Kaum Schutz für die Demokratie

Der Entwurf für eine EU-Verordnung zur Regulierung Künstlicher Intelligenz geht nicht weit genug

Die geplante EU-Verordnung zur Künstlichen Intelligenz wird vermutlich weltweit die erste Regulierung dieses Feldes sein und soll die mit KI verbundenen Risiken für die Menschen begrenzen. Die WZB-Forscherinnen Jelena Cupać und Mitja Sienknecht warnen jedoch, dass der vorgelegte Entwurf der Kommission keinen ausreichenden Schutz vor Manipulation und Desinformation bietet, mit denen KI-Technologien der Demokratie schaden könnten.

Jelena Cupać und Mitja Sienknecht

Nach mehrjährigen umfassenden Konsultationen mit Expert*innen und Interessenvertretungen hat die Europäische Kommission im April 2021 den Entwurf für eine EU-Verordnung zur Künstlichen Intelligenz veröffentlicht. Diese Verordnung wird vermutlich der weltweit erste Rechtsrahmen für Künstliche Intelligenz (KI) sein und zielt darauf ab, ein Gleichgewicht zwischen der Förderung des technologischen Fortschritts und dem Schutz vor dessen vielfältigen Gefahren herzustellen. Eine gründliche Analyse des Entwurfs zeigt jedoch, dass er keinen ausreichenden Schutz vor Schäden bietet, die KI-Technologien der Demokratie zufügen können. Er lässt zu viel Raum für Manipulations- und Desinformationstechniken, die die Qualität von Informationen, die für das angemessene Funktionieren des demokratischen Prozesses benötigt werden, beeinträchtigen. Wenn die Verordnung in ihrer endgültigen Fassung dieser Art von KI-Schäden nicht mehr Aufmerksamkeit widmet, wird die Zukunft der Demokratie in Europa einer Handvoll von Unternehmen überlassen – und all denjenigen, die bereit sind, ihre Datenbestände und KI-Kompetenzen für persönliche politische Vorteile zu nutzen.

In den letzten zwei Jahrzehnten sind digitale Technologien, auch die durch Künstliche Intelligenz betriebenen, zu einem integralen Bestandteil der politischen Kommunikation geworden. Sie sind untrennbar mit der Demokratie verbunden. Sie bestimmen nicht nur, wie die Bürger die Informationen erhalten, die sie zur Teilnahme am politischen Leben ihrer Gesellschaft benötigen, sondern bestimmen auch die Qualität dieser Informationen. Wie sie die Demokratie verändern werden, ist noch offen, aber die bisherigen Erfahrungen geben wenig Anlass zu Optimismus.

Der bisher bekannteste Fall, in dem soziale Medien und KI-Technologien zur Beeinflussung des demokratischen Prozesses eingesetzt wur-

den, ist die Arbeit von Cambridge Analytica in Donald Trumps Wahlkampf 2016. Dies ist jedoch bei Weitem nicht der einzige Fall von technologiebasierter Wahlmanipulation. Allein für das Jahr 2019 haben die Forscher*innen Samantha Bradshaw und Philip Howard Manipulationen in 48 Ländern nachgewiesen.

Zu den dabei am häufigsten eingesetzten KI-basierten Techniken gehören Profiling, Targeting, Social Bots und Deepfakes. Unter Profiling versteht man die Segmentierung von Wähler*innengruppen auf der Grundlage allgemeiner Merkmale wie Alter und Bildung oder persönlicher Eigenschaften wie Introvertiertheit oder Extrovertiertheit. Die Profilerstellung wird jedoch nicht um ihrer selbst willen durchgeführt. Ihr Hauptzweck besteht darin, eine präzise Ziel-

„Allein für das Jahr 2019 konnten KI- und Social-Media-basierte Wahlmanipulationen in 48 Ländern nachgewiesen werden“

gruppenansprache – ein sogenanntes Targeting – zu ermöglichen, die Menschen dazu bringen soll, in einer bestimmten Weise zu denken und sich zu verhalten. Das Targeting erfolgt in der Regel über soziale Medien und kann zur Bildung von sogenannten Echokammern und Filterblasen führen, in denen Menschen hauptsächlich mit Gleichgesinnten interagieren und daher von alternativen Informationen und Gegenargumenten abgeschirmt sind.

