auch die Ausgestaltung der regulatory sandboxes ließe sich innovationsfreundlicher gestalten. Dazu verweise ich auf meine Antwort zu Frage 14.

Zuletzt sollte geprüft werden, wie sich Mehrfachregulierungen und daraus resultierende Überregulierung und Rechtsunsicherheit reduzieren ließen, die gerade KMUs und Start-ups in besonderer Weise belasten (siehe auch Frage 13).

9) Die Bundesregierung hat deutlich gemacht, dass sie KI-Anwendungen für Sicherheitsbehörden vom Hauptvertragstext getrennt regeln will. Denken Sie, dass dies sinnvoll ist und was sind die Vor- und Nachteile einer getrennten Regulierung?

Die aktuelle Fassung des Entwurfs sowie der vorliegende Kompromissvorschlag nehmen KI-Systeme für militärische und Zwecke der nationalen Sicherheit vom Anwendungsbereich der KI-VO explizit aus.

Eine getrennte Regelung für diesen Bereich ist nicht notwendigerweise von Nachteil: KI im militärischen Einsatz wie KI-gestützte Aufklärung, Überwachung und Waffensysteme gehen einher mit höchsten Risiken für Menschen. Insofern erscheinen umfangreichere Vorgaben und Pflichten, gegebenenfalls die (international koordinierte) Ächtung bestimmter Anwendungen und Einsatzszenarien sinnvoll.

Die getrennte (nationale oder europäische) Regulierung sollte jedoch nicht zu einer Verzögerung führen. Eine Parallelregulierung sollte mit vergleichbarem Nachdruck vorangetrieben werden.

An dieser Stelle sei auch auf den Koalitionsvertrag verwiesen: „Für waffentechnologische Entwicklungen bei Biotech, Hyperschall, Weltraum, Cyber und KI werden wir frühzeitig Initiativen zur Rüstungskontrolle ergreifen."

10) In welchem Maße ermöglicht die KI-Verordnung Bürgerinnen und Bürgern, den Einsatz von KI-Systemen zu erkennen, zu verstehen und ihre Rechte wahrzunehmen, wenn sie von Entscheidungen oder Entscheidungsvorbereitungen durch KI betroffen sind und sind die Transparenzanforderungen aus Artikel 52 Satz 1 ausreichend, um darüber zu informieren, dass KI-Systeme automatisiert oder halb-automatisiert Entscheidungen treffen oder vorbereiten oder beeinflussen?

Bürgerinnen und Bürger sollten durch explizite Hinweise informiert werden, wenn sie mit KI-Systemen interagieren, falls dies aufgrund der Umstände und des Kontexts der Nutzung nicht offensichtlich ist. Bei Chat-Bots kann dies bspw. durch eine explizite Nachricht zu Beginn der Interaktion umgesetzt werden, bei anderen KI-Anwendungen sollten entsprechende Informationsmöglichkeiten entwickelt und umgesetzt werden. Vermieden werden sollte aber eine Situation ähnlich zur aktuellen Umsetzung bei Cookie-Bannern auf Webseiten und in Apps. Diese sind so allgegenwärtig, störend und verwirrend gestaltet, dass Betroffene sie nur noch wegklicken, ohne dass die Information tatsächlich wahrgenommen wird.

Zusätzlich zu Transparenz- und Informationspflichten sollte auch diskutiert werden, ob für bestimmte KI-Systeme mit gravierenden Auswirkungen auf die Rechte der Betroffenen diesen weitere Möglichkeiten zur Selbstbestimmung und Wahrnehmung von Rechten gegeben werden sollten (siehe Antwort auf Frage 5). Dies kann auch eine Einwilligungspflicht oder Opt-out Möglichkeit für bestimmte KI-Anwendungen umfassen.

Die in Art. 52 Abs. 3 vorgesehene Kenntlichmachung von Deep-Fakes ist zu begrüßen. Es steht allerdings zu befürchten, dass diejenigen, die Deep-Fakes zu Zwecken der Desinformation, Manipulation, Hate Speech oder anderen Straftaten erstellen oder weiter verbreiten, diese Kenntlichmachung unterlassen oder bestehende Kenntlichmachungen entfernen. Hier muss ergänzend weiter an Werkzeugen geforscht werden, um es nutzerseitig zu ermöglichen, Deep-Fakes effektiv zu erkennen.