Social Bots – also teilweise oder vollständig von Algorithmen gesteuerte Social-Media-Konten – werden zunehmend als Targeting-Instrumente eingesetzt. Sie sind so konzipiert, dass sie die Praktiken von Menschen bei der Erstellung und Verbreitung politischer Inhalte glaubhaft imitieren; ihr Hauptzweck besteht darin, eine Idee zu übertreiben oder einen politischen Kandidaten zu diskreditieren. Der Wahrheitsgehalt einer Nachricht spielt dabei keine Rolle, was Social Bots zu einem wichtigen Instrument für die Online-Verbreitung von Desinformationen und Fake News macht. Bots breiten sich zunehmend in der öffentlichen Sphäre aus und bergen die Gefahr, Verwirrung, Vorurteile und Misstrauen zu säen und so die Grundlagen unserer Demokratie zu gefährden.

Jelena Cupać *ist wissenschaftliche Mitarbeiterin der Abteilung Global Governance. Ihr Forschungsschwerpunkt liegt bei der Transformation der internationalen Sicherheitsorganisationen.*
jelena.cupac@wzb.eu

Eine andere KI-gestützte Technologie, die ein ähnliches Risiko birgt, sind Deepfakes. Deepfakes sind Multimedia-Fälschungen, bei denen eine Person in einem Bild oder Video durch das Bild einer anderen Person ersetzt wird. Das Ergebnis sind äußerst realistische Video- oder Fotoinhalte, die dazu verwendet werden können, den politischen Gegnern zu schaden, indem sie sie beispielsweise bei diskreditierenden oder kriminellen Aktivitäten zeigen. Deepfakes sind besonders effektiv, wenn sie der Öffentlichkeit kurz vor Wahlen präsentiert werden, da nur wenig Zeit bleibt, um sie rechtzeitig zu entlarven.

Vor diesem Hintergrund kommt dem Vorschlag der Europäischen Kommission für eine EU-Verordnung zur Künstlichen Intelligenz eine hohe Bedeutung zu. Doch obwohl der Skandal um Cambridge Analytica bekannt ist, obwohl man weiß, wie Profiling, Targeting, Social Bots

„Der Entwurf versäumt es, die Demokratie vor KI-basierter Manipulation und Desinformation zu schützen“

und Deepfakes funktionieren, und vor allem, obwohl die Demokratie zu den sechs Grundwerten der EU gehört, hat der Entwurf es versäumt, die Demokratie vor KI-basierter Manipulation und Desinformation zu schützen.

Die Verordnung verfolgt einen risikobasierten Ansatz für KI und unterscheidet zwischen unannehmbaren, hohen, begrenzten und geringen oder minimalen Risiken. KI-Systeme, die als unannehmbar eingestuft werden, sind verboten; Systeme, die als hohes Risiko eingestuft werden, unterliegen Konformitätsbewertungen, bevor sie Zugang zum EU-Markt erhalten; Systeme mit begrenztem Risiko unterliegen einer Reihe von Transparenzverpflichtungen; und Systeme mit geringem oder minimalem Risiko erfordern die Einhaltung von Verhaltenskodizes und, wo nötig, die Beteiligung von Interessengruppen. Wie wir im Folgenden zeigen werden, geht keine dieser Kategorien angemessen auf die Risiken für die Demokratie ein, die sich aus den oben beschriebenen Manipulations- und Desinformationstechniken ergeben.

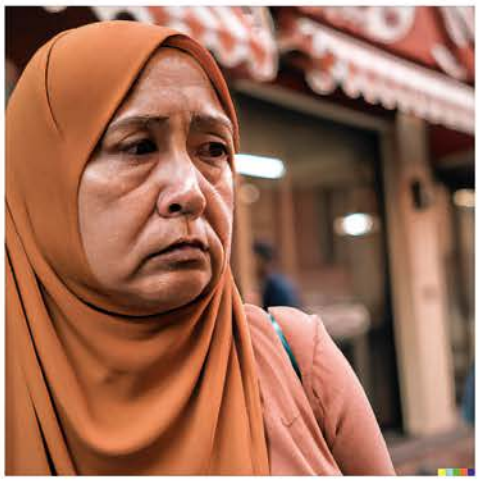
Die KI-Systeme, die der Entwurf als unannehmbares Risiko ansieht, sind Systeme, die für Manipulation, Social Scoring und die Sammlung und Verarbeitung biometrischer Daten verwendet werden können. Wenn man bedenkt, dass die bereits erwähnten Praktiken der Profilerstellung und des Targeting als Manipulationen der politischen Meinung und des politischen Diskurses sowie des allgemeinen demokratischen Prozesses interpretiert werden können, könnte man erwarten, dass sie von dieser Bestimmung des Gesetzes erfasst werden. Das ist jedoch nicht der Fall. Die Verordnung sieht Manipulation nur in den KI-Praktiken, die Personen durch unterschwellige Techniken oder durch Ausnutzung der Schwächen von Kindern oder Menschen mit Behinderung in einer Weise beeinflussen können, die ihnen psychischen oder physischen Schaden zufügen könnten. Die Bestimmungen zur Manipulation konzentrieren sich daher in erster Linie auf die psychische und körperliche Schädigung von Individuen, ohne sich um die allgemeine gesellschaftliche Schädigung und deren Folgen für die Demokratie zu kümmern. In der Begründung des Entwurfs heißt es, dass andere manipulative Praktiken schon durch die bestehenden EU-Rechtsvorschriften zum Datenschutz, zum Verbraucherschutz und zu digitalen Diensten abgedeckt werden könnten, um „natürlichen Personen“ zu garantieren, angemessen informiert zu werden, und ihnen zu ermöglichen, sich gegen Profiling oder andere Praktiken zu entscheiden, die ihr Verhalten beeinflussen könnten. Doch auch dieser Ansatz löst das Problem nicht. Auch bei ihm liegt der Schwerpunkt nach wie vor auf dem individuellen Schaden und nicht auf dem Schaden für die Gesellschaft.

Die Verordnung behandelt risikoreiche KI, indem es acht Bereiche auflistet, in denen die Anwendung von KI unerwünschte Schäden verursachen könnte. Die Liste reicht von der Verwaltung und dem Betrieb kritischer Infrastrukturen bis hin zur Strafverfolgung. Wichtig ist, dass sie auch die „Rechtspflege und den demokratischen Prozess“ umfasst. Indem sie den „demokratischen Prozess“ in die Hochrisikokategorie einordnet, signalisiert die Europäische Kommission ein Bewusstsein dafür, dass KI der

„Die EU-Kommission signalisiert ein Bewusstsein dafür, dass KI der Demokratie schaden könnte, geht aber nicht darüber hinaus“