11) Mit Blick auf große Datensets für Gemeinwohl/Forschungsdaten: Gehen Sie davon aus, dass Forschungssandboxes so gestaltet werden können, dass keine strukturellen Einschränkungen von Datenschutz durch Nutzung von Forschungsdaten (Beispiel europäischer Gesundheitsdatenraum) erfolgen kann? Wenn ja wie und wenn nicht, könnten Sie bitte die Gründe ausführen?

Der Entwurf der KI-VO sieht in Erwägungsgrund 45 explizit vor, Forschern und Forscherinnen für die Entwicklung von Hochrisiko KI-Systemen Zugang zu hochqualitativen Daten, wie beispielsweise im European Health Data Space, zu gewähren. Großforschungszentren in der EU, unter anderem auch das CISPA, haben mehrjährige Erfahrung darin, auch beim Umgang mit hochsensiblen Daten wie beispielsweise Medizindaten höchste Datenschutz- und Datensicherheitsanforderungen umzusetzen. Wie dies im Bereich der KI-Forschung und -entwicklung umgesetzt werden kann (beispielsweise Sicherung der durch KI gewonnenen Modelle gegen Rückschlüsse auf einzelne Trainingsdaten) ist Gegenstand unserer laufenden Forschung, mit der wir bereits große Erfolge für eine weitere Verbesserung des Patientendatenschutzes erzielen konnten.

In der Praxis bestehen also vielfältige organisatorische vor allem aber auch technische Möglichkeiten, um Forschungssandboxes datenschutzfreundlich zu gestalten.

Der Entwurf der KI-VO sieht jedoch selbst gewisse strukturelle Einschränkungen des Datenschutzes gesetzlich vor, was wir aus Sicht der Forschungs- und Innovationsförderung für äußerst sinnvoll erachten. So wird z.B. in Art. 54 KI-VO für die regulatory sandboxes der datenschutzrechtliche Zweckbindungsgrundsatz dahingehend gelockert, dass eine Datenweiternutzung zur Entwicklung und Training von KI-Systemen in Bereichen des öffentlichen Interesses erlaubt wird, auch wenn dies nicht vom ursprünglichen Verarbeitungszweck gedeckt ist.

12) Sind die sozialen Auswirkungen von KI derzeit ausreichend erforscht, oder benötigt es eine spezifische Forschungsethik und strukturelle wissenschaftliche Forschung/Evaluation, um Anwendungsbeispiele z.B. aus dem Bereich der Sicherheitstechnik kritisch zu hinterfragen und sicherzustellen, dass KI nicht diskriminiert und Ungleichheit verfestigt?

KI-Systeme müssen in der Lage sein, Wahres von Falschem, Fakten von Fiktion und Fake News von echten Nachrichten zu unterscheiden. Aktuelle Lösungen nehmen jedoch i.a.R. keine Rücksicht auf die Vertrauenswürdigkeit - geschweige denn den Wahrheitsgehalt- von Informationen. Dies ist in vielen Bereichen problematisch, da es zu Systemen führt, die unwissentlich auf falsche Fakten reagieren können, bis hin zu KI-Systemen für soziale Medien, die aktiv die Verbreitung falscher Nachrichten fördern. Ebenso hat der Erfolg von KI-Systemen eine Schattenseite, da diese dazu verwendet werden können, sehr realistische Medien (z. B. Bilder, Videos oder Texte) zu erzeugen. Diese so genannten Deep-Fakes sind bereits allgegenwärtig und haben zu einem dramatischen Anstieg des Risikos von Täuschung durch Nachahmung, Verleumdung bis hin zu sehr detaillierten, speziell auf die Beeinflussung der öffentlichen Meinung zugeschnittenen falschen Darstellungen geführt. Die Auswirkungen von synthetischen Daten in Form von multimedialen Inhalten auf unsere Informationsgesellschaft, deren Kommunikation und die Demokratie sind noch nicht absehbar und müssen genauer untersucht werden. Uns fehlen fundierte Theorien und Algorithmen, um den Wahrheitsgehalt einzelner Aussagen, Texte, Bilder, Videos bis hin zu ganzen Datensätzen zu bestimmen, um Rückschlüsse auf die Zuverlässigkeit von Informationen aus mehreren Quellen zu ziehen und um Informationen aus Quellen, denen wir unterschiedliches Vertrauen entgegenbringen, optimal zu kombinieren. Entsprechende Forschung und Evaluation zu sozialen Auswirkungen von KI sollte unterstützt werden.