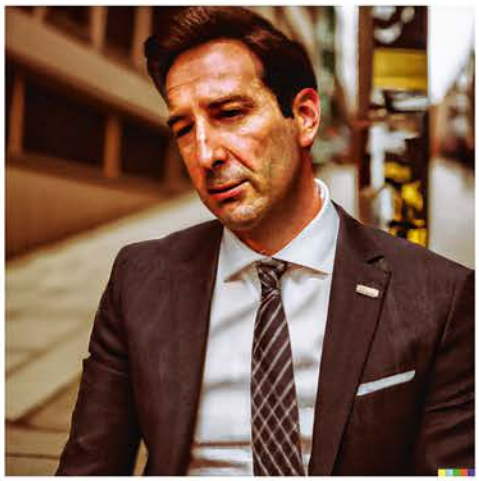
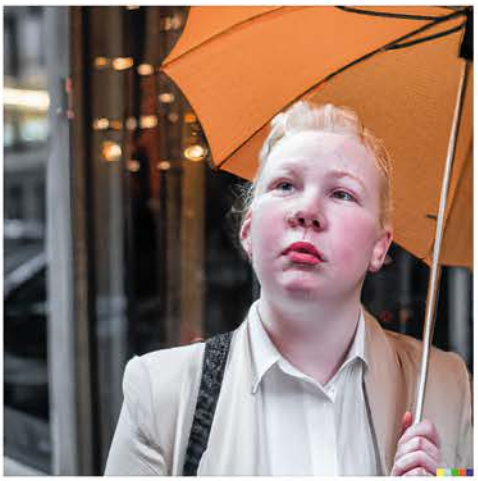
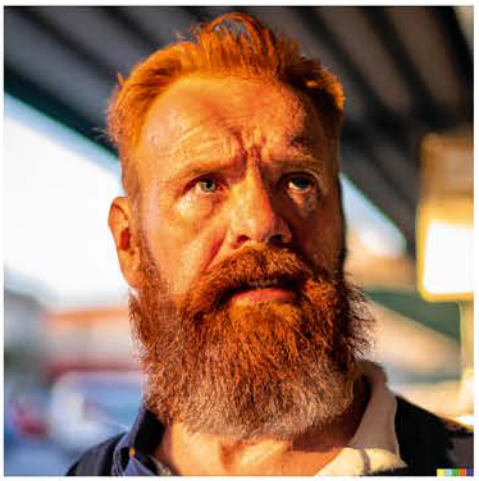
Demokratie schaden könnte. Allerdings geht sie nicht über diese Signale hinaus. Weder im Haupttext des Entwurfs noch in seinem Anhang wird näher ausgeführt, welche Art von KI-bedingten Störungen im demokratischen Prozess zu sehen sein müssten, um sie als hochriskant zu bewerten. Demografisches und psychometrisches Profiling, Targeting, Desinformation oder Deepfakes werden nicht erwähnt. Dieser Ansatz steht im Widerspruch zu der Art und Weise, wie das Gesetz andere Hochrisikobereiche behandelt, einschließlich der Justizverwaltung, die zusammen mit dem demokratischen Prozess aufgeführt wird. Im Gesetz heißt es, dass „KI-Systeme, die Justizbehörden dabei helfen [sollen], Sachverhalte und Rechtsvorschriften zu ermitteln und auszulegen und das Recht auf konkrete Sachverhalte anzuwenden“, ein hohes Risiko darstellen könnten. Es ist rätselhaft, warum der demokratische Prozess nicht in ähnlicher Weise behandelt wurde.

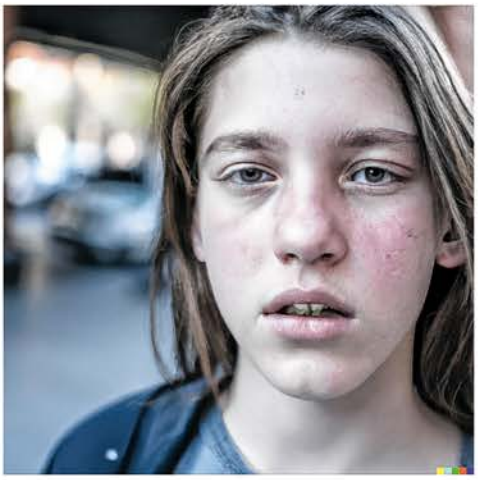
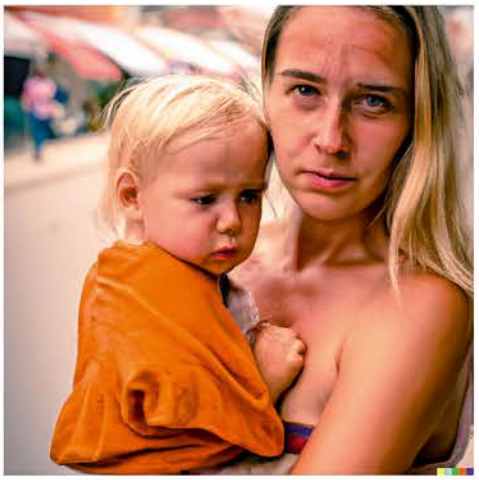
Die KI-Systeme, die in der Verordnung als begrenzt risikobehaftet eingestuft werden, stehen in direktem Zusammenhang mit den Systemen und Praktiken, die oben bereits als größte Bedrohung für die Demokratie erörtert wurden: Social Bots, Emotionserkennungssysteme, biometrische Kategorisierungssysteme und Deepfakes. Diese Systeme, die zur Profilerstellung, Desinformationsverbreitung und Meinungsmanipulation eingesetzt werden, können den demokratischen Prozess tiefgreifend stö-