Auch die Technologiefolgenabschätzung wird durch zunehmend generalisierte KI erschwert, hier besteht ebenfalls noch viel Forschungsbedarf. Sicherheit, Vertrauenswürdigkeit und auch Privacy von

KI spielen ebenso eine zentrale Rolle, obwohl der aktuelle Kompromissvorschlag der Ratspräsidentschaft viele Passagen zu Fehlfunktion und Missbrauch entfernt hat. Aktuelle Forschungsprojekte widmen sich einerseits der Fairness von KI-Systemen und entwickeln Methoden, die einer potentiellen Diskriminierung durch KI-Systeme entgegenwirken. Andererseits spielt ein inklusiver Zugang zu neuen Technologien für alle eine ebenso wichtige Rolle.

Dabei ist zu beachten: die Stärke von KI-Lösungen liegt in ihrer extremen Flexibilität - anstatt sehr restriktive Annahmen darüber zu treffen, wie die Welt funktioniert, verfügen KI-Systeme über Milliarden von Parametern, die es ihnen ermöglichen, praktisch jeden datenerzeugenden Prozess zu erfassen. Diese Stärke hat ihren Preis: Aufgrund ihrer extremen Komplexität sind diese Systeme Black Boxes, bei denen man oft nicht sagen kann, warum und aus welchen Gründen Entscheidungen getroffen werden, wie sicher man sich einer Entscheidung ist, welche Änderung der Eingaben zu einem anderen Ergebnis geführt hätte, geschweige denn, welchem Teil ein Fehler (oder Erfolg) zuzuschreiben ist. Dies erschwert die Analyse von Fehlern eines bestimmten KI-basierten Systems, da die derzeitigen Methoden es uns nicht erlauben, zu untersuchen, wann und warum ein KI-basiertes System versagt hat. Mehr Forschung ist auch in diesem Bereich zwingend erforderlich.

13) An welchen Stellen sehen Sie bei den Kompromissvorschlägen der tschechischen Ratspräsidentschaft vom 15. Juli 2022 noch Verbesserungsbedarf, was die Definition von KI, die Bestimmung von Hoch-Risiko-Systemen und die Einstufung von KI-Anwendungen in Annex III betrifft?

Die Definition von KI im Entwurf war bisher extrem weit gefasst. Dies führte zu dem Problem, dass auch zahlreiche Systeme, die nicht KI im eigentlichen Sinn nutzen, von der Definition umfasst waren. Der Kompromissvorschlag der tschechischen Ratspräsidentschaft unternimmt Anstrengungen, den Begriff des KI-Systems weiter einzuschränken. Insbesondere die Ergänzung des Autonomieaspekts ist zu begrüßen. Allerdings krankt auch die neue Definition daran, dass sie versucht, KI teilweise tautologisch zu definieren: „KI ist, was Methoden der KI nutzt“. Auch wissenschaftlich gibt es keine klare Konsensdefinition von KI. Die Unterscheidung, was ein klassisches System und was ein KI-System ist, ist nicht trennscharf. Zentral ist die technische und vom Einsatz abhängige „Kompetenz“ eines Systems automatisiert und autonom Entscheidungen zu treffen und Vorhersagen zu machen. Insofern setzt die Definition den richtigen Schwerpunkt. Die fehlende Trennschärfe führt allerdings zu dem Risiko, dass IT-Systeme Gegenstand zahlreicher Normen werden (beispielsweise Mehrfachregulierung durch die KI-VO, den Cybersecurity Act, die DSGVO, den Digital Services Act, Vorgaben zu kritischen Infrastrukturen und sektorale Vorgaben wie die geplante Regulierung des European Health Data Space). Diese Vielzahl an parallelen Pflichten und Anforderungen führt zu Rechtsunsicherheit und einem Überhang an mehrfachen Dokumentations- und Berichtspflichten. Beides kann Innovation und Konkurrenzfähigkeit des europäischen Marktes massiv hemmen. Die rechtlichen Pflichten sollten daher unbedingt konsolidiert werden, um eine solche Mehrfachregulierung derselben Technologie zu vermeiden, gerade im Hinblick auf die notwendigerweise unscharfe Definition von KI.

Auch die Definition von Hochrisikosystemen bedarf einer genaueren Prüfung. Der Kompromissvorschlag der tschechischen Ratspräsidentschaft fasst die Definition von Hochrisikosystemen in Art. 6 (3) a extrem weit und erfasst alle KI-Systeme, deren Entscheidungen sofort effektiv werden ohne menschliche Prüfung. Eine solche Loslösung von Einsatzkontext und ohne Betrachtung des tatsächlichen Risikos beim Einsatz birgt jedoch die Gefahr der Überregulierung, indem hier auch KI-Systeme mit sehr niedrigem Risiko mit erfasst werden, nur weil sie keine weitere menschliche Prüfung (gegebenenfalls aber trotzdem Revisionsmöglichkeiten)

vorsehen. Um diese Überregulierung zu vermeiden, sollte die Einstufung als Hochrisikosystem auch von den möglichen Auswirkungen der KI-Entscheidungen auf Grundrechte der Betroffenen abhängig gemacht werden wie es in Alternative Art. 6 (3) b auch vorgesehen ist. Im neuen Entwurf vom 16. September wurde Artikel 6 (3) a gestrichen, diese Änderung ist zu begrüßen (siehe auch Erwägungsgrund 32).

Im neuen Entwurf vom 16. September wurde die Erläuterung zu Hoch-Risiko-Systemen aktualisiert. Die Entfernung der ersten beiden Punkte zu „law enforcement" und „public services" ist als kritisch anzusehen (siehe auch Frage 9).

14) Gibt die KI-Verordnung aus Ihrer Sicht genügend Freiraum für deutsche und europäische KI-Forschung, um mit den Forschungsbedingungen in den USA und China konkurrenzfähig zu sein und wo sehen Sie gegebenenfalls Bestimmungen der KI-Verordnung, die, gerade auch mit Blick auf die vorgesehenen Regelungen für Sandboxes, einschränkend auf zukünftige KI-Forschungsvorhaben, aber auch für den Transfer aus der Forschung in für den Markt zugelassene Produkte wirken könnten?

Freiraum für konkurrenzfähige Forschung

Aus Sicht der KI- und grundsätzlich der IT-Forschung ist besonders die durch den Kompromissvorschlag der tschechischen Ratspräsidentschaft vorgesehene Ausnahmeregelung in Art. 2 Abs. 6 und 7 zu begrüßen. Diese Regelungen nehmen KI-Systeme, die speziell für den alleinigen Zweck der wissenschaftlichen Forschung und Entwicklung entwickelt und betrieben werden sowie Forschungs- und Entwicklungstätigkeiten in Bezug auf KI-Systeme, sofern diese Tätigkeiten nicht das Inverkehrbringen des KI-Systems zum Ziel haben, vom Anwendungsbereich der Verordnung aus.

Eine solche Forschungsausnahme betrachten wir als essentiell, um die Forschungsfreiheit und Innovationsfähigkeit von wissenschaftlichen Einrichtungen zu wahren. Gerade wenn neue Methoden und Modelle für KI-Systeme von Grund auf neu entwickelt werden sollen, ohne bereits auf einen konkreten realen Einsatz abzuzielen, wären die umfangreichen Pflichten nicht nur unverhältnismäßig, sondern auch unnötig. Sofern tatsächlich Risiken entstehen könnten, indem beispielsweise Tests mit Forschungsdatensätzen und damit Datensätzen von realen Personen durchgeführt werden müssen, unterliegen die Forschungsvorhaben ohnehin bereits den umfangreichen Schutzpflichten der DSGVO. Die zusätzlichen Pflichten der KI-VO erscheinen in diesem Kontext insofern unverhältnismäßig, als dass die Forschungsvorhaben zu KI-Systemen im Sinne der Definition der KI-VO gerade die Beeinflussung des Umfelds durch KI-Vorhersagen, Entscheidungen oder Empfehlungen untersuchen und regelmäßig manuell begleitet werden und insofern in einem geschützten Raum stattfinden. Die Risiken für die Grundrechte Betroffener oder andere schwerwiegende Risiken sind daher schon durch den Einsatzkontext ausgeschlossen oder weitestgehend minimiert.