Diese Personen gibt es nicht.
Ihre Gesichter sind geschaffen von Gesine Born mit Hilfe der Software DALL-E 2.
Ihre Eingabe: "portrait photograph of {... description; z.B. woman with a baby}, looking worried, street photography, Leica style, 35 mm, warm colors".

ren, indem sie die Wähler*innen abschotten und polarisieren, Emotionen wie Angst und Wut bedienen und das Vertrauen in die Demokratie insgesamt untergraben. Der Entwurf zeigt jedoch kein Bewusstsein für diese Folgen. Er stellt lediglich fest, dass Bots, Emotionserkennung, Kategorisierungssysteme und Deepfakes ein „spezifisches Risiko der Manipulation" darstellen können. Dieses spezifische Risiko wird aber nicht erläutert, obwohl der Entwurf postuliert, dass einfache Transparenzpflichten zum Schutz davor ausreichen sollten. So sollen laut Entwurf Anbieter von Social Bots sicherstellen, dass ihre Systeme so konzipiert und entwickelt werden, dass natürliche Personen darüber informiert werden, dass sie mit einer KI und nicht mit einem Menschen interagieren; Personen, die einem Emotionserkennungssystem oder einem biometrischen Kategorisierungssystem unterliegen, sollten darüber informiert werden, dass sie solchen Systemen ausgesetzt sind; und Nutzer*innen von Deepfakes sollten deutlich machen, dass die Inhalte künstlich erzeugt oder manipuliert wurden. Mit dem Argument, dass diese Transparenzmaßnahmen es den Menschen ermöglichen würden, eine informierte Entscheidung zu treffen oder sich aus einer bestimmten Situation zurückzuziehen, zeigt die Verordnung erneut, dass es KI-Risiken in erster Linie als individuelle Risiken und nicht als soziale oder politische Risiken ansieht.

Insgesamt ist der Entwurf für die EU-KI-Verordnung ein schwaches Instrument zum Schutz der Demokratie vor bestehenden und zukünftigen KI-Schäden. Indem er die Manipulation durch Techniken wie Profiling, Targeting, Bots und Deepfakes uneingeschränkt zulässt und den von ihnen verursachten Schaden als individuell und nicht als gesellschaftlich definiert, ebnet der Entwurf den Weg für beispiellose und disruptive Veränderungen in der europäischen Demokratie.

Mitja Sienknecht ist Lehrstuhlvertreterin am Lehrstuhl für Europäische und Internationale Politik der Europa-Universität Viadrina Frankfurt (Oder). Sie war Gastwissenschaftlerin in der Abteilung Global Governance am WZB.
sienknecht@europa-uni.de

Noch ist jedoch nicht alles verloren. Das Gesetzgebungsverfahren läuft noch, und noch besteht die Möglichkeit, die Demokratie vor Manipulation und Fehlinformation zu schützen. So müssen vor der Verabschiedung zwei weitere Organe der EU ihre Stellungnahmen und Änderungsanträge abgeben: das Europäische Parlament und der Europäische Rat. Und es gibt Anzeichen dafür, dass diese einen umfassenderen Schutz der Demokratie fordern werden, indem sie Targeting, Profiling, Bots und Deepfakes nicht nur Transparenzanforderungen, sondern auch Risikobewertungen unterwerfen.

Literatur

Bradshaw, Samantha/Howard, Philip N.: The Global Disinformation Order: 2019 Global Inventory of Organized Social Media Manipulation. Oxford: University of Oxford 2019.

Edwards, Lilian: Regulating AI in Europe: Four Problems and Four Solutions. Expert Opinion. Ada Lovelace Institute 2022.

Edwards, Lilian: The EU AI Act: A Summary of Its Significance and Scope. Expert Opinion. Ada Lovelace Institute 2022.

Veale, Michael/Zuiderveen Borgesius, Frederik: „Demystifying the Draft EU Artificial Intelligence Act – Analysing the Good, the Bad, and the Unclear Elements of the Proposed Approach". In: Computer Law Review International, 2021, Jg. 22, H. 4, S. 97-112.