Zwar mag die wissenschaftliche KI-Forschung und der wissenschaftliche Einsatz und die Analyse von KI-Systemen auch durch eine enge Auslegung einzelner Definitionen und Normen vom Anwendungsbereich der KI-VO oder einzelner Pflichten ausgenommen werden, allerdings wären dennoch Chilling Effects für Forschungsvorhaben zu befürchten, wenn hier Rechtsunsicherheit herrschen würde. Eine klare Ausnahme wie im Kompromissvorschlag der tschechischen Ratspräsidentschaft vorgesehen, gäbe hier Forschenden die nötige Sicherheit, um sich in der internationalen Konkurrenz behaupten zu können.

Markttransfer und Sandboxregelungen

Als Instrument zur Innovationsförderung und zur Unterstützung von Start-ups und KMUs sieht der bisherige Verordnungsentwurf und auch der Kompromissvorschlag der tschechischen

Ratspräsidentschaft in Art. 53 sogenannte AI regulatory sandboxes („KI Reallabore“) vor. In diesen sandboxes sollen unter Aufsicht und Beratung der zuständigen Behörde(n) neue KI-Systeme unter Echtweltbedingungen erprobt werden bevor sie in den Markt eingeführt werden.

Die regulatory sandboxes sind ein neues regulatorisches Instrument. Grundsätzlich ist der Ansatz einer Inkubatorphase für neue KI-Systeme, in der diese durch behördliche Begleitung getestet und weiter entwickelt werden können bis zur Einsatzreife, interessant und begrüßenswert. Gerade auch aus Wissenschaftssicht könnte dies ein wertvolles Instrument im Transfer von wissenschaftlichen Forschungsprojekten hin zu marktreifen KI-Systemen sein. Ob die sandboxes in der Praxis angenommen werden und tatsächlich fördernd für Innovation und Adaption von KI-Systemen wirken können, wird entscheidend von der Ausgestaltung abhängen. Eine Förderung von KI-Reallaboren sowie entsprechenden Infrastrukturen ist begrüßenswert.

Der bisherige Entwurf sieht vor, dass sandboxes bei den zuständigen Behörden eines oder mehrerer Mitgliedstaaten oder vom Europäischen Datenschutzbeauftragten eingerichtet werden sollen. Der Kompromissvorschlag der tschechischen Ratspräsidentschaft ändert dies dahingehend, dass die zuständigen nationalen Aufsichtsbehörden sandboxes einrichten können. Auch wenn nationale Behörden eventuell niedrigschwelligeren Zugang zu sandboxes ermöglichen können, birgt dieser dezentrale Ansatz doch die Gefahr einer stark unterschiedlichen Ausprägung des sandbox-Prozesses. Unterschiedliche Anforderungen und unterschiedliche Betreuungsqualität könnten jedoch zu einem „sandbox shopping“ führen, bei dem sich Hersteller und Betreiber den Mitgliedsstaat mit den geringsten Anforderungen aussuchen. Ähnliche Effekte ließen sich vor der Harmonisierung auch im Datenschutz beobachten.

Der Erfolg der sandboxes wird auch davon abhängen, ob während des Prozesses tatsächlich Erleichterungen im Hinblick auf Rechtsanforderungen, Haftung und die konkreten Pflichten der KI-VO gewährt werden. Vorstellbar und sinnvoll erscheint beispielsweise die sukzessive Erreichung von Konformität als Bestandteil des sandbox-Prozesses. Zwar sollen die Behörden in der sandbox-Phase von Bußgeldern absehen, wenn der Betreiber/Hersteller kooperiert und den gemeinsam erstellten sandbox-Plan einhält. Unklar ist allerdings, ob während des Prozesses tatsächlich rechtliche Privilegierungen greifen. Dies allein in das Ermessen der nationalen Behörden zu stellen, könnte zu dem oben erwähnten Risiko des „sandbox shopping“ führen. Beispielhaft für eine solche unionsweite und innovationsfördernde Privilegierung während der sandbox-Phase ist die in Art. 54 vorgesehene Lockerung des datenschutzrechtlichen Grundsatzes der Zweckbindung.

Die Dezentralisierung der sandboxes wirft auch Fragen zur Machbarkeit und Skalierbarkeit der sandboxes auf. So wie die sandboxes derzeit konzipiert sind, steht zu befürchten, dass der Zugang bereits durch finanzielle und personelle Ressourcen der Mitgliedsstaaten vielerorts beschränkt sein wird. Damit durch dieses Instrument allerdings tatsächliche Innovationsförderung erzielt werden kann, muss der Ansatz skalierbar und leicht zugänglich sowie die Betreuung durch die Aufsichtsbehörde(n) agil sein. Wenn sich Sandboxverfahren mehrere Monate hinziehen bis die Unternehmen Rückmeldungen zu ihren Fragen und Vorschlägen erhalten, würde der eigentlich innovationsfördernde Ansatz der regulatory sandboxes scheitern.

Generell sollten solche Prozesse mit möglichst wenig Aufwand umsetzbar sein, um Innovationen nicht zu behindern und die Chancen sowie Potentiale von „Trustworthy KI made in EU“ besser zu nutzen. Mehr öffentliche Förderung in diesem Themenbereich ist essentiell, um im Wettbewerb mit anderen Staaten konkurrenzfähig mitwirken zu können.

15) Wie kann aus Ihrer Sicht eine einheitliche Auslegung des AI-Acts in allen EU-Mitgliedstaaten erreicht werden?

Es gibt zahlreiche positive Beispiele, wie eine einheitliche Auslegung von Verordnungen in den EU-Mitgliedsstaaten erfolgreich umgesetzt werden konnte. Als Beispiele sei hier auf die Harmonisierung in den Bereichen Datenschutz (EU Data Protection Board und Datenschutzgrundverordnung) sowie Cybersecurity (Zusammenarbeit und Informationsaustausch der nationalen Aufsichtsbehörden) verwiesen.

Dabei ist zu beachten: Die Harmonisierung weist aber auch Schwächen auf, bspw. die Langsamkeit der Verfahren und die Konsenslösungen auf dem kleinsten gemeinsamen Nenner. Ein besseres Instrument ist in der aktuellen Situation allerdings schwierig.

In den beiden oben genannten Fällen existiert eine zentrale EU-Behörde für Koordinationsaufgaben (EDPB und ENISA). Im Bereich AI fehlt eine solche europäische Behörde bisher. Der Aufbau einer solchen Institution ist potentiell sinnvoll, um die Koordination der EU-Mitgliedsstaaten in diesem sehr komplexen Themenfeld besser zu steuern. Neben Aufgaben wie der Umsetzung von Sandbox-Verfahren könnte eine solche Behörde auch die Standardisierung und Normierung vorantreiben (siehe auch Fragen 16 und 17).

16) Wie sollte Ihrer Meinung nach die Governance bei der Aufsicht und Kontrolle für KI-Anwendungen aussehen, konkret, was die Ausgestaltung des europäischen AI-Boards angeht, dessen Zusammenarbeit mit den nationalen Behörden und die Kompetenzverteilung zwischen dem AI-Board und den nationalen Behörden, und welche Kriterien sollten für die Auswahl der nationalen Behörden Ihrer Meinung nach angesetzt werden?

Ähnlich wie in der DSGVO und im Dezember 2020 vorgeschlagenen Digital Service Act, der die Schaffung eines EU Digital Services Board vorsieht, will die EU-Kommission auch für die Regulierung der KI einen European Artificial Intelligence Board nach Art. 56–58 KI-VO vorsehen, der die EU-Kommission im Wesentlichen beraten und für eine Koordinierung des Vollzugs und der Überwachung durch die nationalen Aufsichtsbehörden sorgen soll. Das European AI Board soll aus den nationalen Aufsichtsbehörden sowie den Europäischen Datenschutzbeauftragten bestehen unter Vorsitz der EU-Kommission (Art. 57 KI-VO). Während der bisherige Entwurf vorsieht, Vertreter von Wirtschaft und Wissenschaft direkt im AI Board als Mitglieder zu beteiligen, sieht der Kompromissvorschlag der tschechischen Ratspräsidentschaft lediglich vor, Experten im Rahmen von Subgroups zu beteiligen.

Unabhängig davon, wie dies organisatorisch ausgestaltet wird, muss sichergestellt werden, dass die Einbeziehung von Expertinnen und Experten verpflichtend ist und ihre Stimme entsprechendes Gewicht bei den Entscheidungen des Gremiums hat. Nur so kann der Stand von Wissenschaft und Technik angemessen in dem Gremium abgebildet werden. Insbesondere Chancen und Potentiale sollten auch berücksichtigt werden, um eine Überregulierung und Hemmung von Innovationen zu vermeiden.

Welche nationale Behörde in Deutschland diese Aufgabe der Zusammenarbeit mit dem AI-Board übernehmen kann ist unklar, ggf. müsste in Deutschland eine entsprechende Institution mit den nötigen Ressourcen und Kompetenzen geschaffen werden.

17) Wäre es Ihrer Auffassung nach sinnvoll, die geplante Verordnung um einen eigenen Titel zu „Normen und Standards“ zu erweitern? Schließlich hat die Normung im Entstehen respektive Wachstum begriffener Technologien entscheidenden Anteil an der Marktfähigkeit konkreter Lösungen und damit Marktchancen einzelner gegenwärtiger und künftiger Anbieter. Sollte die Kommission es als ihre Aufgabe begreifen, die (Normungs-)Interessen deutscher und europäischer Akteure im Bereich der Künstlichen Intelligenz in den einschlägigen internationalen Gremien mit Nachdruck zu vertreten?

Aus meiner Sicht ist eine Hinzuziehung technischer Standards über Rechtsakte der Kommission ausreichend. Technische Normen sind Instrumente der Selbstregulierung, insofern sollten hier allenfalls Zielvorgaben im Rahmen der KI-VO gemacht werden. Der Vorschlag des JURI Ausschusses des EU Parlaments zu Zielvorgaben und Werten von „Trustworthy AI“, die dann durch technische Standards weiter konkretisiert werden können, wäre ein möglicher Ansatz.

Die Kommission regt bereits seit mehreren Jahren finanziell geförderte Forschungs- und Entwicklungsprojekte dazu an, ihre Erkenntnisse und Interessen auch im Rahmen internationaler Standardisierung einzubringen. Eine zusätzliche politische Vertretung gesamteuropäischer Interessen im Bereich der Künstlichen Intelligenz durch Vertreter der Kommission selbst, einer europäischen Aufsichtsbehörde oder des EU AI Board wäre von großem Vorteil, um den europäischen Interessen Gewicht zu verleihen und gerade auch die Interessen von KMU und der Zivilgesellschaft einzubringen. Diese sind aus Gründen von finanziellen und personellen Ressourcen von zeitaufwändigen Normungsprozessen häufig ausgeschlossen. Generell sollte die EU eine stärkere Position bei KI-Normierung und -Standardisierung einnehmen, vor allem um den internationalen Anschluss nicht weiter zu verlieren und die Chancen und Potentiale europäischer Lösungen besser zu nutzen.

18) Wäre es Ihrer Auffassung nach sinnvoll, die möglichen KI-Lösungen in der geplanten Verordnung nicht nur defensiv in Risikoklassen einzuordnen, sondern komplementär in Chancen- oder Wertigkeitsklassen? Ließe sich auf diese Weise nicht das enorme Innovations- und Schöpfungspotential von KI auf einem hoch dynamischen Markt betonen, was in Deutschland und in der EU im Gegensatz zu den USA und zu China zu selten und zu zaghaft geschieht?

Eine Einteilung von KI-Systemen zusätzlich zur Risikobewertung in Chancen- oder Wertigkeitsklassen ist ein sinnvoller Vorschlag, um auch das Potential sowie den möglichen Impact von KI-Lösungen besser beurteilen zu können. Das Spannungsfeld hierbei ist, dass wir natürlich deutsche und europäische Werte beachten sowie die Sicherheit und Vertrauenswürdigkeit von KI-Lösungen (Stichwort: „Trustworthy AI“) sicherstellen müssen. Dennoch sollte das Innovations- und Wertschöpfungspotential nicht außer Acht gelassen und bei der Einschätzung ebenfalls beachtet werden. Bei der Einteilung in Chancen und Wertigkeitsklassen sollte überprüft werden, wie ein solcher Ansatz kompatibel mit der Konzeption des Produktsicherheitsrechts, wie ihn der aktuelle Entwurf der KI-VO verfolgt, gestaltet werden kann. Die Einteilung in Risikoklassen sollte beibehalten werden, hier kann Europa durch die KI-VO eine internationale Vorreiterrolle einnehmen und ein starker Brüssel-Effekt, also die faktische Übernahme von Rechtsnormen, Regulierungsmaßnahmen und Standards der EU in anderen Teilen der Welt, ist für KI-Systeme zu erwarten (analog zur DSGVO).